# WHAT IS IT ABOUT LOCATION?

*Kirsten Martin*† *& Helen Nissenbaum*††

## TABLE OF CONTENTS



---

DOI: https://doi.org/10.15779/Z382F7JR6F


† William P. and Hazel B. White Professor, University of Notre Dame's Mendoza College of Business.

†† Professor of Information Science, Cornell Tech. The authors would like to thank the participants of the 2017 Privacy Law Scholars Conference for their helpful comments on an early study in this series. We are supremely grateful to colleagues who read earlier drafts and offered invaluable suggestions: Paul Ohm, Joel Reidenberg, Frederik Zuiderveen Borgesius, Deborah Estrin, Mainack Mondal, and Eran Toch provided critical insights, particularly into the technical correlates, in turn spurring ideas on the empirical and policy issues. We are grateful for support from the National Science Foundation under Grants No. 1311823, No. 1649415, CNS-1801501, and NSA H98230-18-D-006. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or NSA.

## I. INTRODUCTION

This Article reports on a set of empirical studies that reveal how people think about location data, how these conceptions relate to expectations of privacy, and consequently, what this might mean for law, regulation, and technological design. Despite the great debates, published commentary, court action, regulatory activity, and scholarly literature, not enough is known about how people understand location data, and what specifically about it affects people's judgments about others' access to their whereabouts.[1] Further, despite

1. *See* Matthew B. Kugler & Lior Jacob Strahilevitz, *Actual Expectations of Privacy, Fourth Amendment Doctrine, and the Mosaic Theory*, 2015 SUP. CT. REV. 205, 261–63 (2015) (calling for more empirical research on people's perception of location data and the psychological basis of privacy expectations).

efforts to stem location tracking, it remains rampant. Stern rules[2] aimed at curtailing location tracking are a poor match for the ingenuity of seekers of this information who, among other tactics, exploit enormous ambiguity in how location is interpreted and operationalized to make end runs around these rules.[3]

Filling this gap is critical to a frontier of privacy regulation that has been sorely neglected. This neglect exists in part because the significance of location data was not fully appreciated until the recent ubiquity of technology-enabled location tracking, and in part because its murkiness has suited the beneficiaries of location surveillance. Although our findings alone do not support specific lines of legal regulation, they leave little doubt of a damaging rift between how these beneficiaries of location surveillance communicate their practices and how we, its subjects, understand these practices. Only when this rift is repaired will it be possible to adequately regulate location surveillance—through policy, law, and technology—to meet privacy expectations and promote privacy's societal value.

### A. BACKGROUND AND MOTIVATION

The set of empirical studies on which this Article reports is the third in a series, initiated in 2015, which challenges the role of the public-private dichotomy in privacy law and regulation by scrutinizing the extent to which

2. *See, e.g.*, *Privacy, Security, and Deception*, GOOGLE PLAY DEVELOPER POL'Y CTR., https://play.google.com/about/privacy-security-deception/ [https://perma.cc/DG49-XBPG] (last visited Dec. 30, 2019); *App Store Review Guidelines*, APPLE DEVELOPER, https://developer.apple.com/app-store/review/guidelines/ [https://perma.cc/MZ47-J7P3] (last visited Dec. 30, 2019).

3. Several companies collect and monetize location data, including precise GPS coordinates, the name of Wi-Fi routers, and whether users have Bluetooth on or off. *See, e.g.*, Michael Grothaus, *Google Tracks Your Movements Even if You've Turned Location History Off*, FAST COMPANY (Aug. 13, 2018), https://www.fastcompany.com/90217689/google-tracks-your-movements-even-if-youve-turned-location-history-off [https://perma.cc/4CNY-M2KZ]; Adrianne Jeffries, *Why Is This Company Tracking Where You Are on Thanksgiving?*, OUTLINE (Nov. 15, 2017, 9:50 AM), https://theoutline.com/post/2490/why-is-this-company-tracking-where-you-are-on-thanksgiving [https://perma.cc/4WN5-D3T4] (last visited Nov. 16, 2017); Taylor Hatmaker, *Users Dump AccuWeather iPhone App After Learning It Sends Location Data to a Third Party*, TECHCRUNCH (Aug. 22, 2017, 1:19 PM), http://social.techcrunch.com/2017/08/22/accuweather-revealmobile-ios/ [https://perma.cc/8NLX-3RPA]; Robbie Gonzalez, *The "Thanksgiving Effect" and the Creepy Power of Phone Data*, WIRED (May 31, 2018, 2:29 PM), https://www.wired.com/story/the-thanksgiving-effect-and-the-power-of-phone-data [https://perma.cc/CMV3-JUQ8]; Frank Bajak, *Mobile Carriers Cut Off Flow of Location Data to Brokers*, AP NEWS (Jun. 19, 2018), https://apnews.com/8582857aff8146f8ac81d247533b2177/APNewsBreak:-Verizon-to-end-location-data-sales-to-brokers [https://perma.cc/9Q5E-7SDV].

peoples' privacy expectations align with the dichotomy.[4] Contrary to received views,[5] we found that they do not align very well at all. Utilizing concepts from the theory of contextual integrity,[6] the first two sets of studies revealed that in the right circumstances (defined by social domains, recipients, and purposes), people are quite ready to share information deemed private with others. However, for information deemed public (so defined by its placement in public records), people maintain highly modulated privacy expectations.[7]

These studies extended over diverse categories of information types, but, quite early in their design, we set aside location, realizing that this category deserved special and separate attention. For one, location has had strong historical associations with both the private (e.g., one's home) and the public (e.g., the proverbial public square). For another, it has become a target of great interest and value as a raft of existing and emerging technologies have rendered location information accessible to an unprecedented degree. In so doing, these technologies and associated practices have muddied historical lines between public and private spaces, both by giving public exposure to that which was considered private, and also by revealing legitimate privacy interests in erstwhile public locations.

The focus of our studies here is the latter; that is, privacy interests in location data gleaned from spaces deemed public and historically not warranting legal or other forms of protection. Although novel capabilities eroding the sanctity of historically private spaces are deeply worrying,[8] the

4. For a fuller discussion of this point, see generally HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE (2010), especially Chapters 4, 5, and 6.

5. The clearest articulation of the private-public dichotomy is in the plain view and third-party doctrines; or, as summarized by Monu Bedi, the Fourth Amendment Disclosure Doctrines, which equate making something available to be seen as, therefore, relinquishing privacy expectations. Monu Bedi, *The Fourth Amendment Disclosure Doctrines*, 26 WM. & MARY BILL RTS. J. 461, 461–63 (2017); *see also* Ian Kerr & Jena McGill, *Emanations, Snoop Dogs and Reasonable Expectations of Privacy*, 52 CRIM. L.Q. 392, 407–11 (2007); Orin S. Kerr, *The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution*, 102 MICH. L. REV. 801 (2004); Orin S. Kerr, *The Case for the Third-Party Doctrine*, MICH. L. REV. 561, 566 (2009).

6. NISSENBAUM, *supra* note 4. For a definition of privacy as contextual integrity, see *infra* Section III.A. According to the theory of CI, whether privacy has been preserved or violated depends on whether a given flow of information (or data) is *appropriate*, which in turn depends on whether this flow conforms with entrenched and contextual informational norms (sometimes abbreviated as "privacy norms").

7. Kirsten Martin & Helen Nissenbaum, *Privacy Interests in Public Records: An Empirical Investigation*, 31 HARV. J.L. & TECH. 111 (2017); Kirsten Martin & Helen Nissenbaum, *Measuring Privacy: An Empirical Test Using Context to Expose Confounding Variables*, 18 COLUM. SCI. & TECH. L. REV. 176 (2017).

8. *See* Kerr & McGill, *supra* note 5, at 393–94 (describing how bodily emanations like sweat and scents can be harnessed by new technologies for surveillance purposes); Kerr, *The*

erosion of freedom in spaces deemed public seems to defy standard labels. The way we see it, regulation (or absence of regulation), guided by a principle of laissez-faire or "up for grabs," reflects intuitions based on the material capabilities of prior eras. Details aside, the so-called plain view or public disclosure doctrine is one such—a comfortable fit for traditionally-defined public spaces viewed through human eyes and recorded by notes on paper.[9] We should not be surprised, therefore, to discover that these ideas are desperately inadequate for public spaces of the present day—monitored by sophisticated systems of fixed and mobile networked sensors and recorded into computerized databases. Regulation that embodies intuitions and norms of past eras is bereft of concepts for handling present day privacy threats in historically public spaces, in turn handicapping courts and other regulatory efforts to identify, grasp, acknowledge, and protect against them. While people struggle to convey the nature of these wrongs, stakeholders continue to exploit this convenient lacuna.

Our studies offer insights into how people think about location data and the factors affecting how we evaluate common location-tracking practices. In so doing, these studies may serve the needs of courts, regulators, and system designers seeking to address diverse challenges without compromising the normative standing of privacy interests in location data. One important instance is the need to flesh out the meaning of "reasonable expectation of privacy" in the myriad of privacy cases that reach courts. Studies such as ours serve decision makers, including judges and regulators, who could benefit from robust empirical findings rather than intuition, hearsay, or anecdote as grounds for deciding whether practices in question either meet or do not meet societal expectations.[10] Likewise, social actors using and offering digital devices and

---

*Fourth Amendment and New Technologies*, *supra* note 5, at 865–66 (offering examples of how technological developments allow for increasing intrusion by law enforcement into private spaces).

9. *See* Bedi, *supra* note 5, at 470 ("[T]he public disclosure doctrine, which says that there is no privacy protection for a person's movements in public."); Kerr, *The Fourth Amendment and New Technologies*, *supra* note 5, at 827–28.

10. Professors Kugler and Strahilevitz nicely summarize why actual beliefs (as measured in surveys) are relevant to court opinions. Kugler & Strahilevitz, *supra* note 1, at 220 ("[W]e show how scientific polling can alleviate concerns that, in undertaking such an inquiry, judges will place undue weight on their own beliefs or on the beliefs of people in their social orbits."). Around location data specifically, Kugler and Strahilevitz quote Justice Alito, who argued that reasonable expectations of privacy are "the average person's expectations" or "popular expectations." *Id.* at 207 (quoting United States v. Jones, 132 S. Ct. 945, 964 (Alito, J., concurring)); *see also* Christopher Slobogin & Joseph E. Schumacher, *Reasonable Expectations of Privacy and Autonomy in Fourth Amendment Cases: An Empirical Look at Understandings Recognized and Permitted by Society*, 42 DUKE L.J. 727 (1992).

services would do well to heed these findings in order to comply with them and avoid scandals of noncompliance when discovered.[11]

Despite the great debates, published commentary, court action, regulatory activity, and scholarly literature, not enough is known about how people understand location data, what specifically about location tracking affects their judgments of it, and what their expectations are regarding others' access to their whereabouts.[12] Given breakneck development of location tracking systems and the fundamental importance of a reasonable expectation standard in deciding legal and regulatory questions about privacy, answers to these questions are urgently needed. Our studies seek to fill some of the gaps in knowledge by focusing on location data and location tracking in public places. One of the most dramatic findings is that people's expectations of privacy are not correlated with the traditional dichotomy of private versus public. Moreover, privacy expectations in public spaces are far from haphazard but are tied systematically to factors that our studies reveal.

## B. OUTLINE

Part II of this Article provides a backdrop for our studies showcasing related work on privacy and location data. We have highlighted work on location privacy in technology design, regulation, and the courts that has particularly informed and influenced our own. We also explain how our studies extend past and contemporaneous empirical work on location and privacy.

Part III describes the design of our studies, including the factorial vignette methodology. It also outlines the theory of contextual integrity, which provides the framework structuring our survey instrument.

In Part IV, we describe a series of pilot studies that guided the design of the main survey and were critical in informing its structure, such as the study's 'voice' and the ordering of the questions. Results, some of which were quite surprising, shaped our main studies.

