Forthcoming, 35 BTLJ (2020)
11/8/2019 Draft

CAN YOU PAY FOR PRIVACY? CONSUMER EXPECTATIONS AND THE BEHAVIOR OF FREE AND PAID APPS

Kenneth A. Bamberger,† *Serge Egelman,*†† *Catherine Han,*††† *Amit Elazari Bar On*‡ *&*
Irwin Reyes‡‡

ABSTRACT

"Paid" digital services have been touted as straightforward alternatives to the ostensibly "free" model, in which users actually face a high price in the form of personal data, with limited awareness of the real cost incurred and little ability to manage their privacy preferences. Yet, the actual privacy behavior of paid services, and consumer expectations about that behavior, remain largely unknown.

This Article addresses that gap. It presents empirical data both comparing the true cost of "paid" services as compared to their so-called "free" counterparts, and documenting consumer expectations about the relative behaviors of each.

We first present an empirical study that documents and compares the privacy behaviors of 5,877 Android apps that are offered both as free and paid versions. The sophisticated analysis tool we employed, AppCensus, allowed us to detect exactly which sensitive user data is accessed by each app and with whom it is shared. Our results show that paid apps often share the same implementation characteristics and resulting behaviors as their free counterparts. Thus, if users opt to pay for apps to avoid privacy costs, in many instances they do not receive the benefit of the bargain. Worse, we find that there are no obvious cues that consumers can use to determine when the paid version of a free app offers better privacy protections than its free counterpart.

We complement this data with a second study: we surveyed 1,000 Android mobile app users as to their perceptions of the privacy behaviors of paid and free app versions. Participants indicated that consumers are more likely to expect that the free version would share

DOI: https://doi.org/10.15779/Z38XP6V40J

 This research was supported by The Center for Long-Term Cybersecurity (CLTC) at UC Berkeley, the National Science Foundation (NSF) under grant CNS-1817248, and the National Security Agency's (NSA) Science of Security Program under contract H98230-18-D-0006. Sanjana Parikh provided superb research assistance.

† The Rosalinde and Arthur Gilbert Professor of Law, UC Berkeley; Faculty Co-Director, Berkeley Center for Law and Technology.

†† Research Director, Usable Security & Privacy Group, International Computer Science Institute (ICSI); Research Scientist, Electrical Engineering and Computer Sciences, UC Berkeley.

††† Undergraduate student, UC Berkeley.

‡ Lecturer, UC Berkeley School of Information.

‡‡ Researcher, Usable Security and Privacy Group, International Computer Science Institute (ICSI).

their data with advertisers and law enforcement agencies than the paid version, and be more likely to keep their data on the app's servers when no longer needed for core app functionality. By contrast, consumers are more likely to expect the paid version to engage in privacy-protective practices, to demonstrate transparency with regard to its data collection and sharing behaviors, and to offer more granular control over the collection of user data in that context.

Together, these studies identify ways in which the actual behavior of apps fails to comport with users' expectations, and the way that representations of an app as "paid" or "ad-free" can mislead users.

These findings have important implications for law and policy intended to protect personal data. Specifically, they underscore existing research about the ways that consumers develop understandings (and misunderstandings) about data use, and the ways in which regulatory regimes purportedly premised on consent can thwart user intentions and vitiate that notion.

More specifically, they point to a wide divergence between expectations and reality and provide important evidence about the pervasive and unanticipated collection and sharing of data. While some of that behavior might be captured in evolving legal regimes, notably the GDPR, much is not. This failure suggests the importance of expanding regulatory approaches that on one hand reflect the ways that inaccurate expectations might lead to user deception, and on the other recognize that consumer confusion can vitiate notions of consumer choice by eliminating privacy as a salient factor in consumer decision-making altogether.

Finally, our research suggests the important role that the tools developed in our research can play in empowering and protecting users by making app behavior more transparent. These tools can reduce the opacity of app behavior and information asymmetries that corrupt the market for privacy.

Our study demonstrates that, at least in the most dominant example of a free versus paid market—mobile apps—there turns out to be no real privacy-protective option. Yet the failures of transparency or auditability of app behaviors deprive users, regulators, and law enforcement of any means to keep developers accountable, and privacy is removed as a salient concern to guide user behavior. Dynamic analysis of the type we performed can both allow users to go online and test, in real-time, an app's privacy behavior, empowering them as advocates and informing their choices to better align expectations with reality. The same tools, moreover, can equip regulators, law enforcement, consumer protections organizations, and private parties seeking to remedy undesirable or illegal privacy behavior.

TABLE OF CONTENTS

I. INTRODUCTION

Users pay a high price to enjoy "free" digital services as they engage in the most prominent *quid pro quo* of the digital age: the exchange of personal information and privacy for utility and comfort. At the same time, amid ensuing media attention to growing data abuses, businesses have come to recognize that users are often willing to expend a small monetary sum for an "ad-free" experience. Many companies have promoted such "paid" services as a straightforward alternative, in which users can choose to pay with money, instead of with personal privacy.[1]

The free model has dominated many provinces of digital space. In aggregate, free mobile applications attract over ten times the volume of downloads as paid apps.[2] Rather than charging consumers directly, free app developers generate significant revenue in other ways, such as partnering with advertising networks to provide ads to users. Google's AdMob, for instance, is found in more than 1 million apps, and has yielded more than $1 billion collectively to developers.[3]

Scholarship has increasingly critiqued the free model, revealing its true cost.[4] Properly understood, "free" transactions are anything but. They are, instead, exchanges between consumers and services collecting their data[5]—unequal exchanges, moreover, in which users possess limited awareness of the real costs incurred, and little ability to manage their privacy preferences.[6] This

1. *See infra* Part II.A.

2. D. Amalfitano, A. R. Fasolino, P. Tramontana, B. D. Ta, &A. M. Memon, *MobiGUITAR: Automated Model-Based Testing of Mobile Apps*, IEEE SOFTWARE 2015.

3. *Get paid to show relevant ads from over a million advertisers with Google AdMob*, GOOGLE, https://developer.android.com/distribute/best-practices/earn/show-ads-admob [https://web.archive.org/web/20181129004421/https://developer.android.com/distribute/best-practices/earn/show-ads-admob] (last visited Nov. 28, 2018).

4. *See, e.g.*, John M. Newman, *The Myth of Free*, 86 GEO.WASH. L. REV. 513, 520 (2018); Chris Jay Hoofnagle & Jan Whittington, *Free: Accounting for the Costs of the Internet's Most Popular Price*, 61 UCLA L. REV. 606, 613 (2014) [hereinafter *Free*]; Katherine J. Strandburg, *Free Fall: The Online Market's Consumer Preference Disconnect*, 2013 U. CHI. LEGAL F. 95, 95 (2013).

5. Hoofnagle & Whittington, *Free*, *supra* note 4, at 608. *See also* Jan Whittington & Chris Jay Hoofnagle, *Unpacking Privacy's Price*, 90 N.C. L. REV. 1327 (2012) (applying transaction cost economics to define the relationship between consumers and social networks as an exchange).

6. *See* Jonathan A. Obar & Anne Oeldorf-Hirsch, *The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services*, 2018 INFO., COMMC'N & SOC. 1, 16 (noting that of more than 500 surveyed users, 93% accepted a "first-born child assignment" term and 98% ignored or missed it).

scholarship documents "the many reasons consumers find it impossible to account for the risk of harm from online data collection." [7] The opacity of app behavior obscures the extensive processing of personal information and the fact that the collection and sale of user information "is the main business proposition."[8] The misdirection of privacy policies and the framing effects of the "myth" of free, moreover, exacerbate the "privacy paradox," [9] by which consumers behave inconsistently with their actual privacy-protective preferences when it comes to decisions about personal information.

The paid model of digital services has been touted as a means to solve these inadequacies, and enhance consumer choice regarding privacy.[10] User fees offer a substitute for ad revenues, and—potentially—from the intrusive data collection that fuels targeted advertising.[11] Users paying for apps generally

---

7. *See*, *e.g.*, Paul Ohm, *Free for the Taking (or Why Libertarians are Wrong about Markets for Privacy)* (May 26, 2014), JOTWELL, http://cyber.jotwell.com/free-for-the-taking-or-why-libertarians-are-wrong-about-markets-for-privacy/ ("*Free* and *Free Fall* document the many reasons consumers find it impossible to account for the risk of harm from online data collection.").

8. Chris Jay Hoofnagle & Jan Whittington, *Free*, *supra* note 4, at 606, 613, 628; *see also id.* at 620 ("[B]ut firms online are like firms offline—both spend money to generate their products and both must recoup costs to survive."). *See also id.* at 628 ("For these types of businesses and the third parties they market to, the collection and sale of personal information about consumers is the main business proposition."). *See also id.* at 633. The article stated:

If, for any reason, a firm is unable to earn enough revenue from either ads or paying customers, the firm can simply sell the personal information on the market. The firm may not even have intended to capitalize on the personal information it collected with each transaction, free or otherwise.

Id.

9. *See e.g.*, Alessandro Acquisti, Laura Brandimarte, & George Loewenstein, *Privacy and Human Behavior in the Age of Information*, 347 SCIENCE 509, 510 (2015) ("This discrepancy between attitudes and behaviors has become known as the 'privacy paradox.'").

10. *See*, *e.g.*, David Z. Morris, *Sheryl Sandberg Says Facebook Users Would Have to Pay for Total Privacy*, FORTUNE (Apr. 7, 2018), *available at* http://fortune.com/2018/04/07/sheryl-sandberg-says-facebook-users-would-have-to-pay-for-total-privacy/.

11. *See* Stacy-Ann Elvy, *Paying for Privacy and the Personal Data Economy*, 117 COLUM. L. REV. 1369, 1373, 1373 n.16 (2017) (describing the pay-for-privacy (PFP) approach, "which requires consumers to pay higher fees to avoid data collection and targeted advertisements while offering discounts to consumers who consent to these practices") (citing Letter from Senator Elizabeth Warren to Tom Wheeler, Chairman, FCC 2 (June 21, 2016), "describing Internet service provider discount plans as 'requir[ing] consumers to pay hundreds of dollars extra each year so that [a company] does not collect and sell information on the websites they visit, the ads they see, and the terms they enter into search engines'"), *available at* http://www.warren.senate.gov/files/documents/2016-6-21_Letter_to_FCC_re_Privacy_Rulemaking.pdf [http://perma.cc/9WWT-7362]; *see also* Michael R. Hammock & Paul H. Rubin, *Applications Want to be Free: Privacy Against Information*, COMP.

expect them to be of higher quality compared to free versions, and the removal of ads in the paid version may be understood (rightly or wrongly) as implying freedom from the associated extensive data collection.[12] Media outlets, moreover, have reflected such expectations, crediting paid apps with having better security and privacy assurances than free apps.[13]

Yet, unlike the free business model, the paid model for digital services has largely evaded scholarly attention. Moreover, the actual privacy behavior of paid apps, and consumer expectations about that behavior have largely evaded scholarly attention. This Article addresses that gap. It presents empirical data both comparing the true cost of "paid" services as compared to their so-called "free" counterparts, and documenting consumer expectations about the relative behaviors of each.