Part V describes our main study. This study presented a series of scenarios involving the capture and flow of location data to a nationally representative sample of 1,500 respondents. Respondents were asked to rank these scenarios in terms of how appropriate they judged the practices to be.[13]

---

11. *See generally* Ira S. Rubinstein & Nathaniel Good, *Privacy by Design: A Counterfactual Analysis of Google and Facebook Privacy Incidents*, 28 BERKELEY TECH. L.J. 1333 (2013) (analyzing the ways in which digital services are being designed to violate privacy); WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES (2018).

12. *See* Kugler & Strahilevitz, *supra* note 1.

13. These results are reported in the full Article to explain the study design.

Part VI takes up a question that emerged from findings in both Pilot and Main studies. It was clear that we needed to learn more about how respondents conceptualized location and how this affected their expectations of location privacy. To this end, we investigated different ways of describing location tracking, from merely numeric representations to semantically meaningful descriptions of *place*. To isolate the importance of adding place to vignettes describing a generic location, we ran two factorial vignette surveys: the first merely referenced *location*, and the second referenced a meaningful *place* (e.g., school, hospital).

In Part VII, we discuss the significance of the findings of all three studies for technology, regulation, and the courts. Our results immediately debunk the idea that people have no expectations of privacy in public.[14] The findings call common practices of amassing location data by government and commercial entities into question by showing that these practices flout expressed privacy expectations in systematic and specific ways.

The studies further reveal that *how* we ask about location in surveys makes a difference to how people react. Details such as duration of collection, place, and inferences drawn significantly affect respondent ratings. Strikingly, the respondents were far more attuned to location tracking when it revealed place (e.g., home, work, shopping) than GPS coordinates. By implication, regulating standard technical markers (e.g., GPS) representing location in technical systems may not assuage location privacy worries. Another surprising result is that the duration of location-tracking loses significance when inferences are drawn, which suggests that inference trumps duration and that concerns over duration may be proxies for more fundamental concerns over what can be inferred from longer-term location surveillance.

Finally, in line with our earlier studies, respondents consistently found most repugnant data capture and flow practices involving data aggregators or data brokers.

## II. BACKGROUND AND RELATED WORK

Our work has been prompted and shaped by much that has come before, including the developmental trajectories of technology, regulation, and court decisions. It has also drawn from related empirical work, which like ours has sought to understand the influence of diverse factors over privacy expectations concerning location. A caveat (for which we hope to be forgiven) is that in

14. *See* Helen Nissenbaum, *Protecting Privacy in an Information Age: The Problem of Privacy in Public*, 17 LAW & PHIL. 559 (1998) (examining why theories of privacy neglect or dismiss questions of privacy in public).

acknowledging influences from all four domains—technology, courts, regulation, and empirical studies—we have had to be selective in reviewing each of them.

### A. TECHNOLOGY

This Section provides a selective survey of technologies that enable and facilitate the monitoring and tracking of individuals through space, with a focus on mobile devices or "smartphones." A person's whereabouts may be noted, tracked, and recorded by a variety of means, ranging from the plain sight of other people to technology-enabled image capture. The class of digital technologies that generate and record location data is broad and diverse, including fixed sensors that locate individuals within their ranges to mobile location sensors that people increasingly carry around with them. Such technologies span traditional CCTV systems to newer forms of networked cameras (still and video), license plate readers, RFID tags associated with traditional forms of identification (e.g., credit cards or passports), mobile phones, Internet-of-Things (IoT) devices, location-specific social media, and more. The emerging arena of urban tech—so-called "smart cities"—which, by definition, involves a myriad of system-integrated sensors interacting with physical bodies in motion as well as signals from mobile devices, introduces acute privacy challenges.[15] Few are more urgent than those associated with the capture of location data generated by individuals via innumerable transceivers "communicating" with an equally diverse range of transmitters from familiar mobile phones to novel, smart (driverless) vehicles.

This expanding array of location-generating and location-capture technologies requires a full reckoning outside the scope of this Article; however, a closer examination of one case, namely smartphones, helps to showcase at least one reason why location privacy has fallen into a mire of confusion. We further confine the examination under this heading to devices powered by Apple's iOS and Google's Android OS.[16] Without doing justice to all relevant developments, it is fair to say that since we began our studies of the determinants of privacy expectation roughly four year ago, advances in the scope and sophistication of consumer mobile technologies have been staggering.

For the two major competing mobile operating system (OS) platforms, numbers, one might say, are the tail that wags the dog. The more apps and app

15. *See, e.g.*, BEN GREEN, THE SMART ENOUGH CITY: PUTTING TECHNOLOGY IN ITS PLACE TO RECLAIM OUR URBAN FUTURE (2019).

16. The discussion of mobile privacy owes a huge debt to Mainack Mondal and Eran Toch, who should not, however, be blamed for any inaccuracies.

developers are attracted to respective operating systems, the greater the value to users and, so the argument goes, the greater the likelihood they will choose respective operating systems. To take one slim measure, the number of offerings in Apple's app store jumped from 800 in 2008 to 1 million in 2013.[17] And within the four-year timespan of our studies, the number jumped from 1.3 million in 2014 to 2.1 million by 2017.[18] With respect to the Android operating system, while slower to introduce third-party apps, Play Store offerings grew from 1.38 million in 2014 to 2.7 million by 2017.[19]

It is not surprising that, in reverse symbiosis, Apple and Google extend capacity and power to developer communities through Application Programming Toolkits (APIs) and Software Developer Kits (SDKs)[20] to capitalize on data naturally generated by their respective systems. For location, the OS provides not only GPS, but other markers such as position in relation to nearby Wi-Fi routers[21] and the closest cellular service towers. In addition to location markers, iOS and Android OSs are constantly updating, refining, and augmenting their offerings with a myriad of others (gyroscope, compass, identity verification, time, etc.) in service of the nearly 5 million total apps in the App Store and Play Store. Various uses of these developer kits and interfaces have stoked public outcry. For example, the popular Brightest Flashlight app was discovered to be tracking users' location and selling it to

17. Caroline McCarthy, *Apple: One Million iPhones Sold, 10 Million App Store Downloads in First Weekend*, CNET (July 15, 2008), https://www.cnet.com/news/apple-one-million-iphones-sold-10-million-app-store-downloads-in-first-weekend/ [https://perma.cc/AL6Z-6JAB]; *App Store Sales Top $10 Billion in 2013*, APPLE (Jan. 7, 2014), https://www.apple.com/newsroom/2014/01/07App-Store-Sales-Top-10-Billion-in-2013/ [https://perma.cc/B6TZ-DQC2].

18. Nick Summers, *The App Store Now Boasts 1.3 Million iOS Apps*, NEXT WEB (Sept. 9, 2014), https://thenextweb.com/apple/2014/09/09/now-13million-apps-app-store/ [https://perma.cc/U2A4-BX9R]; Shannon Liao, *Apple's Total Number of Apps in the App Store Declined for the First Time Last Year*, VERGE (Apr. 5, 2018, 6:07 PM), https://www.theverge.com/2018/4/5/17204074/apple-number-app-store-record-low-2017-developers-ios [https://perma.cc/ZJ6R-ZBLV].

19. *Number of Android Applications*, APP BRAIN STATS (Oct. 5, 2014), https://web.archive.org/web/20141006142446/https://www.appbrain.com/stats/number-of-android-apps [https://perma.cc/L5EA-GJ7G]; *Number of Android Applications*, APP BRAIN STATS (Feb. 9, 2017), https://web.archive.org/web/20170210051327/https://www.appbrain.com/stats/number-of-android-apps [https://perma.cc/ZB8D-SLCD].

20. APIs and SDKs provide convenient programming interfaces that aid application developers in making their systems function within operating systems, such as mobile operating systems, or platforms, such as Facebook.

21. *See, e.g.*, WIGLE.NET, https://wigle.net/ [https://perma.cc/9LZF-GLF6] (last visited Dec. 30, 2019) (offering geolocated Wi-Fi network services).

third parties,[22] the Weather Channel was sued by the city attorney of Los Angeles for passing its users' location data to other IBM-owned services as well as outside entities,[23] and Accuweather stirred ire when investigators discovered that it was recording and selling location data even after users had said no.[24]

To rein in practices where app developers extract ostensibly unnecessary data, government regulators and OS providers have tightened policies for accessing various classes of information. Because of growing public distaste over stealth capture of device-generated data, regulators and OS providers are suggesting, and in some cases requiring, just-in-time, explicit requests for access to specific categories of data, with location data an important category among those singled out for special treatment.[25] Should we be satisfied that, with these explicit requests, websites, services, and mobile apps are finally doing right by their users? Can users be confident that their expressed preferences will, in fact, determine how location data is handled "in the machine" and beyond? Will their expectations be met?

In our view, the only correct answer to these questions is "we don't know," because the internal practices of OS providers, as well as the data flowing back and forth between the OS and app providers, remain opaque to the vast majority of users and to regulators. Only with considerable ingenuity have experts developed tools, such as Serge Egelman's AppCensus, to ferret out some level of insight, far from complete.[26] But another reason, not previously recognized, why these questions are impossible to answer directly, is the conceptual ambiguity of location. In turn, this conceptual ambiguity poses challenges even to good faith efforts to regulate location tracking and to

22. Robert McMillan, *The Hidden Privacy Threat of…Flashlight Apps?*, WIRED (Oct. 20, 2014, 6:30 AM), https://www.wired.com/2014/10/iphone-apps/ [https://perma.cc/HQC4-GY8A].

23. *See* Complaint for Injunctive Relief and Civil Penalties for Violations of the Unfair Competition Law, *People v. TWC Prod. and Tech., L.L.C.* (2019), https://int.nyt.com/data/documenthelper/554-l-a-weather-app-location/8980fd9af72915412e31/optimized/full.pdf [https://perma.cc/57CT-RU5X].

24. *See* Hatmaker, *supra* note 3.

25. Currently, there are twenty-eight such categories requiring special permissions, out of a total of ninety-one possible. *Permissions Overview*, ANDROID DEVELOPERS, INTERNET ARCHIVE, https://web.archive.org/web/20190303040327/https://developer.android.com/guide/topics/permissions/overview [https://perma.cc/K7MU-PCFM] (last visited Dec. 30, 2019).

26. *See* Irwin Reyes et al., *"Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale*, PROC. ON PRIVACY ENHANCING TECH., June 2018, at 63–83 (analyzing the privacy behavior of the mobile apps by "dynamic test," which contains App Corpus, Analysis Environment, Event Extraction, etc.). The AppCensus search is available at https://search.appcensus.io [https://perma.cc/6TSR-9VGY] (last visited Dec. 30, 2019).

represent and enforce it in systems in concert with the ways people conceive, interpret, and value it. In other words, technical efforts to protect location privacy may stumble because of a failure to map the meaning that people assign to location with its representations in technical systems.

To illustrate the discrepancy between the meaning that people assign location with its representation in technical systems, let us return to the Accuweather scandal and consider a hypothetical explanation that gives Accuweather the benefit of the doubt. To begin, let's assume that Accuweather represented location in the system as coordinates derived from GPS. When users answered "no" to location tracking, Accuweather respected this expressed preference by ceasing to attach GPS coordinates to their respective records. Still wanting information about users' whereabouts, it sought alternative markers; in this instance, lookup tables from closest Wi-Fi routers. While users might be outraged by the workaround, Accuweather could counter that by ceasing to collect GPS signal, they were dropping location as it is normally represented in its system. Although we have not seen evidence of this precise dialog, the indignation registered in reports of this incident suggests that people are neither attuned to nor impressed by such distinctions. Our hypothetical account could continue. Even assuming that Accuweather has taken this criticism to heart and now eschews location markers drawn from GPS, Wi-Fi, and cellular towers, they have not exhausted all sources: in particular, semantic sources. Consider, for example, a user paying with anything but cash at a CVS branch on Bleecker Street, New York City. In this case, location is rendered semantically as "a CVS drugstore on Bleecker." Such data also may have been shared in the text message to a friend, "I am just finishing up at the CVS on Bleecker!" or tagged in a selfie posted on Instagram. Accuweather could hypothetically purchase such information from CVS, or one of the many location data brokers.

The point is that location can be characterized in many different ways, from GPS coordinates to semantically rich labels. Viewed in this light, one could conceive of the constellation of location-tracking mobile apps as a massive and distributed system for producing layer upon layer of meaning to numeric location coordinates. This system is akin to Geographic Information Systems, which attach meaningful labels to numerical geographical coordinates, but far more varied and potentially threatening. Similarly, meanings that apps attach to particular locations may be rich and complex, and potentially uncomfortably revealing. For example, in a familiar case, an app may identify a given set of coordinates as a person's "home" or "work." In more complicated instances, it may connect locations with app-labeled activities (e.g., "exercise" or "having sex") or even through co-presence with

other people (e.g., in social apps).[27] With greater sophistication, these systems may infer even more.

To put the conundrum plainly, when people respond "no" to location tracking, what is it that they believe, expect, and want to be happening? And whatever this is, does it map onto how systems developers represent and enforce this? For anyone committed to privacy-by-design or, more concretely, committed to ensuring that people's location privacy expectations can be represented and enforced in technical systems, a sound mapping between those expectations and those systems is a necessary condition. The goal of such a mapping between technical representations and people's privacy expectations is a key motivator of our work. As such, we have also sought to demonstrate where revealed expectations currently are asynchronous with efforts on the technical side.

### B. REGULATION

In this Section, we discuss the regulation of the location-tracking practices of commercial entities. The sources of this regulation are far less clear than the constitutional principles that apply to governmental actors, as discussed below in Section II.C.

As we know, the information technology and service industry functions under a model of "self-regulation,"[28] particularly in relation to privacy. Following concerns over the information practices of apps, the major mobile operating systems have issued sets of policies and guidelines for app developers.[29] Acknowledging deep anxiety over location, as noted above in Section II.A, they have become more demanding in requiring mobile app developers to provide finer grained notices about the data fields they seek to

---

27. *See* Jennifer Valentino-De Vries et al., *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. TIMES (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html [https://perma.cc/RF53-BSZT].

28. Mary J. Culnan, *Protecting Privacy Online: Is Self-Regulation Working?*, 19 J. PUB. POL'Y & MARKETING 20, 20–26 (2000); FED. TRADE COMM'N, PRIVACY ONLINE: FAIR INFORMATION PRACTICES IN THE ELECTRONIC MARKETPLACE 1–7 (2000), https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission-report/privacy2000.pdf [https://perma.cc/SE88-33SK]; Siona Listokin, *Industry Self-Regulation of Consumer Data Privacy and Security*, 32 J. MARSHALL J. INFO. TECH. & PRIVACY L. 15, 15 (2015).

29. *See, e.g.*, *Developer Policy Center*, GOOGLE PLAY, https://play.google.com/about/developer-content-policy/ [https://perma.cc/S2B8-AKPU] (last visited Dec. 30, 2019) (describing Android policies); *App Review*, APPLE, https://developer.apple.com/app-store/review/ [https://perma.cc/EXP8-VL3M] (last visited Dec. 30, 2019) (describing iOS policies).

collect as well as finer grained choices for users, particularly as applied to location data.[30]

Although these policies and guidelines have somewhat constrained app developer access to user data generated by mobile devices,[31] by no means do they address the full scope of vulnerability to location tracking. First, quite obviously, location tracking is not limited to mobile apps; for example, fitness trackers may provide users with information about their runs by mapping and measuring their routes.[32] Second, the guidelines have still not stopped controversial practices that have raised eyebrows, if not vocal protest.[33] For example, having secured users' permission to monitor location data, companies may then provide this data to brokers.

One might argue that the status quo is not surprising, given the general backdrop of weak privacy regulation in the United States. Over the past decade, however, due to increasing pressure from advocacy organizations[34] and the public exposure of high-profile industry missteps,[35] the appetite for

30. *See Permissions Overview*, ANDROID DEVELOPER, *supra* note 25.

31. For example, from Google Play Developer Policy Center:

> Limit your collection and use of this data to purposes directly related to providing and improving the features of the app (e.g. user anticipated functionality that is documented and promoted in the app's description).
>
> Post a privacy policy in both the designated field in the Play Console and within the app itself. The privacy policy must, together with any in-app disclosures, comprehensively disclose how your app collects, uses, and shares user data. Your privacy policy must disclose the type of parties to which any personal or sensitive user data is shared.
>
> Handle all personal or sensitive user data securely, including transmitting it using modern cryptography (for example, over HTTPS).

*Personal and Sensitive Information*, GOOGLE PLAY, https://play.google.com/about/privacy-security-deception/#!?zippy_activeEl=personal-sensitive#personal-sensitive [https://perma.cc/MXM8-JDXW] (last visited Dec. 30, 2019).

32. *See, e.g.*, Alex Hern, *Fitness Tracking App Strava Gives Away Location of Secret US Army Bases*, GUARDIAN (Jan. 28, 2018), https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases [https://perma.cc/29CQ-SPZP]; Liz Sly, *U.S. Soldiers Are Revealing Sensitive and Dangerous Information by Jogging*, WASH. POST (Jan. 29, 2018, 2:22 AM), https://www.washingtonpost.com/world/a-map-showing-the-users-of-fitness-devices-lets-the-world-see-where-us-soldiers-are-and-what-they-are-doing/2018/01/28/86915662-0441-11e8-aa61-f3391373867e_story.html [https://perma.cc/V3DH-ZNEQ].

33. *See* Valentino-DeVries et al., *supra* note 27.

34. *See, e.g.*, ELECTRONIC PRIVACY INFORMATION CENTER, https://epic.org [https://perma.cc/NT5T-E2WP] (last visited Dec. 30, 2019); ELECTRONIC FRONTIER FOUNDATION, https://www.eff.org [https://perma.cc/TD8A-WGDH] (last visited Dec. 30, 2019).

35. *See, e.g.*, Matt Warman, *Google: We Failed to Delete All Streetview Data*, TELEGRAPH (July 27, 2012), https://www.telegraph.co.uk/technology/google/9432518/Google-we-failed-to-delete-all-Streetview-data.html [https://perma.cc/5NQA-42CU]; Ritchie S. King & Mika

privacy regulation is slowly growing, with location privacy at the leading edge. A 2013 FTC Staff Report defined geolocation as "critical information" in need of greater regulation,[36] and location data was the focus of the Future of Privacy Forum's "Mobile Location Analytics Code of Conduct."[37] Yet even while warning that location data as generated by and garnered from mobile devices may be deeply revealing, these documents did not disrupt the reigning notice-and-choice model and merely offered "suggestions" and "recommendations" for how to communicate location data practices with greater salience, such as with "just-in-time" notices. Although this model allowed the FTC to issue a complaint against Goldenshores Technologies, LLC, maker of the "Brightest Flashlight" Android app, for misrepresenting its privacy practices,[38] it is impotent against accurate representations that are nevertheless incomplete and difficult to follow.

There is sufficient alarm over the insidious practices surrounding location data that it has gained the attention of lawmakers. Notably, in the European Union, the General Data Protection Regulation (GDPR), implemented in May 2018, singled out location data for special attention along with other types of data in the tightly regulated category of personally identifying information.[39]

---

Gröndahl, *How Google Collected Data from Wi-Fi Networks*, N.Y. TIMES (May 23, 2012), https://archive.nytimes.com/www.nytimes.com/interactive/2012/05/23/business/How-Google-Collected-Data-From-Wi-Fi-Networks.html [https://perma.cc/KP2P-NFYB]; Matthew Rosenberg et al., *How Trump Consultants Exploited the Facebook Data of Millions*, N.Y. TIMES (Mar. 17, 2018), https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html [https://perma.cc/U67M-TN3K]; Matthew Rosenberg & Sheera Frenkel, *Facebook's Role in Data Misuse Sets Off Storms on Two Continents*, N.Y. TIMES (Mar. 18, 2018), https://www.nytimes.com/2018/03/18/us/cambridge-analytica-facebook-privacy-data.html [https://perma.cc/4A7L-YSJM].

36. FED. TRADE COMM'N, MOBILE PRIVACY DISCLOSURES: BUILDING TRUST THROUGH TRANSPARENCY 3 (2013), https://www.ftc.gov/sites/default/files/documents/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission-staff-report/130201mobileprivacyreport.pdf [https://perma.cc/JT6B-QB2A].

37. *Mobile Location Analytics Code of Conduct*, FUTURE PRIVACY F. (Oct. 22, 2013), https://fpf.org/wp-content/uploads/10.22.13-FINAL-MLA-Code.pdf [https://perma.cc/3CZE-JUNT].

38. *Android Flashlight App Developer Settles FTC Charges It Deceived Consumers*, FED. TRADE COMMISSION (Dec. 5, 2013), https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived [https://perma.cc/ZAY9-6JYS].

39. In the GDPR, personal information includes "name, identification number, location data or online identifier . . . ." *Frequently Asked Questions about the GDPR*, EU GDPR PORTAL, http://eugdpr.org/gdpr-faqs.html [https://perma.cc/F27U-H2K6] (last visited Sept. 5, 2018); *Overview of the General Data Protection Regulation (GDPR)*, INFO. COMMISSIONER'S OFF. (2016), https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/ [https://perma.cc/853C-GQTA] (last visited Dec. 30, 2016); ICO, *What is personal data?*, ICO'S GUIDE GDPR (2019), https://ico.org.uk/for-organisations/guide-to-data-protection

Specifically, processing identifiable information is regulated. To do so, data processors must meet one of a few criteria: the processing of the data must be necessary (1) to complete a contractual obligation, (2) to protect vital interests of the data subject or other person, (3) to perform a task in the public interest, (4) to comply with a law or regulation, or (5) "for the legitimate interests" pursued by the data controller or a third party.[40] These requirements would, for example, clearly and immediately rule out Brightest Flashlight.

Our assessment is that as the hardware, software, and political economy of data advance, the practices of location tracking are diverging from people's expectations of appropriate behaviors. These discrepancies between expectations and common practices, despite efforts to regulate, suggest at least two possibilities, not necessarily mutually exclusive. First, the crafters of regulation, government and industry, are knowingly trading off privacy expectations and interests of data subjects in favor of location data collectors. Or, second, they do not properly grasp how people understand and value location data. Although our studies mainly shed light on the latter possibility, in so doing, they raise the stakes by revealing the nature and extent of the tradeoff.

### C. COURTS

In this Section, we consider how the courts have dealt with privacy and location data. Historically, the "third-party doctrine" has reflected the idea that individuals have no reasonable expectation of privacy in information they willingly give to others. But two landmark court cases have suggested that the doctrine is stretched thin in the face of location tracking technologies. First, in *United States v. Jones*, the Supreme Court held that police could not attach a GPS device to a defendant's vehicle and track its movement for a period of twenty-eight days. While the majority focused on the trespass to property, Justice Sotomayor wrote in a concurring opinion that the third-party doctrine is "ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks."[41] In a second case, *Carpenter v. United States*, the Court held that a

/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/ [https://perma.cc/5MAY-ASQJ] (last visited Dec. 30, 2019); *GDPR FAQs*, EUGDPR.ORG, http://eugdpr.org/gdpr-faqs.html [https://perma.cc/CML4-77DL] (last visited Sept. 5, 2018).

40. *Lawful Basis for Processing*, INFO. COMMISSIONER'S OFF., https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/ [https://perma.cc/N95S-GWNB] (last visited Dec. 30, 2018).