We first present an empirical study that applied a scalable analysis framework to document and compare the privacy behaviors of thousands of Android apps that are offered both as free and paid versions.[14] Our method employed static and dynamic analysis: static analysis to determine the third-party Software Development Kits (SDKs) bundled with each app and the permissions that each app requests; and dynamic analysis to monitor what sensitive data is collected by which remote services in realtime, as each app is executed. From a random sample of free apps listed on the Google Play Store's category-level top charts, we examined thousands of pairs of free apps and their paid counterparts. The sophisticated analysis tool we employed, AppCensus, developed through a collaboration at the International Computer Science Institute

---

POL'Y INT'L. (2011) (suggesting that regulators ignore customer concerns and allow the market to provide solutions to concerned consumers to enhance their privacy). *Id.*

12. *See* M. Panzarino, *Why You Should Want to Pay for Apps*, THE NEXT WEB (Apr. 24, 2011), https://thenextweb.com/apps/2011/04/24/why-you-should-want-to-pay- for-apps/ [https://web.archive.org/web/20181129005820/https://thenextweb.com/apps/2011/04/24/why-you-should-want-to-pay-for-apps/]; Max Van Kleek, et al., *X-Ray Refine: Supporting the Exploration and Refinement of Information Exposure Resulting from Smartphone Apps*, ACM CHI CONF. ON HUM. FACTORS COMPUTING SYSS. (2018) ("Free apps (or freemium versions of apps) were naturally expected to send data to more companies than their paid counterparts because paid apps were perceived to need less ad support.").

13. S. Angeles, *Are Free Apps Safe?*, BUS. NEWS, businessnewsdaily.com/4868-free-app-security-risk.html [https://web.archive.org/web/20181129010454/https] (last visited Nov. 28, 2018).

14. *See infra* Part II.B(1).

(ICSI), [15] allowed us to detect exactly which sensitive user data is accessed by each app and with whom it is shared.[16]

Utilizing our framework to compare 5,877 paired apps in their "free" and "paid" versions, we examined whether the cost paid by the user (in privacy terms) was in fact lower in paid services, challenging the common conception that paid services are more secure, offer stronger privacy protections, and share less personal data. Our results show that paid apps often share the same implementation characteristics and resulting behaviors as their free counterparts: 48% of the paid apps we examined carried all of the same third-party code (e.g., for advertising, analytics, graphics rendering, logging, etc.) as their free versions; 56% of paid apps had all the same privileges to access sensitive system resources; and 38% of paid apps collected all the same personal and tracking information about users as their free versions. Thus, if users opt to pay for apps to avoid privacy costs,[17] in many instances they do not receive the benefit of the bargain. Worse, we find that there are no obvious cues that consumers can use to determine when the paid version of a free app offers better privacy protections than its free counterpart.[18]

We complement this data with a second study: we surveyed a large sample of mobile app users as to their perceptions of the privacy behaviors of paid and free app versions.[19] The survey explored consumers' expectations around privacy with regard to different app versions, and specifically considered whether, when apps are advertised as "ad-free," most consumers believe that this is synonymous with "better privacy." Our findings suggest that consumers

---

15.AppCensus was started as a research project at the International Computer Science Institute (ICSI), which is a research institute affiliated with UC Berkeley, and has since been spun off as an independent startup, see https://search.appcensus.io/about and https://www.appcensus.io/.

[16] AppCensus AppSearch analyzes free publicly-available Android apps and reports the private and personally identifiable information that different apps access and share with other parties over the Internet, what personal data is being accessed by an app, and then with whom that app shares it. The results reflect the actual behavior of the apps when they are used. *See* https://search.appcensus.io/about.

17. *Data Privacy: What the Consumer Really Thinks*, ACXIOM (February 2018), https://dma.org.uk/uploads/misc/5a857c4fdf846-data-privacy---what-the-consumer-really-thinks-final_5a857c4fdf799.pdf ("a sizeable proportion of consumers indicate that they would prefer to pay for online services so that they do not have to share any personal data.").

18. *See generally Future of Privacy Forum*, FPF MOBILE APPS STUDY, http://www.futureofprivacy.org/wp-content/uploads/Mobile-Apps-Study-June-2012.pdf (finding that only 48% of free apps and 32% of paid apps provide in-app access to a privacy policy).

19.*See infra* Part II.B(2).

not only expect more privacy-preserving behaviors from the paid app in practice, but are likelier to assume a higher level of transparency from the paid version about its data collection and sharing behaviors, and more granular control over the collection of their data. Thus, the mere act of paying for apps—regardless of how much—is associated with receiving better privacy. Nonetheless, when asked in open-ended questions at the beginning of the survey (before being primed to consider privacy and security), a large majority (83%) of participants indicated that they would choose to buy the free app version, and nearly none referenced privacy or security behaviors as a driving consideration. So while our results indicate that consumers do care about privacy and believe that paying for apps yields better privacy, this is far from the only factor that consumers consider when choosing whether or not to purchase a given app.

Our combined research identifying ways in which the actual behavior of apps fails to comport with users' expectations has important ramifications for policy and practice.

First, our findings that paid apps often conduct equally extensive levels of data collection and sale as free ones offers important information about how the "paid" model operates in practice. This model is already critiqued for privileging those with more resources by letting them "buy out" of certain exploitative practices to which others are subject.[20] If, in fact, the economics of this model also relies on widespread data collection and sharing, its promise as a privacy-protective alternative is questionable.

So, moreover, may be its legality. The treatment of data in ways that are not necessary for the provision of an app's service triggers the European General Data Protection Regulation's (GDPR) requirement that there be another legal basis for this processing of personal data (e.g., explicit consent).[21] While

20. *See*, *e.g.*, Sophia Cope & Jeremy Gillula, *AT&T is Putting a Price on Privacy. That is Outrageous*, GUARDIAN (Feb. 20, 2015), https://www.theguardian.com/commentisfree/2015/feb/20/att-price-on-privacy (citing AT&T's practice of charging customers extra money for increased privacy protections).

21. Recital 43, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L. 119) 1. Recital 43 states:
Consent is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual case, or if the performance of a contract, including the provision of a service, is made dependent on the consent despite such consent not being necessary for such performance.

commentators have suggested that these and similar provisions in other recent privacy laws could prove the death knell for the free model and privilege the position of paid apps,[22] our data suggests that much paid app behavior also fictionalizes notions of consent, calling into question its legality.

Second, by providing empirical foundations for better understanding both corporate behavior and consumer expectations, our findings offer important insights for legal approaches to privacy protection.[23] Specifically, they further undermine the legitimacy of legal regimes relying on fictive "notice" and "consent" that do not reflect user understandings as bases for the collection, sale, and processing of information. At a minimum, they fortify demands for a privacy law that focuses on vindicating actual consumer expectations, and prohibiting practices that exploit them.[24] More broadly, they strengthen the argument for ex ante regulation of exploitative data practices where consumers are offered no opportunity for meaningful choice or consent. Amid the adoption of a new statutory privacy regime in California and increased discussion regarding omnibus Federal Privacy legislation, these findings can inform policymakers as they seek to implement broader privacy protections for users.[25]

Finally, building on our evidence that users often misunderstand technological models, to their detriment, this paper demonstrates the need for technical tools that offer transparency about app behaviors, and empower consumer choice.[26] Our study demonstrates that, at least in the most dominant example of a free versus paid market—mobile apps—there turns out to be no real privacy-protective option. The failures of transparency or auditability of

---

Id.

22. *Will Free Apps Soon be Dead in Europe?*, MINTZ (Feb. 2016), https://www.mintz.com/insights-center/viewpoints/2826/2016-02-will-free-apps-soon-be-dead-europe.

23. *See infra* Part III.B.

24. Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 667 (2016) (identifying actions by the agency that could form a possible basis for a move in enforcement towards preventing "broken expectations of consumer privacy"); *See also generally* KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE 64 (2015) (discussing an evolving orientation among privacy leaders towards consumer expectations).

25. *See* California Consumer Privacy Act of 2018, Assembly Bill No. 375, ch. 55: An Act To Add Title 1.81.5 (commencing with Section 1798.100) to Part 4 of Division 3 of the Civil Code, relating to privacy (Filed with Secretary of State June 28, 2018) [hereinafter CCPA].

26. *See infra* Part IV.B.

app behaviors, moreover, mean that consumers have no way to identify instances in which paid apps might actually be limiting data collection, and actually offering more privacy-protective options. Those failures also deprive users, regulators, and law enforcement of any means to keep developers accountable. Despite the touted potential for paid models to support consumer choice and privacy protection, then, these information failures destroy any opportunity for a meaningful privacy market. Without information about app privacy practices, privacy is removed as a salient concern for users faced with a free or paid choice.

Accordingly, this work demonstrates how dynamic analysis of the type we performed in this study could serve as a tool for empowering app users. Building in such tools can allow users to go online and test, in real-time, an app's privacy behavior, revealing collection practices formerly obscured within the "black box." This technology could therefore empower users to be advocates, and inform their choices to better align their expectations with reality. More systemically, these tools could facilitate the creation of third-party certification markets, which could then test apps and label them based on their observed behaviors, thus offloading the burden from consumers. The same tools, moreover, could equip regulators, law enforcement, consumer protections organizations, and private parties seeking to remedy illegal privacy behavior through the civil system.

While automation and technology are often viewed as counter-beneficial to privacy in the context of data collection, we suggest they can serve a vital function in streamlining and scaling both user decision making and regulatory enforcement. Wedding laws strengthening consumer protection with the types of privacy-enhancing technologies often overlooked by privacy regulations should be front and center in any consumer-focused legislative effort.

II. PAYING FOR PRIVACY: OUR RESEARCH IN CONTEXT

A. PAYING FOR PRIVACY

When Facebook COO Sheryl Sandberg said in interviews last year that allowing users the option to completely opt-out of tracking and data profiling would require a "paid product," [27] her comments echoed increased traction for

27. Andrew Albanese & Annie Coreno, *Are We Headed for a Pay-for-Privacy World?*, PUBLISHERS WKLY. (Apr. 6, 2018), https://www.publishersweekly.com/pw/by-topic/industry-news/libraries/article/76530-are-we-headed-for-a-pay-for-privacy-world.html.

the proposition of offering users the option of paying (or paying more) to limit the use and dissemination of their personal information.[28] Broadband Internet Service Providers (ISPs) have piloted offerings that permit customers to opt-out of surveillance for an extra fee.[29] The landmark California Consumer Privacy Act—the farthest-reaching privacy legislation in the United States—opens the door for differential pricing based on the right to sell or share some kinds of data.[30]

The availability of paid versions of digital services provides a substitute for targeted advertising revenues that drive the free services model.[31] Commentators therefore have touted it as an important means to empower consumer choice regarding the use of their personal information.[32] Indeed, some argue, preserving a choice between free and paid options is important in increasing access for low-income consumers, who might otherwise be priced out of services.[33] Especially in light of the fact that most users do not actually choose to

28. Constine, *How ad-free subscriptions could save Facebook*, TECH CRUNCH (Feb. 17, 2018), https://techcrunch.com/2018/04/15/would-it-make-us-love-or-hate-ads/ (updated Apr. 10, 2018).

29. *See* Sophia Cope & Jeremy Gillula, *AT&T is Putting a Price on Privacy. That is Outrageous*, THE GUARDIAN (Feb. 20, 2015), https://www.theguardian.com/commentisfree/2015/feb/20/att-price-on-privacy. (describing AT&T's piloting of a service which allows gigabit service customers to opt-out of surveillance for $29 per month.).