41. United States v. Jones, 565 U.S. 400, 417 (2012) (Sotomayor, J., concurring) ("[I]t may be necessary to reconsider the premise that an individual has no reasonable expectation

defendant had a reasonable expectation of privacy in his cell phone's location data, even though it was in the hands of his service provider, a third party.[42]

In the legal literature, much has been written about these two important cases and others involving location tracking.[43] Insofar as they relate to and influence our work, we have focused on factors that have systematically affected how courts have resolved questions about reasonable expectations of privacy in location, and how these factors have evolved over time. Guided by the theory of contextual integrity and characterizing location tracking practices as special cases of information flow, we considered how courts took the following features into consideration in determining whether practices involving the collection and uses of location data were legally acceptable: (1) who collects the data, (2) how it is gathered, and (3) the meaning that can be extracted from it.

*1. Who Collects Location Data Is Important*

An initial factor critical to determining whether reasonable expectations of privacy have been respected is *who* collects the data (or in contextual integrity terms, who receives the data). The courts have often differentiated between law enforcement versus private actors, with the former subject to rigorous constitutional constraints and the latter to far fewer.[44] The rise of commercial information intermediaries such as data brokers and credit agencies drove an active discussion in the courts and among legal scholars about the third-party doctrine.[45] This discussion focused specifically on the legal issues when intermediaries, with whom one has no reasonable expectations of privacy, provide information to government agencies, with whom one has a

---

of privacy in information voluntarily disclosed to third parties . . . . This approach is ill-suited to the digital age . . . .").

42. Carpenter v. United States, 138 S. Ct. 2206, 2222 (2017) (holding that individuals, in "rare case[s]," may have "a legitimate privacy interest in records held by a third party"); *see also* Bedi, *supra* note 5, at 486–88.

43. *See, e.g.*, Christopher Slobogin, *Making the Most of United States v. Jones in a Surveillance Society: A Statutory Implementation of Mosaic Theory*, 8 DUKE J. CONST. L. & PUB. POL'Y 1 (2012); Kevin S. Bankston & Ashkan Soltani, *Tiny Constables and the Cost of Surveillance: Making Cents Out of United States v. Jones*, 123 YALE L.J. 335 (2013); Paul Ohm, *The Many Revolutions of Carpenter* 32 HARV. J.L. & TECH. 357 (2019); Orin S. Kerr, *Initial Reactions to Carpenter v. United States* (USC Law Legal Studies Paper No. 18-14, 2018).

44. Kiel Brennan-Marquez, *Outsourced Law Enforcement*, 18 U. PA. J. CONST. L. 797, 797–99 (2016) (explaining that Fourth Amendment protections extend only to law enforcement seeking to gain information about citizens; commercial entities are able to surveil citizens at any time).

45. *See, e.g.*, Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 528 (2006); Susan Freiwald, *Cell Phone Location Data and the Fourth Amendment: A Question of Law, Not Fact*, 70 MD. L. REV. 681, 733 (2010).

constitutionally-based reasonable expectation of privacy.[46] One analogy supporting the third-party doctrine was likening private firms providing information to government actors to confidential informants, thus putting the onus on individuals such as clients, customers, and consumers for their misplaced confidences in untrustworthy actors or firms with whom they interact.[47] One problem with a focus on the actor as determinative of the norms of collecting and using location data is that law enforcement can then simply get the information from private parties.[48]

No matter what one's view on past cases, it would take willful avoidance to ignore epic transformations in the informational landscape. Writing about the burgeoning data broker industry, ranging from general brokers (such as Acxiom) to specialized providers (including some that focus on location data),[49] Chris Hoofnagle and others warn against private actors serving as government surrogates, calling them "Big Brother's Little Helpers."[50]

Another aspect of this transformation is the gradual elimination of choice in the transfer of data from individuals such as subscribers, consumers, and customers, to third parties, which are increasingly online, as a condition of a diverse array of services and transactions. This has led to a literature debating the idea of information intermediaries as fiduciaries.[51] Without pursuing this debate further, to us, significant progress will not be made that makes the actors in question—government or private—determinative of appropriate

46. *See* Kerr, *The Case for the Third-Party Doctrine*, *supra* note 5; Bedi, *supra* note 5.

47. In other words, the individual is at fault for sharing information with informants who, in turn, share that information with the government, whether a confidential informant in a criminal conspiracy or an untrustworthy firm with whom data is shared. *See* David C. Gray & Danielle Keats Citron, *The Right to Quantitative Privacy*, 98 MINN. L. REV. 62, 86 n.458 (2013) (describing the "misplaced trust rationale"); Kerr, *The Case for the Third-Party Doctrine*, *supra* note 5, at 568 (explaining that the Fourth Amendment does not protect defendants' misplaced confidence) (citing Lewis v. United States, 385 U.S. 206 (1966)).

48. Gray & Citron, *supra* note 47, at 140 ("If the government lacks legal authority to install and monitor a GPS-enabled tracking device, then it can get the same information by securing locational data from OnStar, Lojac, a cellular phone provider, or any number of 'apps' that gather and use locational information as part of their services.").

49. *See* Valentino-DeVries et al., *supra* note 27.

50. Chris Jay Hoofnagle, *Big Brother's Little Helpers: How Choice Point and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement*, 29 N.C. J. INT'L L. & COM. REG. 595 (2003); Michael D. Birnhack & Niva Elkin-Koren, *The Invisible Handshake: The Reemergence of the State in the Digital Environment*, 8 VA. J.L. & TECH. 1 (2003).

51. *See, e.g.*, Kiel Robert Brennan-Marquez, *Fourth Amendment Fiduciaries*, 84 FORDHAM L. REV. 611 (2015); Neil M. Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431 (2016); Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183 (2016); Lina Khan & David Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497 (2019).

action; that is, action that conforms with reasonable privacy expectations. This point is elaborated in Part III below.

*2. How Location Data Is Collected Is Important*

Some legal scholars have focused on *how* location data is collected as determinative of the norms of data collection. David Gray and Danielle Citron focus on the investigative technique used to surveil the individual.[52] They argue that the technological advancements around indiscriminate data collection, aggregation, and storage remove the practical limitations on surveillance and, by this capability, run afoul of the traditional Fourth Amendment prohibition on dragnets.[53] Similarly, Margaret Hu shifts to a non-intrusion test to justify surveillance.[54] Hu focuses on big data technologies that facilitate horizontal cybersurveillance as a new technique.[55] Katherine J. Strandburg also argues that courts should apply a principle of technosocial continuity to respect privacy expectations of individuals.[56] The principle of technosocial continuity "requires that courts consider both the ways in which technology facilitates intrusive surveillance and the ways in which technology spurs social change that may make citizens more vulnerable to existing surveillance technologies."[57]

Arguments to tie privacy expectations of location data to how the data is collected—if the technique is too invasive or pervasive, then privacy expectations are violated—closely align with Harry Surden's theory of

52. Gray & Citron, *supra* note 47, at 102 ("Among the important factors that a court would need to consider are: (1) the inherent scope of a technology's surveillance capabilities, be they narrow or broad; (2) the technology's scale and scalability; and (3) the costs associated with deploying and using the technology.").

53. *Id.* at 102.

54. Margaret Hu, *Cybersurveillance Intrusions and an Evolving Katz Privacy Test*, 55 AM. CRIM. L. REV. 127, 131 (2018) ("During oral argument in *Jones*, and in concurrences by Justices Alito and Sotomayor, the Court suggested that a nonintrusion test may be more appropriate given the scope of developing technology. A nonintrusion test is grounded in customary law, replacing an interpretation of the Fourth Amendment that is currently grounded in property and tort law, and presents a way to untether concepts of privacy from nondisclosure.").

55. *Id.* at 361 ("Horizontal cybersurveillance makes possible what has been termed as 'sentiment analysis.' Sentiment analysis can be described as opinion mining and social movement forecasting. Through sentiment analysis, mass cybersurveillance technologies can be deployed to detect potential terrorism and state conflict, predict protest and civil unrest, and gauge the mood of populations and subpopulations. Horizontal cybersurveillance through sentiment analysis has the likely result of chilling expressive and associational freedoms, while at the same time risking mass data seizures and searches.").

56. *See* Katherine J. Strandburg, *Home, Home on the Web and Other Fourth Amendment Implications of Technosocial Change*, 70 MD. L. REV. 614, 619 (2010).

57. *See id.*

structural privacy rights.[58] According to Surden, physical, societal, and technological constraints combine to make certain activities, including surveillance, difficult to complete without heavy costs; when one of these constraints is penetrated, we see our privacy as violated.[59] For Gray and Citron, technological advances serve to remove the structural constraints previously curtailing mass surveillance;[60] whereas obscurity, as defined by Professors Frederic Stutzman and Woodrow Hartzog, can be seen as adding to structural constraints.[61]

Along the line of location data collection and duration, Matthew Kugler and Lior Strahilevitz have examined if the duration of GPS data collection impacts people's reasonable expectations of privacy. Specifically, Kugler and Strahilevitz test the importance of duration in how the public regards the appropriateness of law enforcement needing a warrant to gather GPS data; they find it has no significant effect.[62] Importantly, these scholars frame the technology used to collect the data as critical to understanding whether privacy expectations are violated in the collection of location data.[63]

---

58. Harry Surden, *Structural Rights in Privacy*, 60 SMU L. REV. 1605 (2007).

59. *Id.*

60. Gray & Citron, *supra* note 47, at 63–67.

61. Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CALIF. L. REV. 1, 35–36 (2013) ("[W]e have identified four of these key factors: (1) search visibility, (2) unprotected access, (3) identification, and (4) clarity. The presence of these factors diminishes obscurity, and their absence enhances it. Thus, in determining whether information is obscure online, courts should consider whether any of these factors were present. Information that is entirely unobscure is completely obvious, and vice versa.").

62. The authors ask a single question: Would it "violate people's reasonable expectations of privacy if law enforcement" (1) used a car's onboard GPS system to locate it on public streets without the owner's permission? (2) used a car's onboard GPS system to track its movements on public streets for one day without the owner's permission? (3) same, but for one week? (4) same, but for one month? Kugler & Strahilevitz, *supra* note 1, at 246.

63. Rachel Levinson-Waldman argues that the following are important factors to consider in examining surveillance technologies:

> (1) the duration of the surveillance; (2) the lowering of structural barriers to pervasive surveillance, reflected in the greatly reduced cost of tracking; (3) the recording of an individual's or group's movements; (4) the elicitation of information from within a protected space such as a home; and, as appropriate, (5) whether the technology undermines core constitutional rights and (6) whether surveillance technologies are piggy-backed on each other. Pulling out and articulating these factors, and analyzing how and why they should be considered, seeks to add rigor to the improvisatory method that has defined the judiciary's consideration of these questions.

Rachel Levinson-Waldman, *Hiding in Plain Sight: A Fourth Amendment Framework for Analyzing Government Surveillance in Public*, 66 EMORY L.J. 527 (2016). The article goes on to examine various types of surveillance technologies (e.g. GPS, cellular phones, video cameras, drones, license plate readers, and body-worn cameras). *Id.*; *see also* Christopher Slobogin, *Making the*

*3. What May Be Inferred on the Basis of the Location Data in Question*

Pertinent to our work is the way courts have increasingly acknowledged the power of information technologies to transform information about one thing into another. Thus, in addition to how intrusive or pervasive are the *modes* of information collection, an important question is *what more can be inferred* from the information collected.