30. Allen St. John, *How California's New Privacy Law Could Affect You (Even If You Don't Live There)*, CONSUMER REP. (June 29, 2018), https://www.consumerreports.org/privacy/how-californias-new-privacy-law-could-affect-you/ ("The most controversial provision of the new law allows companies to provide a discount in exchange for the right to sell or share some kinds of data."); Adam Schwartz, *The Payoff From California's "Data Dividend" Must Be Stronger Privacy Laws*, EFF (Feb. 15, 2019), https://www.eff.org/deeplinks/2019/02/payoff-californias-data-dividend-must-be-stronger-privacy-laws.(cautioning against moves towards "pay-for-privacy" in the CCPA).

31. *See* Benjamin Edelman, *Priced and Unpriced Online Markets*, 23 J. ECON. PERSP. 21, 34 (2009) (discussing the economic reality that zero pricing can be sustainable "when there are adequate profits in complementary businesses like advertising or technical support").

32. Omri Ben-Shahar, *Your Internet Privacy Should Be Up for Sale*, FORBES (Aug 8, 2016), https://www.forbes.com/sites/omribenshahar/2016/08/08/your-internet-privacy-should-be-up-for-sale/#2ec1f66d7ef2

33. *See id.*; Thomas M. Lenard, *'Pay-for-Privacy' Internet Actually Benefits Low-Income Consumers*, THE HILL (Aug. 16, 2016), https://thehill.com/blogs/pundits-blog/technology/291549-pay-for-privacy-internet-actually-benefits-low-income-consumers.

pay for services despite their articulated privacy concerns,[34] and given the overall consumer surplus from online applications, [35] commentators argue that this is exactly the type of market-driven solution to which policymakers should defer in protecting the privacy of concerned consumers. [36]

At the same time, the turn to paid models as an antidote to the information abuses by providers of free services has received significant criticism. Two sets of concerns—distributional and operational—have received significant attention.

The fundamental distributional concern with relying on price-based market models (especially in place of government regulation) involves the disproportionate barriers placed on exercising a fundamental right.[37] As Julie Cohen noted nearly twenty years ago, "[i]f data privacy costs money—or, conversely, if surrendering privacy saves money—access to privacy will be more unequal than if it did not."[38] Stacy-Ann Elvy's foundational work on the pay-for-privacy model, moreover, explores the ways that this divide is particularly pernicious, in that it can lead to the collection of more data about precisely those

34. Anthony Spadafora, *Americans reluctant to pay for privacy*, TECHRADAR (Jan. 16, 2019), https://www.techradar.com/news/americans-reluctant-to-pay-for-privacy (discussing a Center for Data Innovation study finding that only 27 percent of the surveyed would pay a monthly subscription fee in exchange for less data collection, despite the fact that 80 percent of respondents said they wanted online services to collect less data, and 63 percent of the surveyed opposed receiving free applications or services in exchange for more intensive data collection).

35. Michael R. Hammock & Paul H. Rubin, *Applications Want to be Free: Privacy Against Information*, COMP. POL'Y INT'L 1 (2011), https://techpolicyinstitute.org/wp-content/uploads/2011/03/applications-want-to-be-free-p-2007462.pdf ("The costs of online privacy-related harm (such as identity theft) and of protective activities are small relative to the benefits from applications that are supported by online advertising, which depends on the collection of personal information.").

36. *Id.* at 2 ("If consumers do have valid privacy concerns, markets can and do respond to them.")..

37. *See* Adam Schwartz, *The Payoff From California's "Data Dividend" Must Be Stronger Privacy Laws*, EFF(Feb. 15, 2019), https://www.eff.org/deeplinks/2019/02/payoff-californias-data-dividend-must-be-stronger-privacy-laws ("Pay-for-privacy schemes undermine this fundamental right. They discourage all people from exercising their right to privacy. They also lead to unequal classes of privacy 'haves' and 'have-nots,' depending upon the income of the user."); Elvy, *supra* note 11, at 1409 ("[U]se of this model is likely to contribute to the divide between those that can afford privacy and those that cannot.").

38. Julie E. Cohen, Examined Lives: Informational Privacy and the Subject as Object, 52 STAN. L. REV. 1373, 1398 (2000).

consumers who are particularly susceptible to predatory and discriminatory behavior.[39]

The operational concerns derive from longstanding research into the practical and cognitive barriers skewing the market for privacy. Because of the inscrutability of the behavior of companies that collect, process, and share data—and the inability to predict and understand how it will be used in the future—consumers have little access to a true understanding of the ways that their data will be used, and the personal implications (indeed, in many cases companies themselves do not know how they will use the data in the future).[40] Consumers, then, simply do not have the capacity to estimate the value of their own data, or the costs of handing it over.

These information asymmetries are compounded by the opacity of privacy policies,[41] and—especially under such conditions of uncertainty—the malleability of consumer choices in light of the framing of "consent" to data use and sharing, and of very small transaction costs, or frictions.[42] Moreover, even those who may wish to take more active measures to protect their privacy are often unable to succeed because of technical barriers, collective action problems, monitoring costs, third-party leakage, and data interconnectedness.[43] Thus, privacy—even for consumers who might feel strongly about protecting their data—might not actually be a salient feature of a free or paid choice, because it is frequently difficult for consumers to negotiate for privacy protec-

39. *See* Elvy, *supra* note 11 at 1421-28 (providing examples of the ways that companies use consumer lifestyle data and data analytics in making choices about services offered to consumers and of predictive data to discriminate against individuals deemed "less valuable" or "risky") (*quoting* Amy J. Schmitz, *Secret Consumer Scores and Segmentations: Separating "Haves" from "Have-Nots"*, 2014 MICH. ST. L.REV. 1411, 1414.

40. *See* Strandburg, *supra* note 4, at 143, 148, 150.

41. Kevin Litman-Navarro, *We Read 150 Privacy Policies. They Were an Incomprehensible Disaster*, N.Y.TIMES (June 12, 2019), *available at* https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html; *see also* Aleecia McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*. IN PROCS. TECH. POL'Y RES. CONF., Sept. 26–28 2008 (concluding that if the average Internet user reads every word of all privacy policies they come across, the user would spend 201 hours reading, worth roughly $3,534 annually)

42. Susan Athey, Christian Catalini, & Catherine Tucker, *The Digital Privacy Paradox: Small Money, Small Costs, Small Talk*, MIT SLOAN RES. PAPER No. w23488 (2017) at 12, *available at* https://www.ftc.gov/system/files/documents/public_comments/2017/09/00010-141392.pdf. (analyzing the choices of a group of MIT undergraduates on data use and sharing); *see also supra* text at nn. 84-88 (discussing framing effects);

43. *See* Strandburg, *supra* note 4, at 156-57, 164.

tions in the current market. This lack of transparency and choice is why analyses of "revealed preferences" are unlikely to accurately explain consumers' true preferences.

Generally, the lack of information and the malleability of consumer choices has provided some explanation for the "privacy paradox," by which consumers care about personal privacy, yet frequently exhibit privacy-compromising behaviors.[44] More specifically, it suggests the deep shortcomings of the pay-for-privacy model for offering consumers a meaningful option and an informed choice. In the absence of necessary information, consumers may rationally be unwilling to pay for more privacy because they cannot accurately value it, or assess whether or how it may be protected. Nonetheless, when consumers are given clear privacy indicators, the privacy paradox disappears: they *do* pay for increased privacy, thereby better aligning their behaviors with their stated preferences.[45]

Elvy's work, moreover, raises two additional questions about the paid services model, regarding which far less empirical research exists.[46] First, she points out, companies using a paid model may still not actually refrain from monetizing consumer data.[47] Understanding actual company behavior is particularly important in the context of mobile apps, for example, that offer ad-free paid versions, but whose data collection, processing, and sharing behaviors still remain hidden from the consumer.

Second, and in light of this possibility, Elvy notes the possibility that "[p]rivacy-conscious consumers who elect to pay for" services may be misled about data practices.[48] Research has found consumers' privacy choices to be highly susceptible to suggestions pre-disclosure: when treated with a very modest pri-

44. *See* Alessandro Acquisti. *Privacy in Electronic Commerce and the Economics of Immediate Gratification*, *in* PROCEEDINGS OF THE ACM ELECTRONIC COMMERCE CONFERENCE (EC '04), 21–29 (ACM Press 2004), *available at* http://www.heinz.cmu.edu/~acquisti/papers/privacy-gratification.pdf.

45. Janice Y. Tsai, Serge Egelman, Lorrie Faith Cranor, & Alessandro Acquisti. *The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study*, 22.2 INFO. SYS. RES. 263-64 (2011); Serge Egelman, Janice Tsai, Lorrie Faith Cranor, Alessandro Acquisti, *Timing Is Everything? The Effects of Timing and Placement of Online Privacy Indicators*, *in* PROCEEDINGS OF THE SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 324-325 (CHI '09) (ACM Press 2009).

46. *See* Elvy, *supra* note 11, at 1413–19.

47. See *id.* at 1419.

48. *Id.*

vacy-enhancing accommodation, subjects counter-intuitively increased disclosure incommensurately.[49] In light of the broader pay-for-privacy discourse, and the implicit trade-off between payment with money and payment with data, then, consumers might reasonably (and mistakenly) expect paid services to be more privacy-protective.

B. App Behavior And Consumer Expectations Studies

We present two studies that fill this empirical gap, and begin to answer these two questions—whether companies using a paid model actually refrain from monetizing consumer data, and whether consumers might expect paid services to be more privacy-protective—with data from the mobile app context. The first study employs a dynamic analysis tool to track the actual privacy behaviors of paid and free apps. This tool, currently employed by regulators and watchdog groups, provides crucial information largely unavailable to consumers making privacy decisions, and—as we discuss in Part III—can provide a broader means for empowering consumers in making more informed decisions about the use and dissemination of their data. Almost half of the paid apps that we examined in this study shared the same types of personal information with the same third parties as free ad-supported versions.

The second study consists of an online survey of mobile app users. In this survey, we presented 1,000 respondents with screenshots of two apps from the Google Play Store: a "free" app and its paid counterpart. We then asked them questions about which app they would be more likely to install, as well as what differences they would expect to exist between the two versions. Our results demonstrate that consumers are significantly more likely to expect strong privacy protections when purchasing apps, as compared to installing their free ad-supported counterparts.

1. Paid and Free App Behavior Study

In this analysis, we generalize different app monetization models into two overarching categories: we define "free apps" as those that are available for download on the app store at no up-front cost; and we define "paid apps" as apps that require a one-time payment to download. Our focus is on paid apps in which the consumer pays for the app as a single discrete product, not for a continuously renewed service. We acknowledge that apps may employ other monetization strategies, such as the "freemium" or "paidmium" models, in

49. Athey, et *al.*, *supra* note 42, at 17–18.

which potentially recurring in-app purchases generate revenue for the developer. Though we are aware that some apps do offer in-app purchases to disable ads, these are beyond the scope of this study.

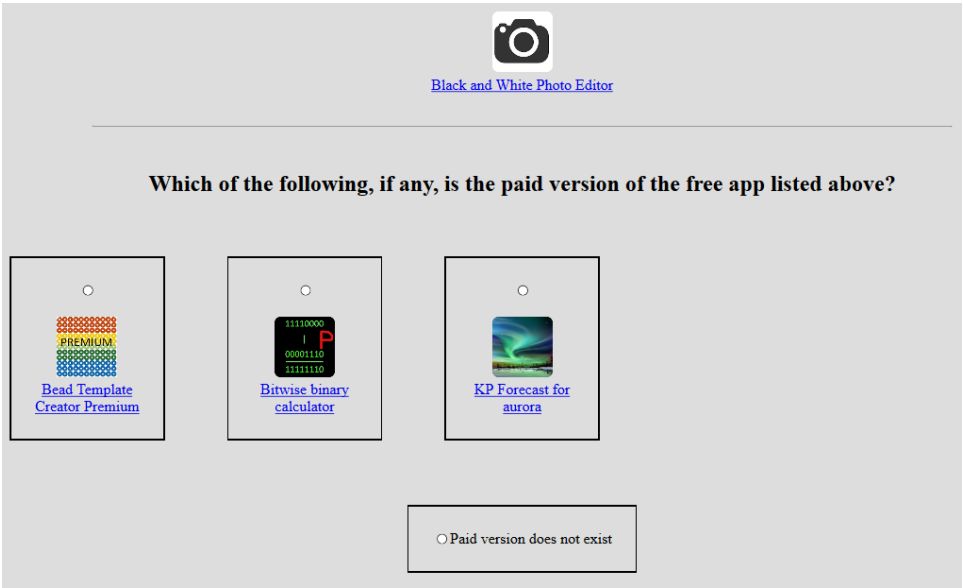
The Google Play Store does not reliably link free apps to their paid versions, or even indicate if a corresponding paid version exists at all. Therefore, we first developed our own method to identify pairs of free and paid versions of the same general app (e.g., "Quick PDF Scanner FREE" and "Quick PDF Scanner PRO"). We evaluated and compared the behavior of these pairs using both static and dynamic analysis techniques.[50]

We formed our app corpus by consulting the AppCensus database,[51] which is regularly updated by crawling the "Top Free" charts in each of the Play Store's categories. We then created a labeling task on Amazon Mechanical Turk. We presented workers with a free app and a list of all paid apps from the same developer, asking them to select the paid version that most closely resembled the given free app (Figure 1). In order to increase the likelihood of valid free and paid pairings, we only presented workers with free apps whose titles or package names contained the words "free" or "lite," as those keywords would suggest that a "paid" or "full" version exists. If the free app did not have a corresponding paid version, workers were instructed to select the "Paid version does not exist" option. We presented each free app to three different workers, then manually adjudicated the responses for agreement and correctness. We paid workers $0.10 for each match in consensus with the others, yielding a corpus of 5,877 pairs of apps.

[50] "Static analysis" refers to the examination of programs without executing them, in order to rapidly detect whether they contain certain instructions or data. Static analysis often yields false positives because some detected instructions may not ever get executed in practice. "Dynamic analysis" refers to the examination of program behavior by executing it to monitor what it does, which may lead to false negatives, if certain functionality is not executed during the testing period.

51. *See generally* APPSEARCH, https://search.appcensus.io/

Figure 1: Amazon Mechanical Turk task in which participants identified the paid counterpart of the given free app



We looked for similarities across pairs of free and paid apps along three dimensions: (1) the portion of Android permissions designated by the operating system as "dangerous"—signifying that they control access to sensitive data or personally identifiable information—declared by the free app also declared by the paid app; (2) the portion of third-party packages (i.e., SDKs) found in the free app that are also included in the paid version; and (3) the portion of sensitive network transmissions performed by the free app also performed by the paid app. We believe these three aspects are a good representation of apps' data collection and sharing behaviors. We employed the following methods to evaluate these:

a) Static Analysis

We used the Android Asset Packaging Tool to extract the permissions apps request for various device resources.[52] We then identified differences in dangerous permissions within pairs of free and paid apps. Additionally, we relied on Apktool to examine apps' file structures for the package names that

52. *See generally AAPT2*, ANDROID, https://developer.android.com/studio/command-line/aapt2.

comprise the app.[53] We identified third-party libraries by eliminating package names that shared the same first two levels as the app package (i.e., disregarding code belonging to the core app). This revealed what third-party libraries—possibly used for monetization and data collection—are shared between free apps and their paid counterparts.

b) Dynamic Analysis

We used dynamic analysis methods derived from earlier work to automatically evaluate apps by executing them in an instrumented environment (deployed on identical Nexus 5X smartphones) that captures apps' network traffic.[54] We relied on the Android SDK's Application Exerciser Monkey tool to automatically explore apps without user intervention.[55] Although there is no guarantee that paired apps have identical user interfaces, we controlled for differences in app execution by providing both apps with the same random input stream at the same time. This increases the likelihood that observed differences in app behavior arose from implementation differences, rather than differences in input.

At the end of each paired execution, we analyzed the captured network data to identify which sensitive data types were sent to which remote services—services that could be for advertising, profiling, crash reporting, etc. We focused on detecting the transmission of sensitive data that can be used to uniquely track a user over time and across different services: persistent identifiers, such as the Android Advertising ID (AAID), IMEI, and Wi-Fi MAC address; as well as personally identifiable information (PII), such as geolocation, name, and phone number. In order to detect the transmission of sensitive data, we not only used simple string matching, but also relied on methods from previous work for the decoding of obfuscated network traffic, which uses regular expressions formed from the manual inspection of different data encoding schemes.

53. *See generally A Tool for Reverse Engineering Android APK Files*, APKTOOL, https://ibotpeaches.github.io/Apktool/.

54. *See* Irwin Reyes, *et al*, *"Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale*, PROCS. ON PRIVACY ENHANCING TECHS 63 (2018).

55. *See generally UI/Appliation Exercise Monkey*, ANDROID, https://developer.android.com/studio/test/monkey.

Most current approaches to detecting suspicious application activity on mobile platforms rely on static analysis[56] or dynamic analysis.[57] However, previous approaches fall short because they either do not observe actual violations—instead only detecting when a program *might* contain violative code (in the case of static analysis), or do not scale (in the case of prior dynamic analysis approaches).

Our dynamic analysis framework allows us to monitor actual program behavior in real-time and at scale. The AppCensus platform allows us to examine how often and under what circumstances apps and third-party libraries access sensitive resources guarded by permissions. By combining this infrastructure with a modified version of Lumen (previously known as Haystack),[58] an advanced network monitoring tool, we obtain a sophisticated holistic view of when sensitive data is accessed and where it gets sent.

2. Consumer Expectations Survey

In addition to examining the differences in behaviors between each free app and its paid counterpart, we also wanted to understand consumer expectations surrounding the two versions of each app. Specifically, our research questions about people's expectations and beliefs about mobile app privacy when presented with a free app and its paid alternative were as follows:

- What differences do consumers expect when downloading an app for free versus purchasing it?
- Given these differences, which app would users be more likely to install?

56. *See, e.g.*, Clint Gibler, Jonathan Crussell, Jeremy Erickson, & Hao Chen, *AndroidLeaks: Automatically Detecting Potential Privacy Leaks in Android Applications on a Large Scale*, *in* PROC. OF TRUST (2012); Michael I. Gordon, Deokhwan Kim, Jeff Perkins, Limei Gilham, Nguyen Nguyen, & Martin Rinard, *Information-Flow Analysis of Android Applications in DroidSafe*, *in*PROC. OF NDSS SYMPOSIUM (2015); Jinyung Kim, Yongho Yoon, Kwangkeun Yi, & Junbum Shin, *ScanDal: Static Analyzer for Detecting Privacy Leaks in Android Applications*, IEEE MOST (2012); Sebastian Zimmeck, *et al*, *Automated Analysis of Privacy Requirements for Mobile Apps*, *in* PROC. OF NDSS SYMPOSIUM (2017).

57. *See, e.g.*, William Enck, Peter Gilbert, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung, Patrick McDaniel, & Anmol N. Sheth, *TaintDroid: An Information-flow Tracking System for Realtime Privacy Monitoring on Smartphones*, *in* PROC. OF USENIX OSDI (2010).

58. *See generally* THE HAYSTACK PORJECT, https://haystack.mobi (an app that "analyzes mobile traffic and helps to identify privacy leaks inflicted by apps and the organizations collecting this information").

- Do users expect different privacy behaviors from a free version and its paid counterpart?

We recruited 1,000 participants from the Prolific Academic survey platform,[59] limiting participation to those within the U.S. who successfully completed at least 95% of the previous tasks that they had undertaken. The survey took approximately five minutes to complete, for which we compensated participants $1.00 for their time. This study was reviewed and approved by the UC Berkeley Institutional Review Board.

We conducted our study during May 2019. We piloted our study with 100 participants, and then ran the main study with 1,000 participants. Our data is drawn only from the latter 1,000 responses. Based on the pilot, we did not make any changes to the survey. Our sample was gender-balanced, with 50% self-identifying as male; the median reported age was 30, with the reported ages ranging from 18 to 76. In addition, approximately 54% of our sample had at least a bachelor's degree, and 56.5% of our sample reported themselves as single.

Our survey was composed of several sections: a mix of open-ended responses, multiple-choice questions, and 5-point Likert scale questions[60] (listed below)—ranging from "Definitely A" to "Definitely B," concluding with a series of demographics questions (See sub-Part 2.b for detailed wording of individual questions).

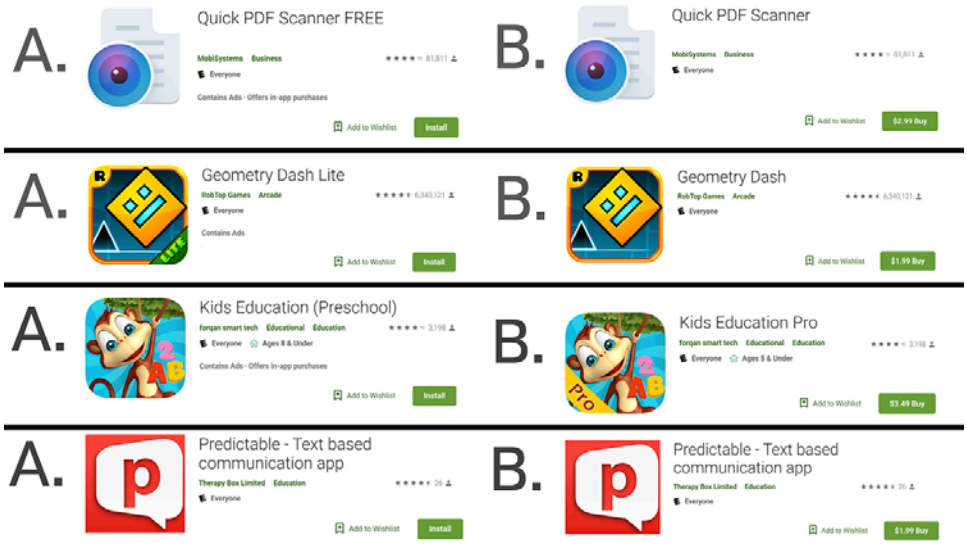
a) Open-Ended Questions

After obtaining participants' consent, we first presented respondents with two images of a free version of an app and its paid counterpart, controlled to have the same rating, where the key differences were in the installation price and the title of the app ("free" or "lite" for the free versions, "pro" or "premium" in the paid versions). We randomly selected a pair of apps from four possibilities (Figure 2).