In other words, the methods for gathering information and the duration of the collection have historically been seen as a technological Peeping Tom peering into previously practically obscure spaces.[64] More recently, however, in both the Jones and Carpenter cases, location data over a period of time has been flagged for its capacity to generate new knowledge. The duration of the surveillance tells a new story about the individual, and individuals have a reasonable expectation of privacy in the whole of their movements.[65] Until now, "the Supreme Court has tended to pay more attention to the nature of the police intrusion required to obtain information than to the nature of the information obtained."[66]

Paul Ohm takes up this shift from the duration of surveillance being a problematic technique to the duration of surveillance capturing new information and quotes the lower court in *Jones*: "[p]rolonged surveillance reveals types of information not revealed by short-term surveillance, such as what a person does repeatedly, what he does not do, and what he does ensemble. These types of information can each reveal more about a person than does any individual trip viewed in isolation."[67]

For Ohm, the recent rulings validate the mosaic theory where the "accumulation of so many individual bits about a person's life" results in a "personality picture that is worthy of conditional protection."[68] Importantly,

---

*Most of United States v. Jones in a Surveillance Society: A Statutory Implementation of Mosaic Theory*, 8 DUKE J. CONST. L. & PUB. POL'Y 1 (2012) (concerning the duration of collection as important to understand privacy expectations around location data).

64. *See, e.g.*, Levinson-Waldman, *supra* note 63, at 561–62 (arguing that duration could work as "a substantial intrusion on individuals' privacy and diminish[] the obscurity that many people take for granted in their day-to-day movements . . . . The addition of technology has thereby both raised the stakes and lowered the barriers to intensive, intrusive surveillance").

65. *See* Carpenter v. United States, 138 S. Ct. 2206, 2217 (2018) ("A majority of this Court has already recognized that individuals have a reasonable expectation of privacy in the whole of their physical movements.") (quoting United States v. Jones, 400, 430 (Alito, J., concurring)).

66. Paul Ohm, *supra* note 43, at 362.

67. *Id.* at 373 (citing United States v. Maynard, 615 F.3d 544, 562 (D.C. Cir. 2010), *aff'd in part sub nom.* United States v. Jones, 565 U.S. 400 (2012)).

68. Slobogin, *supra* note 43, at 3–4; *see also* United States v. Jones, 565 U.S. 400, 416 (2012) (Sotomayor. J., concurring) ("I would ask whether people reasonably expect that their

this line of argument focuses on a new type of information that is revealed through the collection of location data as animating privacy concerns.[69]

### D. RELATED EMPIRICAL WORK

Finally, we connect our studies with important instances of prior empirical work around privacy expectations and location data that has inspired and influenced it. We include work from the survey research literature examining privacy expectations primarily for purposes of influencing social science, law, and regulation. Further, we include empirical work in the user experience literature, primarily informing and targeting technology developers and designers, while aware that regulators are paying attention.

Previous work on location data has focused on the degree to which the method of collection (GPS tracker versus cell phone tower data) or duration of collection matters to reasonable expectations of privacy. The collecting agent is usually explicitly law enforcement. The closest attempt to measure privacy expectations surrounding the collection of location data centers on GPS location, law enforcement, and the duration of the collection.[70] In this study, Matthew Kugler and Lior Strahilevitz conducted a nationally representative survey to test the duration of location data collection that individuals judge as within their privacy expectations.[71] Their specific focus was on law enforcement. They tested whether duration (one day, one week, one month) impacted the degree to which use of "a car's onboard GPS system to locate it on public streets without the owner's permission" met privacy expectations.[72] The authors found that duration "barely affects" the degree to which the public regards geolocation tracking as invading their reasonable expectations of privacy.[73]

Alisa Smith, Sean Madden, and Robert Barton empirically examined how the method of government data collection impacted privacy, and found that respondents disapproved of government intrusion with aerial surveillance, a

---

movements will be recorded and aggregated in a manner that enables the Government to ascertain, more or less at will, their political and religious beliefs, sexual habits, and so on.").

69. There exists a line of regulations focusing on types of information as requiring 'special' consideration including content versus metadata; medical; sensitive; financial, or intimate information. JULIE C. INNESS, PRIVACY, INTIMACY, AND ISOLATION (1996); Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. (2015); Christopher Slobogin, *Cause to Believe What? The Importance of Defining a Search's Object—Or, How the ABA Would Analyze the NSA Metadata Surveillance Program*, 66 OKLA. L. REV. 725 (2014).

70. Kugler & Strahilevitz, *supra* note 1, at 245–46.

71. *Id.*

72. *Id.* at 246.

73. *Id.* at 212.

GPS tracking device, or through cell phone towers.[74] Bernard Chao also compared reasonable expectations of privacy in different scenarios and observed that the highest proportion of respondents found the placement of a GPS device on a car for a duration of eighteen days to be a violation of reasonable expectations of privacy, as compared to seventeen other scenarios.[75] The question centered on the degree of intrusion of a government actor.[76] Similarly, Marc McAllister surveyed respondents with a series of questions involving location tracking through GPS devices versus cell phone tracking to compare the appropriateness of law enforcement surveillance as dependent on the seriousness of the crime.[77]

Outside law enforcement as the collecting agent, Jennifer Urban, Chris Hoofnagle, and Su Li found that "Americans overwhelmingly consider information stored on their phones to be private, and strongly reject systems that would rely on collecting and using contact data from their phones or tracking their locations."[78] They found that 92% of respondents do not think their location data should be used for ads, and 46% say location should not be kept at all, even by cell phone companies.[79] Finally, Kirsten Martin and Katie Shilton compared location data to other data used for advertising, and found that the collection and use of location data for advertising negatively impacts privacy expectations in the mobile context, especially for high-use users.[80]

A series of studies has measured consumer behavior directly around location data to inform the tech industry. Eran Toch et al. employed a location

---

74. Alisa Smith et al., *An Empirical Examination of Societal Expectations of Privacy in the Digital Age of GPS, Cell Phone Towers, & Drones*, 26 ALB. L.J. SCI. & TECH. 26 111, 133–35 (2016). While Chao et al. dismiss these findings as not representative enough (Smith et al. have 54% women and 25% African American respondents), their own re-weighting in Chao et al. did not impact their results. Bernard Chao et al., *Why Courts Fail to Protect Privacy: Race, Age, Bias, and Technology*, 106 CALIF. L. REV. 263, 294, 297 (2018).

75. Chao et al., *supra* note 74, at 308–09. Chao's examination of other forms of surveillance did not include duration. Among the seventeen other scenarios, accessing data stored in the cloud was second-highest, email was fifth, and roadblock was lowest. *Id.*

76. *Id.* at 303.

77. Marc McAllister, *GPS and Cell Phone Tracking: A Constitutional and Empirical Analysis*, 82 U. CINCINNATI L. REV. 207, 212 (2013). Kugler and Strahilevitz rightly identify the methodological issues and open questions of McAllister's work, including no explanation of the sample. Kugler & Strahilevitz, *supra* note 1, at 223 n.113.

78. Jennifer M. Urban et al., *Mobile Phones and Privacy*, BERKELEY CTR. L. & TECH. 6 (2012), https://www.ftc.gov/system/files/documents/public_comments/2013/12/00007-89101.pdf [https://perma.cc/A5HL-E7DC].

79. *Id.* at 19, 20.

80. Kirsten Martin & Katie Shilton, *Why Experience Matters to Privacy: How Context-Based Experience Moderates Consumer Privacy Expectations for Mobile Applications*, 67 J. ASS'N INFO. SCI. & TECH. 1871, 1877–80 (2016).

sharing system to examine the actual behavior of study participants. The authors found that users were more willing to share location data when their location was frequented by a large and diverse set of people, thus suggesting a preference for areas where their identity would be obscured by others.[81] Michael Benisch et al. conducted a user study to measure when and where users would be willing to share their location data.[82] The authors found that day, time, and exact location are the significant factors driving users' willingness to share information rather than user activity, identity, or general concern as found in previous studies.[83] These findings suggest that users are quite nuanced about when and where they are willing to share their location data.[84]

There are three important gaps in the existing literature. First, location data has been operationalized in empirical studies as GPS without any explanation as to the types of inferences drawn about the user or the meaning of location data. Second, the majority of surveys have focused on law enforcement as the collecting actor, though the majority of location data collectors are actually private actors. Finally, the user studies have suggested that individuals have specific privacy expectations about how, when, and where location data should be gathered. Our study seeks to extend this important work on privacy by focusing on a diverse set of collecting actors and measuring the normative judgment of the respondents when the inferences drawn from location data are clear.

---

81. Eran Toch et al., *Empirical Models of Privacy in Location Sharing*, UBICOMP '10 PROC. 12TH ACM INT'L CONF. ON UBIQUITOUS COMPUTING, 129–138 (2010).

82. Michael Benisch et al., *Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs*, 15 PERS. & UBIQUITOUS COMPUTING 679, 679 (2011).

83. *Id.*

84. *Id.*; *see also* Adrienne Porter Felt et al., *I've Got 99 Problems, but Vibration Ain't One: A Survey of Smartphone Users' Concerns*, PROC. SECOND ACM WORKSHOP ON SECURITY & PRIVACY IN SMARTPHONES & MOBILE DEVICES 33 (2012) (finding that respondents differentiated their privacy expectations around location data based on who was receiving it; they were more concerned when friends, advertisers, or the public received it than when the server received it); Irwin Reyes et al., *supra* note 26, at 69–70 (discussing how apps targeted at children collected location data without consent); Primal Wijesekera et al., *Android Permissions Remystified: A Field Study on Contextual Integrity*, PROC. 24TH USENIX SECURITY SYMP. 499, 508 (Aug. 12–14, 2015) (discussing situations where respondents did not find requests for location data from apps to be appropriate).

## III. STUDY DESIGN

### A. CONTEXTUAL INTEGRITY

Our earlier work challenging the role of the private-public dichotomy revealed previously ignored factors that systematically affect people's privacy expectations. We called these confounding variables because they explained some of the inconsistencies between what people say and what they do, which commentators commonly—mistakenly in our view—call a "paradox."[85] This work was guided by the theory of contextual integrity (CI), which pointed to variables that both refined and confounded the blunt categories of public and private. We have taken a similar approach in the present set of studies, in which we demonstrate that people's judgments about appropriate flows (in other words, their expectations) of location data are far more nuanced, in systematic ways, than the dichotomy would predict. Focusing solely on locations traditionally conceived as public, our study is able to hone in on what location means to respondents and the contextual parameters systematically affecting their judgments about location tracking and location data capture. Before proceeding, we offer a brief overview of CI, how it has guided our studies, and how, for pragmatic reasons, we have simplified it.

According to the theory of CI, whether privacy has been preserved or violated depends on whether a given flow of information (or data) is *appropriate*, which in turn depends on whether this flow conforms with entrenched and contextual informational norms (sometimes abbreviated as "privacy norms").[86] When flows conform with entrenched norms, we say CI, prima facie, is respected. Otherwise, a further analysis is required in order to establish whether norms that have been contravened should override a practice under consideration or vice versa.

To establish conformance, a CI analysis needs to map actual flows against privacy norms (or expectations). Fully specifying a privacy norm requires specifying five key parameters: information type (about what), subject (about whom), sender (by whom), recipient (to whom), and transmission principle (flow under what conditions). Thus, when describing a given flow for purposes of evaluating its appropriateness, one needs to provide values for all five

85. Martin & Nissenbaum, *Measuring Privacy*, *supra* note 7, at 218; Martin & Nissenbaum, *Privacy Interests in Public Records*, *supra* note 7.

86. *Id.*; *see also* HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE (2010). The part of CI theory that defines a series of steps to establish whether norms should prevail over conflicting practices, or vice versa, is concerned with moral legitimacy of norms or practices, respectively. Although answering questions about legitimacy is a defining component of CI theory itself, we set them aside for purposes of the current study.

parameters, or risk ambiguity resulting from missing variables.[87] An analysis that takes the public-private dichotomy as determinative would assert that reactions to flows of location data could be predicted solely on the basis of whether the location in question is public or private. By contrast, a CI analysis predicts a complex dependency between privacy expectations on the one hand, and the values for all five parameters on the other.

This thesis fundamentally informs the design of our studies. It also contrasts CI with some of the work discussed in Part II, notably efforts to decide cases or regulate data practices with reference to one factor alone (for example, the actor collecting information, the type of information, or the mode of collection) without recognizing that these factors interact. Although technical innovation has posed persistent challenges to institutional norms and structures, we ascribe painfully slow progress in coping with technology-induced privacy threats to an equally persistent failure to grapple with the interdependencies among key contextual factors. The studies reported in this Article (and the two previous articles), attempt to bring these interdependencies to light in the intersecting domains of law, policy, and technology.

Before describing our methodology and the studies themselves, two further points. First, we have not yet addressed the "context" in contextual integrity. The most we can say here, avoiding a long digression, is that context is roughly equivalent to social domain or sphere as theorized in social and political theory and reflected in the organization of societies (e.g., healthcare, family, commerce, finance, politics, etc.). Such domains are also frequently reflected in areas of law such as commercial law, family law, and constitutional law. Contexts in this sense are constituted by respective roles, activities, purposes, values, and norms. Among the norms, those governing information flows are associated with respective contexts in their characteristic ontologies, such as those defining contextual roles or capacities of actors (e.g., student, physician, senator, rabbi, etc.), and types or categories of information (e.g., diagnosis, blood type, vote, grades, marital status, criminal record, etc.). Accordingly, the scenarios we present to study respondents include values for parameters that are clearly associated with particular, familiar contexts (e.g., government, healthcare, etc.).

The second point is a caveat. Ideally, CI would require that the scenarios presented to respondents include the five parameters, with simultaneous variation of values for them. The reality of limited resources, time, human subjects, and requirements of statistical analysis has necessarily required

87. Martin & Nissenbaum, *Privacy Interests in Public Records*, *supra* note 7, at 123; Martin & Nissenbaum, *Measuring Privacy*, *supra* note 7, at 198.

pragmatic simplifications. These decisions were made with careful forethought and a disciplined effort not to claim more than the results allow.

## B. METHODOLOGY

Our study comprises three key parts: (i) a set of pilot surveys which informed the design of the main study; (ii) a main study with a nationally representative sample to shed light on the attributes of location tracking instances that are systematically related to assessments of appropriateness of information flow (how "okay"); and (iii) a follow-up survey to assess the significance of location semantics relating how respondents' understanding (conceptions) of location affect their judgments of the appropriateness of location tracking.

**Table 1: Overview of Studies**

| Study | Sample | Goal |
|---|---|---|
| Pilot Study 1 | Amazon Turk<br><br>N = 1,200 | Measure the impact of (1) the ordering of the control questions, (2) the voice of the vignettes, and (3) two parameters of the factorial vignette: (a) the precision of the location data described, and (b) the significance of frequency of tracking. |
| Main Study | Knowledge Networks<br><br>N = 1,500 | Understand what attributes of information flow are important to respecting contextual integrity in a public space.<br>Survey 1. Actor, Source, Place<br>Survey 2. Actor, Source, Place, Duration<br>Survey 3. Actor, Source, Place, Duration, Inference. |
| Follow-Up Study | Amazon Turk<br><br>N = 300 | Explore how giving meaning to location data (including the place as understood from the location data) impacts consumers' judgment. |

In what follows, we outline general methods and our selection of respondent control ratings. To settle further design issues, we ran a pilot study which informed the factors we chose to include in subsequent surveys, the voice of the vignettes (2nd versus 3rd person), and the question order.

*1. Factorial Vignette Survey*

The method we used for our studies is known as the factorial vignette methodology.[88] Factorial vignette surveys present respondents with a series of vignettes in which multiple factors are systematically varied in order to test their relevance to respondents' assessments. These factors thus constitute the independent variables of our study. The variables chosen for our study correspond to a subset of the contextual factors (or parameters) of CI. For each vignette, values for the parameters are systematically and simultaneously varied. After seeing each vignette, respondents are asked to complete a simple rating task—the degree to which a scenario is appropriate or "okay"—from which we extract the statistical relevance of each of the factors.

The factorial vignette methodology has proven effective for addressing normative research questions which are notoriously difficult to study.[89] Because of the need to respond to several simultaneous contextual factors in the vignette, respondents are less likely to fall victim to two types of respondent bias. First, respondents may adjust answers in order to appear ethical or concerned in a traditional survey and are less likely to do so when many factors are changing simultaneously. This is particularly useful for privacy, which, according to skeptics, people claim to value while their behaviors communicate otherwise.[90] Second, respondents may have difficulty identifying and articulating the reasons behind their judgments, and the factorial vignette survey methodology supports the researcher in analyzing which factors moved the respondent's rating of the vignette without directly asking the respondent for a prioritized list of what is important to them in judging the vignette.[91]

---

88. Guillermina Jasso, *Factorial Survey Methods for Studying Beliefs and Judgments*, 34 SOC. METHODS & RES. 334, 342 (2006); Steven Nock & Thomas Guterbock, *Survey Experiments*, *in* HANDBOOK OF SURVEY RESEARCH (Peter V. Marsden & James D. Wright eds., 2010).

89. *See, e.g.*, Jasso, *supra* note 88.

90. This is sometimes (mistakenly) referred to as the privacy paradox, where individuals are criticized for stating in surveys that they care about privacy while also sharing their data with companies. However, individuals are shown to not realize how their data is being tracked, shared, and used after disclosure, thereby rendering their behavior more closely aligned with their stated preferences. Individuals believe their privacy expectations are respected online and are shown to penalize companies when privacy expectations are violated. *See* Kirsten Martin, *Breaking the Privacy Paradox*, 32 BUS. ETHICS Q. 1 (forthcoming 2019); Kirsten Martin, *Privacy Notices as Tabula Rasa: An Empirical Investigation into How Complying with a Privacy Notice Is Related to Meeting Privacy Expectations Online*, 34 J. PUB. POL'Y & MARKETING 210, 220 (2015); Kirsten Martin, *The Penalty for Privacy Violations: How Privacy Violations Impact Trust Online*, 82 J. BUS. RES. 103, 110 (2018).

91. Martin & Nissenbaum, *Measuring Privacy*, *supra* note 7, at 195.

For our studies, vignettes described a scenario involving the collection, flow, or use of location data in public spaces, which respondents were asked to evaluate. Each respondent was presented with twenty to thirty vignettes, depending on the study. The survey instrument generates vignettes in real time by varying values randomly for each factor.

We asked respondents to rate the degree to which the vignette was "okay." Choosing this language is part of our ongoing effort to elicit a sense of what is expected and what is normative. Although other studies of privacy might reasonably want to learn what people prefer, in taking guidance from CI, we strive to learn about people's perception of norms. Nevertheless, more work is needed in defining an approach that encourages respondents to cast an objective eye.

*2. Respondent Controls*

Outside the vignettes, we also captured respondent-level controls based on previous privacy studies.[92] As before, we were interested in controlling for individual-level differences when the respondents answered a series of vignettes. Respondent-level beliefs and attributes that we selected (and discuss below) have all been shown to correlate with judgments about privacy and trust.

a) Privacy and Trust

Privacy has been examined as impacting trust in prior studies and respondents' general trust disposition has been found to impact their privacy concerns.[93] We captured the respondents' disposition to trust by asking them to rate, on a scale from "strongly disagree" to "strongly agree," their agreement with the statement: "In general, I trust people until proven otherwise." We also captured the respondents' institutional trust in government and business with the degree they agreed with, "In general, I trust the federal government," and, "In general, I trust business." Finally, we asked respondents to evaluate the statement, "In general, I believe privacy is important."

92. *See id.*; Martin & Nissenbaum, *Privacy Interests in Public Records*, *supra* note 7.

93. Kirsten Martin, *The Penalty for Privacy Violations*, *supra* note 90, at 104. For a comparison of Westin's privacy concern measurement to actual privacy expectations as well as individual's trust disposition, see Martin & Nissenbaum, *Measuring Privacy*, *supra* note 7 (finding that respondents rated as low on Westin's privacy concern measurement believed privacy to be important but trusted the firms and, therefore, had low concerns; and finding that Westin's privacy concern measurement was not significant in driving specific judgments about privacy expectations).

b) Authoritarianism

In previous scholarship examining the privacy interests in public space and the privacy expectations around being tracked in public, Kugler and Strahilevitz found that respondents' affinity for authoritarianism impacted their expectations of privacy in regards to being tracked by the government in public.[94] To test the respondents' affinity for authoritarianism, an authoritarianism score was created from two questions based on existing scholarship: (a) "It's great that many young people today are prepared to defy authority" (reverse coded), and (b) "What our country needs most is discipline, with everyone following our leaders in unity."

*3. Analyzing Respondent-Level Variables*

Each control variable was captured using a slider with a scale of Strongly Disagree (-100) to Strongly Agree (+100). To standardize the responses, a new variable was created and assigned to each respondent as to what quartile their rating corresponded to (top 25%, bottom 25%, etc. of all ratings). This analysis was performed for each respondent control and used in the multi-level regressions as well as for splitting the sample when necessary.

**Table 2: General Format of Surveys**

| Q # | Concept | Prompt |
|---|---|---|
| 1 | Trust in Business | In general, I trust business. |
| 2–31 | Vignettes (1 of 3 possible) | Please rate the degree to which this situation is okay, from Definitely Not Okay to Definitely Okay. |
| Respondent Controls: *How much do you agree or disagree with the following statement:* | | |
| 32 | Privacy Important | In general, I find privacy important. |
| 33 | Trust in Government | In general, I trust the federal government. |
| 34 | RevAuthoritarianism 1 | It's great that many young people today are prepared to defy authority. |
| 35 | Authoritarianism2 | What our country needs most is discipline, with everyone following our leaders in unity. |
| 36 | Trust Disposition | In general, I give people the benefit of the doubt until shown otherwise. |

94. Kugler & Strahilevitz, *supra* note 1, at 254–55.

## IV. PILOT STUDY

### A. PILOT DESIGN

In order to study what location data means to individuals, we needed to make decisions about terminology and study design. To this end, we ran a pilot study to test four facets of the survey design: (1) the ordering of control questions, (2) the voice of the factorial vignettes, and (3) two of the vignettes' parameters: (a) the importance of precision when presenting location data, and (b) significance of tracking frequency. Results of this pilot study, which were used to design the main surveys, are briefly described. A full description and analysis are provided in the Appendix.

### B. PILOT RESULTS

1. **Ordering of Controls and Vignettes**. Did placing the controls before or after the vignettes matter to (i) the rating of the vignette or (ii) the respondents' ratings of the controls? To ensure the ordering did not impact the vignette ratings, we ran the pilot survey with the respondent controls both before and after asking the respondents to rate the vignettes. The average vignette rating did not change when the control questions were asked before versus after the vignettes. The average rating remained about -36 ("Not Okay"). Interestingly, the ratings for certain control variables did change when the controls were asked after the vignettes, as shown in Table A2 in the Appendix.

   Specifically,

   - The Authoritarian score decreased from -13.32 to -20.58 when the question is asked after the vignettes. In other words, the respondents are less authoritarian after rating scenarios about commercial and governmental tracking.
   - The average trust in business rating also decreases from -12.12 to -25.95 when the question is asked after the vignettes are rated. This is consistent with previous work on trust and privacy: respondents' institutional trust in business in general is diminished when the gathering and use of data is just explained in vignettes.[95]

2. **Vignette Voice** ("you" versus "a person"). We tested if the 'voice' of the vignette mattered to the judgment of whether the information flow was appropriate. The voice of a second person, third person, or third person plural impacted the privacy judgments of the respondents, as

95. Kirsten Martin, *Do Privacy Notices Matter? Comparing the Impact of Violating Formal Privacy Notices and Informal Privacy Norms on Consumer Trust Online*, 45 J. LEGAL STUD. 191, 206 (2016).

has been suggested before.[96] Voice did make a difference. When the vignettes included a reference to the respondent ("you"), the vignettes were rated less "okay" (-35.32) compared to a third-person voice (-27.05) or a third-person plural voice (-30.45). We decided to use third person voice for the live survey.