Figure 2: Participants were randomly shown one of four pairs of apps, labeled A and B

59. PROLIFIC, https://prolific.ac.

60 In questionnaire research using Likert scales, respondents specify their level of agreement or disagreement on a symmetric agree-disagree scale for a series of statements, *https://www.britannica.com/topic/Likert-Scale.*



We randomized how the two apps were presented to participants across two metrics: (1) what app they were shown (selected from four possibilities shown in Figure 2): a PDF scanning app, a game app, a children's education app, and a text-based communication app, and (2) whether the free app was labeled as A or B. We later recoded this so that in our analysis App A always referred to the free version, while App B referred to the paid version.

After displaying the randomly-selected pair of apps, we asked participants, "In what way, if any, would you expect the above two apps to differ?" Responses to this question were collected using an open-ended text field. Two independent coders later coded these responses as binary values based on whether participants mentioned privacy or related concepts. Next, we asked participants to specify which app they would be more likely to install and why. As before, this question was coded by two independent coders based on whether concepts pertaining to privacy were mentioned.

b) Likert-scale questions

After participants answered these open-ended questions, they proceeded to the next page of the survey. The top of this page once again displayed the same pair of apps as the previous page, and asked participants to answer several Likert-scale questions using the following 5-point scale: "Definitely A (1)," "Likely A (2)," "Equally A and B (3)," "Likely B (4)," and "Definitely B (5)." The statements participants rated were as follows:

Consider the same two apps once again. Based on the images, which app do you believe is more likely to...

- share your data with third-party services?
- share your data with advertisers?
- share your data with law enforcement agencies?
- encrypt your data to protect it from potential breaches?
- be transparent with you about its data collection and sharing behaviors?
- comply with privacy laws and regulations?
- delete all your data from its servers after you uninstall the app?
- keep your data on their servers when no longer needed for the functionality of the app?
- have effective privacy controls (features that allow you to specify which data types you do not want the app to collect)?
- access more resources that it needs for its functionality (i.e., more permissions)?
- protect the data you gave it permission to access?

Following these Likert-scale questions, we included a few related questions for a separate study (which we do not discuss in this paper), and then concluded by collecting demographic information.

3. Limitations

While our findings (see Part III, below) show that on average, paid apps may provide fewer privacy protections than consumers expect, we note several limitations of our methodology. First, all apps were executed by randomly generating UI events (i.e., random taps, swipes, etc.), which means that certain app functionality may not have been executed during the testing period. Thus, it is possible that under more realistic testing circumstances, some of the apps might exfiltrate data to additional third parties. Similarly, if the user interfaces between free and paid apps differ substantially, different functionality may have been executed between the two during the testing period, confounding

our results. Related to this, another confounding factor is simply due to the stochastic nature of mobile advertising: the same app executed multiple times is likely to contact multiple ad networks, which in turn load ad content from different advertisers and attribution trackers. Thus, it is possible that some pairs of free and paid apps were not observed contacting all of the same third parties for this reason. This suggests that our empirical observations may be lower bounds for privacy-invasive behaviors.

Regarding the expectations survey (see Part III, below), we observed that many participants said they would have chosen the free app despite later indicating that they believed that it would have worse privacy practices. Given that we randomly assigned apps to participants so that they could answer questions about the same free and paid pair, it is likely that many participants would not have chosen either of these apps to install under normal circumstances. Similarly, because participants were not exposed to privacy risks (nor financial costs), our results only show relative stated intentions, rather than revealed preferences. Thus, we believe that further study is needed to better understand participants' decision making between free and paid apps (as distinct from their expectations about the apps' privacy behaviors) under more realistic circumstances.

III. FINDINGS

A. APP BEHAVIOR

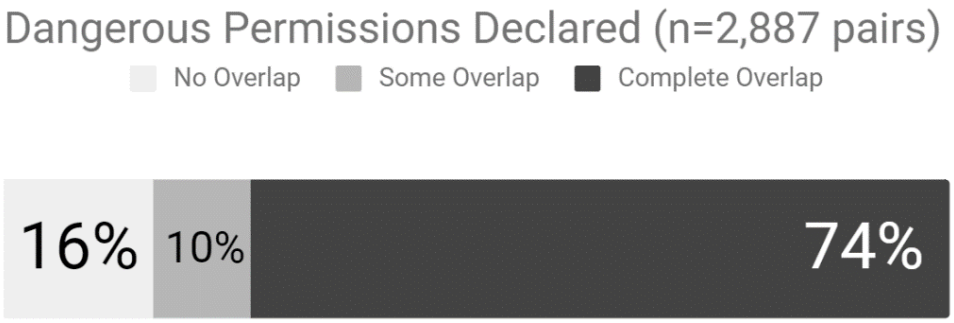
This work focuses on measurable differences in privacy between free and paid versions, so all presented comparisons are conditioned on the free app having at least one observation for any of the corresponding metrics. That is, in each of the following analyses, we disregard pairs in which the free app had no third-party packages, no permission requests, or no sensitive data shared with a third-party service, respectively. As a result, the total sample size in each analysis fluctuates slightly.

We note that there are indeed some paid apps that have observations along these dimensions that were not seen in their free counterparts. However, these represent only a small portion of our corpus: out of the 5,877 studied, 350 paid apps requested *dangerous* permissions not declared by their free versions, and 255 paid apps transmitted data not observed in the free release. We stress that our analysis quantifies the degree to which free apps' behaviors along these three metrics are carried over to their corresponding paid versions.

1. Declared Android Permissions

The Android permission system serves to protect user privacy. Apps must hold appropriate permissions to use various device resources (e.g., Internet access, the camera, etc.) and access sensitive user data (e.g., phone number, location data, various persistent identifiers, etc.). A subset of Android's permissions are deemed "dangerous" because they guard sensitive resources that directly affect user security and privacy, such as the contact list or location information.[61] All of the resources categorized as dangerous permissions require user consent at runtime (though upon approval, the user is never prompted again).

Figure 3: Frequency of dangerous Android permissions inherited between free and paid versions, where the free app requested at least one dangerous Android permission



Of the 5,877 pairs in our corpus, 2,877 had free versions that declared at least one Android-defined dangerous permission. In seventy-four percent of these pairs, the paid version (Figure 3) declared all of the same dangerous permissions held by the free version. That is, paid apps held all the same access to sensitive resources as free versions in a majority of the cases where any dangerous permissions were declared. Since third-party libraries (SDKs) share the same permissions as the main app code, any third-party libraries for tracking and/or user profiling present in the paid version could have access to the same user data as the free counterparts. The most common dangerous permissions that both the paid and free versions requested were those that use disk storage shared between apps, get information about the phone's state (e.g.,

61. *Protection Levels*, ANDROID,

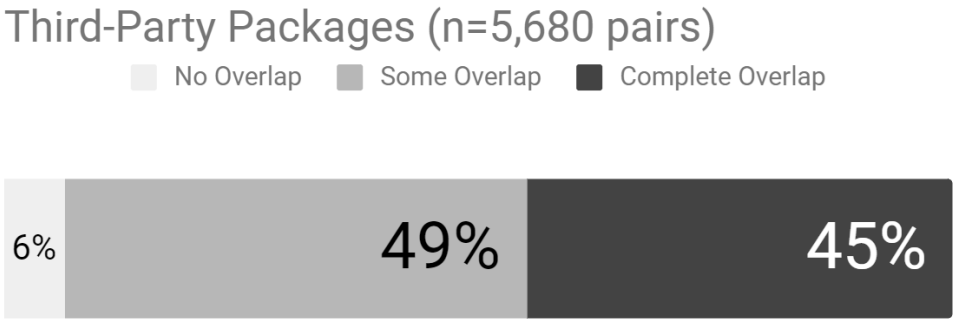
phone number, cellular network information, call status), and access to the device's geolocation.

We also note that sixteen percent of the pairs in our study had paid apps that did not request any of the dangerous permissions declared by their corresponding free versions. This suggests potential over-permissioning of free apps in these cases, in which free apps held access to dangerous permissions that may not have been necessary for those apps' core functionality. Overall, this implies that free apps will likely have access to permissions that the paid app does not, putting users' privacy at higher risk by requesting permissions that are unnecessary to the apps' core functionality.

2. Bundled Third-Party Packages

It is common practice in software engineering to use third-party code to expedite development. That is, developers do not need to "reinvent the wheel" and can instead integrate functionality into their programs written by others. In mobile apps, third-party libraries allow for pre-built functionality like graphics rendering, advertising, and analytics, among others. Third-party code bundled in apps has the same privileges as the host app, and can access all the same device resources and personal data available to the host app.

Figure 4: Frequency of third-party package reuse among free and paid pairs, where the free app had at least one third-party package



Of the 5,877 pairs in our corpus, 5,680 had at least one third-party package in the free version. Of these pairs as Figure 4 shows, forty-five percent of paid apps contained the same third-party libraries as the free versions, while six percent of paid apps showed no third-party libraries carried over from their free versions. The remaining forty-nine percent of paid apps had varying degrees of third-party library reuse from the free version to the paid version. This data suggests that paid apps are likely to contain most, if not all, of the same third-party libraries as the free versions. Although we acknowledge that our

analysis did not account for third-party libraries included but not actually executed (i.e., dead code), these results show that developers of paid apps have little motivation to remove externally-produced code in paid apps. This may leave paying consumers exposed to the same potential for third-party data collection as found in free apps.

Based upon the library categorizations of LibRadar,[62] we analyzed the types of third-party libraries present in free and paid versions of apps, focusing our attention on libraries labeled as "Advertising" and "Analytics"; some of the common libraries categorized as Advertising included advertising companies such as *Unity*, *AppLovin*, *Google's AdMob*, and *Chartboost*.

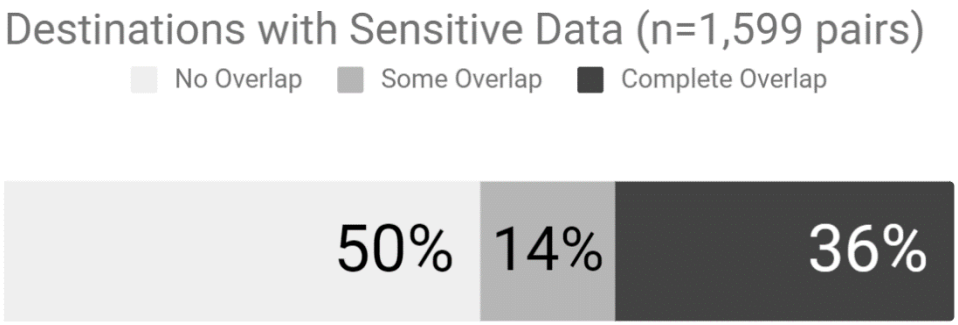
Focusing on advertising libraries specifically, LibRadar detected at least one ad library present in either the free or paid release (or both) in 3,043 pairs. Of these, 2,918 free apps contained ad libraries, while 1,320 paid apps contained ad libraries. Furthermore, 209 paid apps even bundled at least one advertising library that was not present in its free counterpart, suggesting that some paid apps will not only share some of the same advertising libraries included in the free version, but also introduce new ones. Thus, although ad libraries are certainly monetizing most free apps, paying for an app only reduces the likelihood of encountering ad software by half.

3. Network Transmissions

Third-party libraries bundled in apps routinely collect various data from users and their devices, sending it back to the app companies' servers. For example, crash reporting services often gather hardware specifications and usage telemetry to help developers debug their apps, while advertising networks collect persistent identifiers and personal information to better target users with relevant ads. By observing all of the network traffic associated with an app, we can discern the types of sensitive data being transmitted (e.g., Android Advertising ID, e-mail addresses, geolocation information, etc.) and the recipient of that data.

Figure 5: Frequency of unique domain destinations shared between free/paid pairs, where the free app transmitted sensitive data to at least one domain

62. Z. Ma, H. Wang, Y. Guo & X. Chen, *LibRadar: Fast and Accurate Detection of Third-party Libraries in Android Apps*, 2016 IEEE/ACM 38TH INT'L CONF. ON SOFTWARE ENGINEERING COMPANION (ICSE-C) 653 (describing the function and method of LibRadar).



Among the 5,877 pairs of apps that we examined, 1,599 pairs' free version transmitted sensitive data to online services over the Internet. Out of these 1,599 pairs, we observed that fifty percent of these pairs' paid versions (Figure 5) did not communicate with any of the domains that the free version did, while fourteen percent shared some destinations with the free version. Conversely, thirty-six percent of these pairs' paid versions communicated with all of the same domains as the free version. We found that overall, the most frequently-observed sensitive data types shared by both free and paid apps were indeed those that enable persistent tracking, such as Advertising ID (651 pairs), Android ID (570 pairs), device IMEI (65 pairs), and location (39 pairs).

B. CONSUMER EXPECTATIONS SURVEY DATA

We directly tested the null hypothesis that consumers are likely to believe that free and paid versions of the same app offer the same privacy and security protections. To test this hypothesis, we constructed a survey with a mix of open-ended questions, multiple-choice questions, and 5-point Likert scale questions, concluding with a series of demographics questions. We observed that when provided with an open-ended question about the differences between the free and paid versions of an app, few participants mentioned privacy unprompted. This suggests that privacy behaviors may not be their *primary* considerations. However, when explicitly asked to compare the privacy behaviors of the two app versions, we were able to reject the null hypothesis. This indicates that while privacy differences may not be among the participants' primary considerations, they are nonetheless an important secondary consideration for a significant proportion of study participants.

1. Open-Ended Questions

a) Expected Differences

To probe further into consumers' expectations of free and paid app behavior, we began with an open-ended question to avoid priming. After being presented with a free version of an app A, and its paid counterpart, app B, respondents were asked, "In what way, if any, would you expect the above two apps to differ?" Approximately half of the responses (49.4%) mentioned the inclusion or exclusion of ads between versions. One participant stated, "The free one will have ads and the paid one will not" (P77). Another wrote, "Option A will have ads, but I'm not paying for it (it will likely be very annoying). Option B will not, but I would have to pay $3 up front" (P857).

Many responses mentioned differences in app features (48.1%). "The first will not have all the features of the second," one participant stated (P152). Another suggested the possibility of upgrades: "The first one would be free to install with limited features, but could be upgraded by paying" (P146).

However, few participants—without being primed to privacy/security—mentioned security and privacy differences between the versions (1%): "I think that the B app would have less intrusive permission requests than the A app would. The B app would be more trustworthy than the A app" (P457). Others even suspected malicious intent as a difference between the app versions, stating, "I might worry that A has a higher risk of viruses upon download, but otherwise they seem the same" (P174). Similarly, another wrote, "App A will be ad-supported and unlocking the full features would require payments. I would also be more concerned about it having spyware / malware aspects" (P287). These responses suggest that while participants think about the presence of ads unprompted, many do not immediately jump to the privacy implications of those ads.

b) User Preference

We found that most users (81.6%) would be more likely to install the free version of an app over its paid counterpart. To avoid priming, we followed up with an open-ended question: **why**? Of the participants that would be more likely to install the free version, the most prevalent reasoning (58.5%) was simply that the app was free. One participant even noted their willingness to trade security for price, stating, "I hate spending money on apps, especially if the service they offer is simple. So even though there's a higher chance of a virus for app A, I would download it to save the money" (P174). Echoing this

sentiment, another participant wrote, "Because I'm broke, and while I suspect the free app will harvest data beyond what would be appropriate for its function, I don't expect any better of the paid app" (P167). Another common reason was that users would prefer a low-risk option to try an app before committing to purchasing the paid version based on perceived usefulness, quality, and desire to support the developer. Thus, it is likely that the responses to this hypothetical question were influenced by the fact that we were asking participants to choose between two versions of an app, when in reality they may have had little interest in installing either.

However, of the participants that were more likely to state that they would purchase and install the paid version of the app, the most common reason (30.3%) given in the responses was the removal of advertisements. Many users mentioned how advertisements adversely affected the overall user experience, some even describing the advertisements as "annoying" or "disruptive." Almost six percent of the participants who preferred to pay for the app were more likely to say they would purchase it because of perceived privacy and security risks in free apps. A participant wrote, "I would review it for permissions it wants and any customer complaints of spyware, but generally paid apps are safer and I would want the full set of features right away" (P296). Another even stated, "[the paid app] would be less susceptible to security breaches and data mining" (P457). In addition, the responses revealed another facet of consumer preference—advertising to children. Out of the 103 participants who were randomly assigned children's education apps as the focus of the survey, many (28.2%) of the participants expressed that they would purchase the paid version of the app because they would not want advertisements displayed to their children. One participant even made the distinction between their purchasing preferences based on if the advertisements were directed at children or not, stating, "[the app] looks like a child's app and I would want my child not to be bothered by ads. Even if it were for me, I might chose [sic] the ad free version" (P894).

2. Expectations About Privacy Behaviors

For a quantitative perspective on users' expectations on the privacy behaviors between free and paid versions of an app, we asked participants to evaluate

on a Likert scale the differences between the apps based on the provided statements.[63] Overall, while the price of an app did not have any observable effect on whether participants believed an app was likely to request more information than it actually needed to function ($p = 0.68$), we did find that participants were more likely to expect that the free version would share their data with advertisers ($p < 0.0001$) and law enforcement agencies ($p < 0.0001$). In addition to this, users were also more likely to expect the free version to keep their data on the app's servers when no longer needed for the functionality of the app ($p < 0.0001$).

Paralleling this theme, users were more likely to expect the paid version to encrypt their data, protecting it from potential breaches ($p < 0.0001$), and comply with privacy laws and regulations ($p < 0.0001$). Similarly, participants were also more likely to expect the paid version to both protect the data they gave the app permission to access ($p < 0.0001$) and delete all their data from its servers after they uninstall the app ($p < 0.0001$). Not only were they more likely to expect more privacy-preserving behaviors from the paid app in practice, but they also were more likely to expect a higher level of transparency from the paid version with regard to its data collection and sharing behaviors ($p < 0.0001$) and more granular control over the collection of their data ($p < 0.0001$). This suggests that the mere act of paying, regardless of how much, is associated with receiving better privacy.

IV. IMPLICATIONS FOR PRIVACY PROTECTION

Our research shows that while consumers expect that paid apps are likely to have better privacy and security practices than their free counterparts, those expectations may not comport with reality. Specifically, almost half of the paid apps that we examined shared the same types of personal information with the same third parties as the free ad-supported versions. Worse, there was no obvious way for a consumer to understand when paying for an app was likely to lead to better privacy and when it was not.

These findings have important implications for our understandings of the ways that users develop expectations about the privacy behaviors of digital services, and for law and policy intended to protect personal data. Specifically,

[63] See scale and provided statements in Section II(B)(2); all comparisons were made using the one-sample Wilcoxon signed rank test to evaluate the observed data against the null hypothesis.

they underscore existing research about the ways that consumers develop understandings (and misunderstandings) about data usage, and the ways in which regulatory regimes purportedly premised on consent can thwart user intentions and vitiate that notion.

More specifically, they point to a wide divergence between expectations and reality and provide important evidence about the pervasive and unanticipated collection and sharing of data. While some of that behavior might be captured in evolving legal regimes, notably the GDPR, much is not. This failure suggests the importance of expanding regulatory approaches that on one hand reflect the ways that inaccurate expectations might lead to user deception, while on the other hand recognize that consumer confusion can vitiate notions of consumer choice by eliminating privacy as a salient factor in consumer decision-making altogether.

Finally, our research suggests the important role that the tools developed in our research can play in empowering and protecting users by making app behavior more transparent. These tools can reduce the opacity of app behavior and information asymmetries that corrupt the market for privacy. This could better align consumers' expectations with the actual behaviors of the apps they use, increasing the possibility that consumers might have the capacity to make meaningful choices about the digital services they use and the information they share. It also might clarify for policymakers and citizens ways that consumer choice fails as a means to govern information use meaningfully, and the contexts in which repairing market failures and addressing market abuses will require regulation.

A. INSIGHTS FOR LAW AND POLICY

In this light, our findings point to a number of important implications for law and policy.

1. Meaningful Consent, and the GDPR

As an initial matter, the ways that paid apps frequently collect and share personal data likely run afoul of privacy initiatives focused on ensuring that consumers be sufficiently informed about information practices such that consent to use their data is meaningful—notably the GDPR.[64] The finding that

[64] *See also* Proposed Regulations under the CCPA, § 999.305(a)(3), *available at* https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/ccpa-proposed-regs.pdf. The regulation states:

paid apps frequently share data in similar ways as their free counterparts indicates that, where free apps would violate the European law—insofar as they have no legal basis for the processing of personal data—many paid apps would as well. Commentators have predicted that the GDPR's presumption against consent as a legal basis for data processing when the data use in question is not "necessary" for the provision of a service will prohibit many of the information practices engaged in by free services, ending the "Internet's Grand Bargain" by which data is traded for services, and leading to a widespread replacement of free offerings with those that will require payment. Yet paid apps that engage in the behavior we documented would also violate such requirements.

2. Consumer Expectations and Privacy Protection

The divergence our research demonstrates between user expectations and the privacy and security of the apps they use further underscores the shortcomings of legal regimes that purport to rely on consent as the basis for privacy protection, yet fail to contend with the barriers to accurate consumer understandings of the ways technology implicates personal privacy. In the case of the apps we studied, free and paid, users simply had no capacity to pierce the opacity of app behavior, and the apps themselves provided few clues to understand the relative levels of privacy protection necessary for informed consent.

Our findings resonate with broader research into the ways that users develop their (frequently inaccurate or incomplete)[65] understandings of technology, and their interactions with it. The data behaviors of digital services providers are largely black boxes, preventing users from understanding the details of the ways that their information is collected, used, and shared. Privacy policies, moreover, are often deceptive and misleading.[66] Even when they do reflect accurate data practices, they are lengthy, often inscrutable and—despite

---

If the business intends to use a consumer's personal information for a purpose that was not previously disclosed to the consumer in the notice at collection, the business shall directly notify the consumer of this new use and obtain explicit consent from the consumer to use it for this new purpose. *Id.*

[65] Lindsey Barrett, *Model(ing) Privacy: Empirical Approaches to Privacy Law and Governance*, 35 SANTA CLARA HIGH TECH. L.J. 1, 17-20 (2018) (summarizing studies that find, *inter alia*, that most consumers don't understand basic facts including what privacy policies are, that applications continue to run in the background when the user is not directly engaged with them, and that an app can still access their information when not in use).