3. **Location Precision**. This may have been the most surprising finding. We were interested whether precision mattered, ranging from GPS (most precise) to location, street address, and city. The results suggest that the word "location" meant the same to respondents as "GPS" in judging the scenario as appropriate, with no significant difference between the two levels ($p=0.95$). And even where the precision decreased, such as street address (+5.69) and city (+8.19), the degree of difference was only slightly over GPS and generic location. ($p < 0.00$).
4. **Storage versus Frequency of Data Collection**. In order to test if the frequency of the data collection or its storage duration affected subjects' responses, we included both factors in the vignette. The length of storage time was found to be inversely related to how "okay" the vignette was judged, as indicated by the steep negative slope in Figure 2 in the Appendix. Frequency, by contrast, was not significant; respondents did not rate the vignette any differently as the frequency levels varied.[97]

### C. DISCUSSION OF PILOT STUDY

The results of the pilot study were surprising and essential in guiding aspects of the design of our main study. In sum:

1. We used the term "location" in the later studies, knowing that the term is equivalent to "GPS" for the respondent;
2. We dropped the use of frequency;
3. We shifted to the term "duration" for the duration of tracked location information;
4. We used the third-person plural in the later vignettes and asked the control questions after the vignettes in order to break up the control questions.

---

96. *See* Slobogin & Schumacher, *supra* note 10, at 736.

97. Because this result was somewhat surprising, we ran another vignette survey without storage included as a factor to allow the respondent to focus on frequency (from every five seconds to once per day). However, frequency was still not significant; the only difference was the average vignette rating decreased from -35.52 to -31.57 when storage was removed as a factor.

We discuss these decisions further in the Appendix.

## V. MAIN STUDY

Having settled some of these design issues, the purpose of the main vignette study is to identify what contextual factors are important to respondents' judgments of whether location data collected in public were appropriate. The study focused on the factors described below and shown in Table 3:

- Transmission Principles
  - Source: How the location data is gathered (phone signal, mapping app, license-plate reader, etc.)
  - Duration: How long the location data is gathered (from a few minutes to a year)
- Actors: Recipients of the location data (FBI, family, your employer, etc.)
- Attributes: What information can be inferred from the location data (who your friends are, how regularly you vote, etc.)[98]

### A. MAIN STUDY DESIGN

#### *1. Vignette Factors*

a. Source. How the location data is gathered and transmitted has been found to be important.[99] The source of collecting the location data varied across license plate readers, CCTV, phone tracking, social media, or a mapping application. Since sources affect the conditions or constraints of flow from subject to recipient, we took these to be operationalizations of Transmission Principles, as defined in CI.

98. Based on the design pilot, we used third-person voice in each scenario with the word location, which is equivalent to the term "GPS coordinates" for the respondents, given the pilot study described above. We had the respondents rate the vignettes before answering the control questions. We used duration rather than storage or frequency. *See infra* Appendix A.

99. *See generally* Surden, *supra* note 58; Luciano Floridi, *Network Ethics: Information and Business Ethics in a Networked Society*, 90 J. BUS. ETHICS 649 (2009); Kirsten Martin, *TMI (Too Much Information)*, 30 BUS. & PROF. ETHICS J. 1 (2011); Woodrow Hartzog & Evan Selinger, *Surveillance as Loss of Obscurity*, 72 WASH. & LEE L. REV. 1343 (2015); Lawrence Lessig, *The Architecture of Privacy*, 1 VAND. J. ENT. L. & PRAC. 56 (1999).

**Table 3: Vignette Factors Included in National Study**

| **Concept** | Description | As operationalized in Vignette |
|---|---|---|
| **Duration** | How long you are tracked | A year, about six months, a month, a few days, a few minutes |
| **Actor** | Government | A city emergency service (ambulance, fire) |
| | Federal government | The FBI |
| | Employer | Their employer |
| | Commercial data aggregator | A commercial data broker |
| | Commercial | A commercial location-based service (e.g., Yelp) |
| | Family | A family member (e.g., parents, spouse, or sibling) |
| **Source** | License-plate reader | License-plate readers |
| | CCTV | CCTV cameras with facial recognition |
| | Phone | The signal from a mobile phone |
| | Fit Bit | A fitness app (e.g., FitBit or Strava) |
| | Social media | Geo-tagged posts on social media (e.g., Twitter, Facebook, Instagram) |
| | Mapping app | A mapping app (e.g., Google Maps) |
| **Additional Factor** | | |
| | **Place** | **Inferences about individual (Survey 3 Only)** |
| Association | A restaurant or cafe | Who their friends are |
| Protests/rallies | The National Mall | Whether they are active in their political beliefs in attending protests |
| Sin Shopping | A liquor store | Whether they have a drinking problem |
| Shopping | A shoe store | How susceptible they are to shoe ads |
| Home | Home | How often they spend the night away from home |
| Work | Work | Whether they are dedicated workers |
| Medical | A medical clinic | Whether they have a chronic illness |
| Voting | A voting site | How regularly they vote |

b. Duration of Collection. Previous work has found that the duration of data collection can affect privacy expectations.[100] We had the data collation range from a period of a few minutes to a year.

c. Actors. In order to capture both government and commercial actors as well as different purposes of the data collection, the values of the actor (recipient) parameters ranged over FBI, a city planner, a commercial data broker, a location-based commercial service, and family members.

d. Place and Inferences. We added this factor into Survey 2, which is explained below in order to understand the extent to which "bare" location was a stand-in or proxy for other qualitative locational information. Inferred information included a person's associates, whether attending a protest, voting behavior, routine travel, whether frequenting a store, and whether frequenting a medical facility, in addition, simply, to where a person is. This tests whether the attribute of type of information inferred about a person drives expectations surrounding location information.

*2. Vignette Template and Example for Main Study*

The factors in Table 3 are used within a vignette template as described below. A specific level within each factor is randomly assigned as the vignette is generated for the respondent. Below the example vignettes for all three surveys are provided, as well as the general template for each.

a. Survey 1 Template Baseline

{Actor} acquires location data from {Source} and uses this data to figure out if a person was at {Place}.

b. Survey 1 Examples

A city emergency service (ambulance, fire) acquires location data from license-plate readers and uses the data to figure out if a person was at a shoe store.

100. *See, e.g.*, United States v. Jones, 565 U.S. 400 (2012) (finding that the data was collected for twenty-eight days); Shafer v. City of Boulder, 896 F. Supp. 2d 915 (D. Nev. 2012) (finding data was collected for two months); Carpenter v. United States, 138 S. Ct. 2206 (2018) (finding that data was collected for about 127 days); County of Riverside v. McLaughlin, 500 U.S. 44 (1991) (finding that data was collected for twenty-four hours).

A city emergency service (ambulance, fire) acquires location data from a mapping app (e.g., Google Maps) and uses the data to figure out if a person was at a restaurant.

A commercial location-based service (e.g., Yelp) acquires location data from license-plate readers and uses the data to figure out if a person was at a restaurant.

An employer acquires location data from geo-tagged posts on social media (e.g., Twitter, Facebook, or Instagram) and uses the data to figure out if a person was at home.

c. Survey 2 Template – Adding Duration

{Actor} acquires location data from {Source} **for a period of {Duration}** and uses this data to figure out if a person was at {Place}.

d. Survey 2 Examples

A family member (e.g., parents, spouse, or sibling) acquires location data from license-plate readers for a period of a year and uses the data to figure out if a person was at the National Mall.

An employer acquires location data from a mapping app (e.g., Google Maps) for a period of a few minutes and uses the data to figure out if a person was at a shoe store.

A commercial data aggregator acquires location data from license-plate readers for a period of a week or so and uses the data to figure out if a person was at a shoe store.

e. Survey 3 Template – Adding Inference

{Actor} acquires location data from {Source} for a period of {Duration} and uses this data to figure out if a person was at {Place} **and {Inference}**.

f. Survey 3 Examples

An employer acquires location data from a mapping app (e.g., Google Maps) and uses the data to figure out if a person was at a liquor store and whether they have a drinking problem.

A city emergency service (ambulance, fire) acquires location data from geo-tagged posts on social media (e.g., Twitter, Facebook,

Instagram) and uses the data to figure out if a person was at a liquor store and whether they have a drinking problem.

A commercial data aggregator acquires location data from a fitness app (e.g., FitBit or Stava) and uses the data to figure out if a person was at a shoe store and how susceptible they are to shoe ads.

*3. Vignette Rating Task*

For each vignette, respondents were instructed to indicate the degree to which they agreed with the question "Is this okay?" with a slider. The left side of the slider indicated "Definitely Not Okay" and the right of the slider indicated "Definitely Okay." The slider was on a scale of -100 to +100 with the number suppressed so the respondents saw only the labels "Okay" and "Not Okay."

*4. Sample*

In our previous studies, we utilized Amazon's Mechanical Turk, which has become an accepted platform for empirical research such as this. Amazon Mechanical Turk offers a platform for researchers to post surveys (HITs) and respondents or workers to perform HITs they find worthwhile or interesting. Mindful of questions around this choice (not only aimed at our work), for our main survey, we deployed KnowledgeNetworks, an online research panel representative of the entire U.S. population. Approximately 1,500 respondents took one of three possible vignette surveys. KnowledgeNetworks panel members are randomly recruited through probability-based sampling. Households are provided with access to the internet and hardware if needed.[101] Importantly, Amazon Mechanical Turk provided higher quality sample with the same theoretical generalizability as KnowledgeNetworks.

101. For an overview of the KnowledgeNetworks sampling methodology and a comparison to the pilot tests on Turk, see *infra* Appendix C.

**Table 4: Sample Statistics for Surveys 1–3**

| | Survey 1 | Survey 2 | Survey 3 |
|---|---|---|---|
| | Base | + Duration | +Inferred Info |
| Authoritarian Scale | 5.57 | 6.83 | 2.76 |
| Trust Scale | 5.44 | 5.94 | 457 |
| Female | 49% | 53% | 50% |
| Age | 50.1 | 49.5 | 49.4 |
| Privacy Important | 72.32 | 70.97 | 72.14 |
| Trust Government | -23.08 | -22.93 | -27.67 |
| Trust Business | 2.85 | 2.53 | 4.10 |
| DV Mean | -28.66 | -35.96 | -46.16 |
| N (Respondents) | 480 | 483 | 435 |

The sample was analyzed for unresponsive respondents. Since the respondents each rated thirty independently generated vignettes, the pattern of their rating on a sliding scale of -100 to +100 for each vignette could be analyzed as possibly unresponsive. We marked two types of surveys as nonresponsive: those that rated over twenty of the thirty vignettes as "0" (never moved the slider) and those that rated over twenty-five vignettes at one of the end points (moved the slider to the left or the right almost every time). For the KnowledgeNetworks sample, this resulted in 10% of Survey 1 respondents, 13% of Survey 2 respondents, and 16% of Survey 3 respondents being removed from the pool. The number of respondents listed in Table 4 above does not include those respondents removed from the analysis.[102]

102. Appendix C includes a comparison of the KnowledgeNetworks sample with the sample from running the same surveys on Amazon Mechanical Turk. The number of respondents discarded from non-responsive ratings was less for Turk than the national sample from KnowledgeNetworks. For the Turk sample, 2% of Survey 1 respondents, 5% of Survey 2 respondents, and 11% of Survey 3 respondents were found to be unresponsive. In comparison for the KnowledgeNetworks sample, 10% of Survey 1 respondents, 13% of Survey 2 respondents, and 16% of Survey 3 were unresponsive and removed from the sample. The Turk sample was higher quality than the KnowledgeNetworks sample with the same theoretically generalizable findings.