66. *See* Ehimare Okoyomon et al., *On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies*, 2019 WORKSHOP ON TECH. & CONSUMER PROTECTION, *available at* https://www.ieee-security.org/TC/SPW2019/ConPro/papers/okoyomon-conpro19.pdf

their length—often fail to disclose behaviors with sufficient granularity as to provide users with material information.[67] Even disclosures required by legal regimes seeking to mandate notice with sufficient personalization and specificity to warn consumers about problematic information practices have been plagued with problems of ambiguity that result in confusion about whether the recipient of the communication was at risk and should take action.[68]

Especially in such uncertain contexts, consumers, constrained by the human limits on attention and cognition, are "boundedly rational" decisionmakers.[69] Even if they had the capacity, they simply do not have the incentive to invest the time required to discern and evaluate all the terms of an agreement,[70] nor would it be rational for them to do so.[71] This is especially true—and creates particular opportunities for consumer exploitation—when issues are "nonsalient" to consumer decisions to engage in a transaction. Data use and privacy usually are "nonsalient," in that consumers face a "lack of meaningful choice" about them, or they are "hidden," or "unduly complex."[72] Put differently, they

(discussing the gaps between disclosed data collection practices as articulated in privacy policies, and *de-facto* data collection practices as observed using dynamic analysis tools).

[67] Barrett, *supra* note 65, at 17-18 ("Few people read privacy policies, and those who do are left with little basis to understand the uses of their data.") (summarizing research).

68. Yixin Zou, *et al.*, You '*Might' Be Affected*, PROCS. 2019 CHI CONF. ON HUM. FACTORS IN COMPUTING SYSS. 11 (2019) (presenting a study concluding that 97% of sampled data breach notifications were difficult or fairly difficult to read based on readability metrics).

69. Russell Korobkin, *Bounded Rationality, Standard Form Contracts, and Unconscionability*, 70 U. CHI. L. REV. 1203, 1204–06 (2003). In fact, it is because of the bounded rationality of consumers, the "market . . . will often include terms that are socially inefficient, leav[ing] buyers as a class worse off." *Id.* at 1206.

70. Aaron Perzanowski & Chris Jay Hoofnagle, *What We Buy When We Buy Now*, 165 U. PA. L. REV. 315, 323–27 (2017) (demonstrating that consumers lack an understanding of what it is they are buying when purchasing online digital media); Obar & Oeldorf-Hirsch, *supra* note 6, at 1 (finding that participants who joined a fictitious social network spent 51 seconds on average reading the ToS, with a 93% acceptance rate, and 98% of participants missed the intentional "gotcha clauses" like the assignment of their first-born child).

71. *See, e.g.*, Obar & Oeldorf-Hirsch, *supra* note 6at 1, 7 (finding that 98% of those agreeing to conditions of using a fictitious social network platform missed the intentional "gotcha clauses" the researchers implemented in the terms specifically mentioning users' data will be shared for the purpose of assessing eligibility for "employment, financial service (bank loans, insurance, etc.), university entrance, international travel, the criminal justice system, etc." and that users' first-born child will be assigned to the platform provided as payment for accessing the network).

72. *See* Amit Elazari Bar On, *Unconscionability 2.0 and the IP Boilerplate: A Revised Doctrine of Unconscionability for The Information Age*, BERKELEY TECH. L. J. (forthcoming 2019) (discussing standards for salience).

are not policed by the market since they do not impact the decision-making of consumers, due to their lack of ability to evaluate them.

Without access to reliable knowledge about app behavior, most users we surveyed understood payment for an app to suggest improved data privacy and security practices. This finding underscores research in the field of computer-human interaction that has begun to identify ways in which, in the absence of easily-accessible understandings, consumer expectations about technology instead result from "folk theories:" intuitive causal explanations that people construct to explain the world. Drawing on whatever clues are available from the framing of the technology, the discourse around it, and their interface with it, users develop "non-authoritative conceptions of the world" that can diverge significantly from the designers' views regarding—and the reality about—what a technology system is, and how it works.[73] These understandings, in turn, can result in "inaccurate understandings" of how technology systems work, "mismatches between designer and user intent," and "expectation violation."[74] Users rely on such folk models, moreover, to justify ignoring expert advice, reinforcing behaviors that increase data vulnerability and exploitation.[75]

Relatedly, it has been demonstrated that consumers' expectations regarding digital services' use of their personal data, and consequent privacy choices, are shaped by their perceptions about the trustworthiness of those services,[76]

---

73. *See* Motahhare Eslami, et al., *First I "Like" It, Then I Hide It: Folk Theories of Social Feeds*, Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems 2372–73 (2016) (discussing the development of folk theories about how the algorithms driving social media feeds operate, arising from "seams" in the system that are visible to users); *see also* Benjamin Toff & Rasmus Kleis Nielsen, *"I Just Google It": Folk Theories of Distributed Discovery*, 68 J. Comm. 636 (2018) (identifying folk theories about the way news reaches them through digital platforms, and the way that shapes engagement with public affairs) and Motahhare Eslami et al., *User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms*, *in* 2019 CHI CONF. ON HUM. FACTORS IN COMPUTING SYS. (2019) (surveying various users' attitudes and "folk theories" concerning the operation of algorithms on the Yelp platform).

74. Michael A. DeVito, et al., "*Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media*, Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems 3163 (2017).

75. *See* Rick Wash, *Folk Models of Home Computer Security*, PROCS. SYMP. ON USABLE SECURITY & PRIVACY (2010) (identifying eight 'folk models' of security threats used by home computer users, and how these models are used to justify ignoring expert security advice).

76. *See*, *e.g.*, Valentina Bali, *Tinkering Toward a National Identification System: An Experiment on Policy Attitudes*, 37 POL. STUD. J. 233, 250 (2009) (highlighting the role of trust in government institutions when it came to determining respondents' concerns over personal identification); Kirsten Martin, *Transaction Costs, Privacy, and Trust: The Laudable Goals and Ultimate Failure of*

foregrounding the question of ways that trust is generated in disclosure settings, and the relevance to these impressions of a "paid" versus "free" distinction.[77] In the context of the choice between free and paid digital services in particular, one recent study identified trust that the features would deliver privacy benefits, such as reduced data exploitation, as a significant influence on a customer's attitude toward paying for the premium version.[78]

The divergence we demonstrate between app behavior and user understandings, then, further points to the importance of regulatory efforts targeted at vindicating consumer expectations and preventing their abuse. Daniel Solove and Woodrow Hartzog have documented ways that the Federal Trade Commission's enforcement activity has laid the foundations for a robust privacy regulatory regime prohibiting broken consumer expectations of privacy,[79] pursuant to their authority to prevent deceptive acts under Section 5 of the agency's enabling act.[80] While the Commission's enforcement actions have generally focused on "broken promises," in the form of violations of privacy policies and explicit representations, Solove and Hartzog have urged an expansion of the agency's strategy to prohibit company behaviors that have deceptive *effects*, taking into account expectations reflecting the broader context. Our findings document the existence of these deceptive effects based on the gap between consumers expectations and reality, highlighting a need to address

---

Notice and Choice to Respect Privacy Online, 18 FIRST MONDAY (2013) (discussing methods to develop trust as an alternative to notice in protecting privacy).

77. *See* Patricia A. Norberg, Daniel R. Horne & David A. Horne, *The Privacy Paradox: Personal Information Disclosure Intentions Versus Behaviors,* 41 J. CONSUMER AFF. 100, 119 (2007) ("Future research, on a practical level, must examine how factors such as the physical environment, the media of data collection and responses to human interaction impact our assessment of trust and risk.").

78. Michel Schreiner & Thomas Hess, *Why Are Consumers Willing to Pay for Privacy? An Application of the Privacy-freemium Model to Media Companies*, 164 EUR. CONF. ON INFO. SYS. COMPLETED RES. PAPERS 5, 5-6, 12-13 (2015) (surveying German Facebook users regarding their willingness-to-pay for a premium version of Facebook with increased privacy).

79. Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 667 (2016) (identifying actions by the agency that could form a possible basis for a move in enforcement towards preventing "broken expectations of consumer privacy"); *see also* CHRIS JAY HOOFNAGLE, FEDERAL TRADE COMMISSION PRIVACY LAW AND POLICY 123-25 (2016) (discussing FTC use of surveys on consumer understandings, so as to better understand how consumers perceive statements or representations made to them by businesses); BAMBERGER & MULLIGAN, , *supra* note 24, at 183-196 (detailing ways that the FTC has sought, through a variety of "soft" and "hard" regulatory approaches, to link legal standards to consumer expectations).

80. 15 U.S.C. § 45(a)(1) (declaring unlawful "unfair or deceptive acts or practices in or affecting commerce").

this issue, whether through the FTC's "common law" privacy jurisprudence, or through broader privacy legislation.

3. Risk Salience, Transactional Salience, and Substantive Regulation

Finally, the considerable consumer confusion in the face of the complete app behavior opacity reflected in our findings might suggest taking bolder regulatory steps. Reliance on expectations as a legal backstop against privacy-intrusive behaviors is problematic enough in the face of increasingly widespread technical capacity on the one hand,[81] and market constraints on the options offered to consumers on the other.[82] Looking to unfounded expectations to set the appropriate boundaries of privacy protection is a sham. Alternatives could take the form of actively policing elements of user-app transactions in both the paid and free context that may be deceptive or exploitative, deeming them unconscionable.[83] Relatedly, they could manifest in the conclusion that reliance on consumer choice is ineffective against certain information collection, use, and sharing practices, which must be curbed instead by direct substantive prohibition.

An extensive body of empirical literature has identified privacy's "salience"—its prominence in a person's awareness at the time they are faced with a privacy decision—as an important element in shaping whether or not that individual makes more or less privacy-protective choices.[84] Accordingly, users' privacy preferences are not stable and coherent, but rather highly dependent on context.[85] When users are primed regarding privacy concerns, they are less

81. *See generally* David Gray & Danielle Citron, *The Right to Quantitative Privacy*, 98 MINN. L. REV. 62, 127 (2013) (calling such an approach to the definition of "reasonable expectation" of privacy in the Fourth Amendment "technological determinism run amok").

82. *See* Kenneth A. Bamberger & Orly Lobel, *Platform Market Power*, 32 BERKELEY TECH. L. J. 1051, 1089 (2017) (discussing the ways that tech platform market power can restrict users' privacy choices).

83. *See* Bar On, *Unconscionability 2.0*, *supra* note 72(urging the application of unconscionability doctrine to boilerplate terms in technology transactions in light of the newly revised RESTATEMENT OF THE LAW CONSUMER CONTRACTS, COUNCIL TENTATIVE DRAFT (AM. LAW INST. 2019), which addresses the one-sidedness of a term that unreasonably undermines "the consumer's benefit from the bargain").

84 Meredydd Williams, Jason R. C. Nurse & Sadie Creese, *Privacy Salience: Taxonomies and Research Opportunities*, IFIP INT'L SUMMER SCH. ON PRIVACY & IDENTITY MGMT. 263, 263–278 (summarizing research and defining privacy salience "as whether an individual is currently considering the topic of informational privacy").

85 Leslie K. John, Alessandro Acquisti & George Loewenstein, *Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information*, 37 J. CONST. RES. 858, 858-59 (2011); *see also*

likely to disclose data.[86] Moreover, timing matters. In-app dialogs increase salience more than those shown before an app's installation;[87] and even a 15-second delay between data use disclosures and the relevant decision can generate measurable differences in privacy-protective behavior.[88]

Along with informational asymmetries between users and tech companies, cognitive limitations on individual ability to process privacy policies and fully understand data use, and other decisional biases that discount risk,[89] the phenomenon of risk salience, and the resulting manipulability of consumer decisions it allows, has provided an explanation for the "privacy paradox" by which consumers behave in ways that undermine their stated privacy commitments and concerns.[90]

The findings of our expectations survey comport with these insights. Before being primed to the issue of privacy, only one percent of our respondents mentioned privacy or security as elements in which they would expect the paid and free versions of an app to differ. Yet when asked directly, over half indicated their belief that there would be a difference.

These findings suggest the importance of purposive policy efforts to build privacy "nudges" into the design of user interfaces and default configurations to assist users in overcoming hurdles to meaningful privacy choice,[91] especially

---

HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE (2010) (discussing the role of context in privacy attitudes and behaviors).

[86] John, *supra* note 85 (presenting four studies); Hazim Almuhimedi, et al., *Your Location Has Been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging*, 6 PROCS. 33RD ANN. ACM CONF. ON HUM. FACTORS IN COMPUTING SYSS. 787 (2015) (providing real-time information about lax app data sharing practices prompted over half of studied users to change permissions).

[87] Rebecca Balebako, et al., *The Impact of Timing on the Salience of Smartphone App Privacy Notices*, PROCS. 5TH ANN. ACM CCS WORKSHOP ON SECURITY & PRIVACY IN SMARTPHONES & MOBILE DEVICES 63 (2015).

[88] Idris Adjerid, et al., *Sleights of Privacy: Framing, Disclosures, and the Limits of Transparency*, PROCS. NINTH SYMP. ON USABLE PRIVACY & SECURITY 2 (2013).

[89] Alessandro Acquisti, & Jens Grossklags, *What Can Behavioral Economics Teach Us About Privacy? in* DIGITAL PRIVACY: THEORY, TECHNOLOGIES AND PRACTICES (Alessandro Acquisti, et al., eds.) (2007) (discussing the roles of information asymmetry and bounded cognition in explaining the privacy paradox).

[90] Meredydd Williams, Jason R. C. Nurse & Sadie Creese, *The Perfect Storm: The Privacy Paradox and the Internet-of-Things*, 2016 11th Int'l Conf. on Availability, Reliability & Security 2, 2-4 (2016) (discussing the roles of risk salience, user interface design and default configurations in explaining the privacy paradox).

[91] Alessandro Acquisti, et al., *Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online*, 50 ACM COMPUTING SURVEYS 1 (2017) (discussing research regarding design choices to overcome decision-making hurdles affecting individuals' choices in the presence of

in the mobile context in which multiple parties are often involved in data collection, and the small screen size presents display challenges.

The combination of salience effects with other causes of the "privacy paradox," moreover, suggests that another form of "non-salience" might also be at work—the non-salience of privacy as an element to the prospective bargain entered into by the vast majority of users surveyed—whether they choose paid or free services. Indeed, among our group, before being primed to think about privacy and security issues, only six percent cited them as motivators for their version choice.

Given the hurdles to accurate user comprehension about data practices, the opacity of actual app behaviors, and the way users shape expectations based on folk theories uninformed by necessary information, moreover, the non-salience of privacy (and to what practices they were "consenting") is hardly surprising. Armed with only unsubstantiated and largely inaccurate intuitions about the behavior of apps that provide neither transparency nor clues to their treatment of data, users had no way of reliably factoring privacy into the choice before them.

These findings resonate with the development of the notion of "salience" in the context of standard-form contracts (often termed "contracts of adhesion") more broadly. That context offers tools for preventing opaque app behavior that in practice exploits consumers, without relying on fictive consumer choice. Such contracts, include shrink-wrap licenses, software End User License Agreements (EULAs), and digital-platform Terms of Service, arise in contexts in which consumers face similar challenges to understanding inscrutable disclosures and possess no incentive to invest the time required to understand and evaluate terms—including those regarding privacy and data use—and no ability to negotiate them.

Taking into account the actual hurdles faced by consumers in these contexts, scholars and policymakers informed by behavioral understandings have argued that because non-salient terms—those that, in the words of the draft Restatement of The Law of Consumer Contracts currently under consideration by the American Law Institute, do not "affect consumers' contracting

---

privacy and information security tradeoffs); FTC, *Protecting Consumer Privacy in an Era of Rapid Change, Recommendations for Businesses and Policymakers,* https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf, 59–60 (2012) (recommending just-in-time disclosures and the obtaining of affirmative express consent before allowing apps to access sensitive content through APIs, such as geolocation information).

decisions,"[92] and are therefore not policed by market negotiation. They must thus be viewed with suspicion and either policed ex post by courts, or ex ante by legislation.[93]

In that vein, a recent complaint brought by the Los Angeles County Attorney under the California State Unfair Competition Law against a popular mobile weather app urged a court to disregard the company's privacy policy in sanctioning its behavior, because the app provided in-app disclosures that inaccurately presented users with contradictory privacy information, and these in-app disclosures were much more likely to be read by users. Users, the government argued, "have no reason to seek [geolocation data collection] information by combing through the app's lengthy [privacy policy], buried within which are opaque discussions of [the developer's] potential transmission of geolocation data to third parties and use for additional commercial purposes."[94] Indeed, the complaint recognized, "the vast majority of users do not read those sections at all",[95] effectively invoking the principle of "salience."

Where, as in the case of our findings, expectations diverge widely from reality, and mask pervasive and unanticipated collection and sharing of data, regulatory approaches must better reflect the ways that inaccurate expectations might lead to user deception. Regulatory approaches must also reflect the ways that structural and cognitive barriers can vitiate notions of consumer choice by eliminating privacy as a salient factor in consumer decision making altogether. In those contexts, notice and consent is merely a façade held up by an untethered fantasy of a rational, informed, and empowered consumer.

---

[92] Restatement of The Law Consumer Contracts, Council Draft No. 5, at 95 (Am. Law Inst. 2018). The draft Restatement explains, in discussing the standards for consumer consent, that the "concept of salience underlies the metrics regularly used to determine whether a contract term is unconscionable." Thus consumers' consent is vitiated when they face a "lack of meaningful choice" (if a term was non-salient because it did not "affect consumers' contracting decisions") or when a term constitutes an "unfair surprise," was "hidden", was "unduly complex," or resulted from "uneven bargaining power"—and these tests "are either synonymous with, or direct results of, nonsalience," *id. See* Bar On, *Unconscionability 2.0*, *supra* note **Error! Bookmark not defined.** (discussing the role of salience in the unconscionability doctrines).

[93] See Russell Korobkin, Bounded Rationality, Standard Form Contracts, and Unconscionability, 70 U. CHI. L. REV. 1203, 1204–06 (2003).

94. Complaint at 3, California v. TWC Product & Technology (1/3/2019), https://int.nyt.com/data/documenthelper/554-l-a-weather-app-location/8980fd9af72915412e31/optimized/full.pdf (seeking relief under California state Unfair Competition Law (Cal. Business and Professions Code, ss. 17200 *et seq.*)).

[95] *Id.*

B. OPPORTUNITIES FOR CONSUMER EMPOWERMENT AND ENHANCED OVERSIGHT

Finally, the effectiveness of our tool in piercing the opacity of app data collection practices suggests the promise of technical mechanisms that can foster transparency, increase salience, and empower users' decision making at scale by unmasking complex processes for empowering consumers and policing privacy-preserving design. While research often promotes auditing within the domain of technical experts,[96] by fostering dynamic analysis tools with accessible user-interfaces, such as the AppCensus tool, auditing could be scaled, and even crowd-sourced, to allow the average user to uncover data abuse practices.[97] Such tools could also enable the creation of a market for third-party assurance privacy seal and certification programs to set standards and enable companies to demonstrate privacy accountability and compliance.[98] Using dynamic analysis tools, these programs could centralize the testing of apps and label them based on their observed behaviors, relieving consumers of the individualized information-gathering burden and facilitating market enforcement.

Used in this fashion, such tools would further expose abusive terms and behaviors to the attention of consumer advocacy groups—truly increasing their salience.[99] One can even envision how such tools can be used to train machine-learning algorithms to highlight and spot behaviors in the wild, and

96. Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3.1 BIG DATA & SOC. 1, 3-4 (2016) (categorizing such mechanisms of opacity into various categories and suggesting that some opacity steams from "technical illiteracy", due to the specialized technical skill set needed to evaluate algorithms.

97. *See* Eslami et al., *User Attitudes*, *supra* note 73, at 12-13 (proposing user-auditing enhancing tools and crowd-sourcing algorithmic auditing in the context of potentially abusive machine-learning processes).

98. *See* Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247, 263 (2011) (discussing the rise of online privacy seal programs).

99. *Cf.* MARGARET JANE RADIN, BOILERPLATE: THE FINE PRINT, VANISHING RIGHTS, AND THE RULE OF LAW 16, 243 (2012) (noting that "NGO can organize publicity campaigns to make known to the public what some of the onerous terms in the fine print actually mean. The can take the lead in organizing a rating site that will advise consumers which firms are using reasonable terms and which are not"). *See also Ranking Digital Rights*, NGO https://rankingdigitalrights.org/ (rating leading internet companies human rights accountability posture (on a variety of topics from free expression to privacy) based on their ToS and Privacy Policies, *inter alia*).

flag them for review by consumers, regulators, and lawyers—using code to spot abusive code.[100]

With growing attention to privacy concerns, moreover, regulators, developers, and platform providers, such as the Google Play Store, need better tools to monitor app behavior and hold app developers accountable. Dynamic analysis tools and privacy-enhancing technologies and innovations geared to support greater transparency into data collection practices are a key component to the privacy landscape. Similarly, the tools described in this paper could benefit regulators in investigating the market for noncompliance by making it easier for them to detect violations and bring enforcement actions. If these enforcement actions are brought publicly, it may motivate other app developers to pay more attention to the privacy behaviors of their apps.

V. CONCLUSION

Our findings about consumer expectations and app privacy behavior strengthen the case for combining laws grounding consumer protection in behavioral realities with privacy-enhancing technologies that increase accountability. Privacy has seen its share of democratic degradation, where decades-long research has demonstrated the inability of consumers to comprehend lengthy privacy policies or notices and the ways that this failure largely obviates market competition over the quality of privacy-related contractual clauses.[101]

The wide divergence between expectations and reality counsels the expansion of regulatory approaches that on one hand reflect the ways that inaccurate expectations might lead to user deception, and on the other recognize that consumer confusion can vitiate notions of consumer choice by eliminating privacy as a salient factor in consumer decision-making altogether. Equipping users, third parties, and regulators with analytic tools once reserved only for

100. *See* Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez & Serge Egelman, "*Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale*, 2018(3) PROCS. ON PRIVACY ENHANCING TECH. 63 (explaining how the AppCensus tool is allowing users to search a name of a mobile app and learn about its actual information collection practice thereby empowering users).

101. RADIN, BOILERPLATE, *supra* note 99, at 213 (explaining how boilerplate are acts of "democratic degradation;" they employ mass systems of contracts to restructure and supersede the rights given by legislators, taking away rights granted by the democratic process). In the context of privacy see Alessandro Acquisti & Jens Grossklags, *Privacy and Rationality in Individual Decision Making*, 3.1 IEEE SECURITY & PRIVACY 26 (2005) (providing survey evidence as to how the bounded rationality of users affects their privacy decision-making processes and attitudes).

experts and academic researchers, moreover, would go far to address this phenomenon. Dynamic analysis tools offer the means to audit and contest explanations and notices provided by private parties and uncover actual behavior in an accessible manner—crucial to efforts to introduce more transparency and explainability in the context of machine-learning processing and information collection, and bringing information to bear to shape market practices.