**Table 5: Main Regression of Okay Rating on Vignette Factors and Respondent Controls**

| | Survey 1 BASE | | Survey 2 DURATION | | Survey 3 INFERENCE | |
|---|---|---|---|---|---|---|
| | Coef | p | Coef | p | Coef | p |
| FedGovtActor | 45.55 | 0.00 | 33.22 | 0.00 | 10.08 | 0.00 |
| DataAggregatorActor | 0.77 | 0.61 | 0.03 | 0.98 | -0.88 | 0.48 |
| FamilyActor | 33.70 | 0.00 | 23.04 | 0.00 | 19.08 | 0.00 |
| EmployerActor | -3.00 | 0.05 | -7.52 | 0.00 | -1.28 | 0.31 |
| CityServicesActor | 42.39 | 0.00 | 16.74 | 0.00 | 4.42 | 0.00 |
| *(null = Commercial Actor)* | | | | | | |
| MappingAppSource | -1.53 | 0.31 | -3.97 | 0.00 | -1.01 | 0.41 |
| PhoneSource | -0.19 | 0.90 | -7.26 | 0.00 | -0.73 | 0.56 |
| LPRSource | -6.47 | 0.00 | -7.40 | 0.00 | -3.38 | 0.01 |
| CCTVSource | -4.01 | 0.01 | -5.11 | 0.00 | -2.84 | 0.03 |
| FitBitSource | -7.90 | 0.00 | -10.35 | 0.00 | -3.92 | 0.00 |
| *(null = Social Media Source)* | | | | | | |
| MedicalPlace | 0.43 | 0.80 | -1.01 | 0.52 | 0.58 | 0.68 |
| RalliesPlace | 3.34 | 0.06 | 1.45 | 0.36 | -6.10 | 0.00 |
| ShoppingPlace | -2.57 | 0.14 | -4.39 | 0.01 | -0.22 | 0.88 |
| VotingPlace | -13.85 | 0.00 | -12.74 | 0.00 | -7.22 | 0.00 |
| SinShoppingPlace | -3.09 | 0.08 | -6.38 | 0.00 | 1.11 | 0.45 |
| HomePlace | 5.84 | 0.00 | -0.82 | 0.61 | -1.44 | 0.33 |
| WorkPlace | 4.78 | 0.01 | 0.84 | 0.60 | 2.07 | 0.16 |
| *(null= Restaurant)* | | | | | | |
| DurationScale | n/a | n/a | -1.64 | 0.00 | -0.49 | 0.06 |
| PrivacyImport | -0.27 | 0.00 | -0.38 | 0.00 | -0.26 | 0.00 |
| HighAuthoritarianism | 4.28 | 0.30 | 4.67 | 0.23 | 2.94 | 0.52 |
| TrustScale | 0.42 | 0.00 | 0.44 | 0.00 | 0.23 | 0.00 |
| _cons | -32.04 | 0.00 | -12.37 | 0.00 | -31.14 | 0.00 |
| N | 480 | | 483 | | 435 | |
| Vignettes | 14,400 | | 14,490 | | 13,050 | |
| DV Mean | -28.66 | | -35.96 | | -46.16 | |
| ICC | 32.7% | | 35.4% | | 45.2% | |
| ICC Null | 33.6% | | 39.1% | | 48.0% | |

## B. MAIN STUDY RESULTS

To analyze which vignette factors are significant to respondents' judgments about the appropriateness of the gathering and use of location data, we regressed the dependent variable—the rating that the collection of location data in the given vignette was "Okay"—on the vignette factors and the respondent controls. The results are in Table 5. The factor with the most impact on the rating task is the actor collecting the location data; changing who gathers the information had the largest impact on the rating that the gathering of location data was "Okay." Below, each vignette factor—actor, source, and inference—is analyzed.

For the respondent controls, we found that authoritarianism was not significant to the rating task compared with the general trust scale, which was: the greater the respondents' trust in general (a composite of dispositional trust, trust in business, and trust in government), the more appropriate the respondent judged the collection of location data overall.

### *1. Significance of Vignette Factors*

#### a) Actors

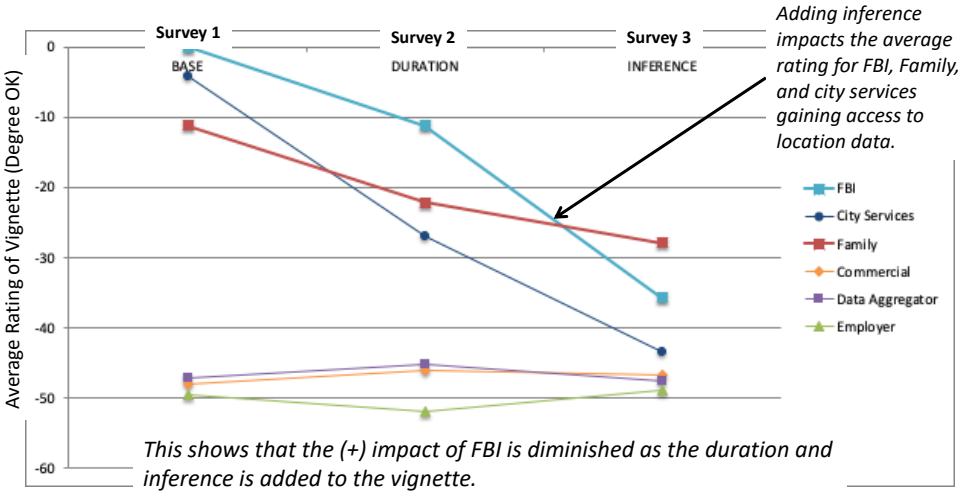
The actor acquiring the location data significantly affected respondents' judgments. As shown in the regression results in Figure 1, for each actor—FBI, commercial location-based service, city planner, and data—it was significantly less appropriate than the null condition for a family member to acquire location data.

Figure 1 also shows that adding inference impacts the average rating for FBI, family, and city services gaining access to location data. Figure 1 further reflects a significantly more negative rating of FBI, city services, and family collecting location data when the vignettes reveal the duration of the collection (Survey 2), and the nature of what the actor can infer about the individual (Survey 3). Even initially, the positive glow surrounding the FBI is extinguished when duration and inference are included in vignettes.

Table 5, with the main regression results, shows that initial differences in ratings between the FBI or city services versus a commercial entity (e.g., Yelp) diminishes as duration and inference are included. The FBI is favored above a commercial actor a when place only is included (+47; $Ave_{FBI}$ = -0.04, $Ave_{Bus}$ = -47.14). But when duration is added (+35; $Ave_{FBI}$ = -11.20, $Ave_{Bus}$ = -46.09) and inferences are drawn, the difference is diminished (+11; $Ave_{FBI}$ = -35.67, $Ave_{Bus}$ = -46.80). While the collection and use of location data by commercial actors such as Yelp or data aggregators is consistently not "okay," the appropriateness of the FBI collecting location data is negatively impacted by

the mere mention of duration and the mere mention of the inferences drawn about the individual surveilled.

**Figure 1: Average Vignette Rating by Actor**

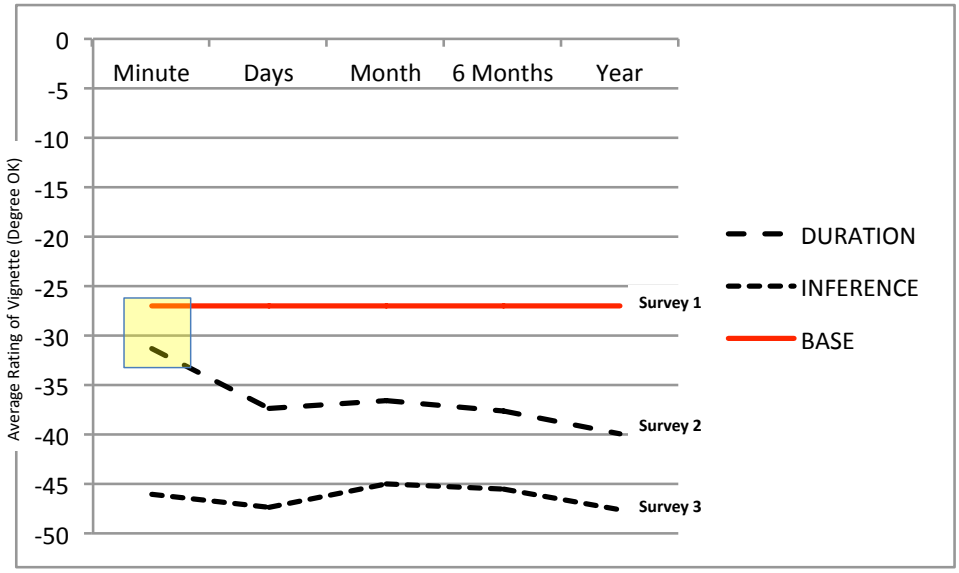



b) Duration

The duration of the tracking of location data was significant with -5.16 lower vignette rating (less "okay") for each incremental step in additional time of tracking as shown in Figure 2. The impact of duration is lessened, (i.e., the slope is shallower) in Figure 2, for Survey 3 where the inferred information is also included.

Importantly, respondents appear to assume the shortest duration when no duration is included in the vignette, as in the base scenario in Survey 1. The average rating for a vignette with the duration set to "a few minutes" is the same as the baseline when no duration mentioned (see the yellow box in Figure 2).

**Figure 2: Average Vignette Rating by Duration Period**

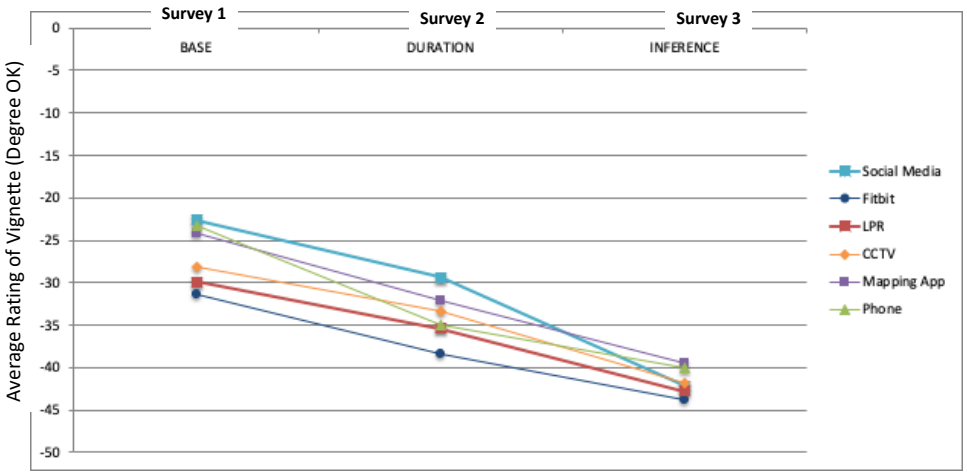



c) Source

How the location data was gathered—through a social media post, a mapping app, a license-plate reader, a phone, or a CCTV with facial recognition—affected the degree to which the vignette was rated "okay," as is shown in Figure 5. Capturing location through a phone or CCTV was rated the lowest, or least "okay"; capturing through social media and a mapping app was the highest rated source (although still negative).[103] In other words, respondents did not significantly differentiate across the different sources of gathering location data, particularly in comparison to the importance of who receives the information. This is shown by how the average rating is actually clustered for each survey across the sources and is also evident in Table 5 above in the general regression, where the coefficients are significantly different at times across types of sources, but not large (e.g., the difference between gathering location data via a phone versus a social networking app is -7.26 in Survey 2 (out of a 200-point scale) and not significant for Surveys 1 and 3.

103. Given the attention to the collection of location data from phones, the difference in respondents' ratings across sources is significant but not a main driver of the appropriateness rating.

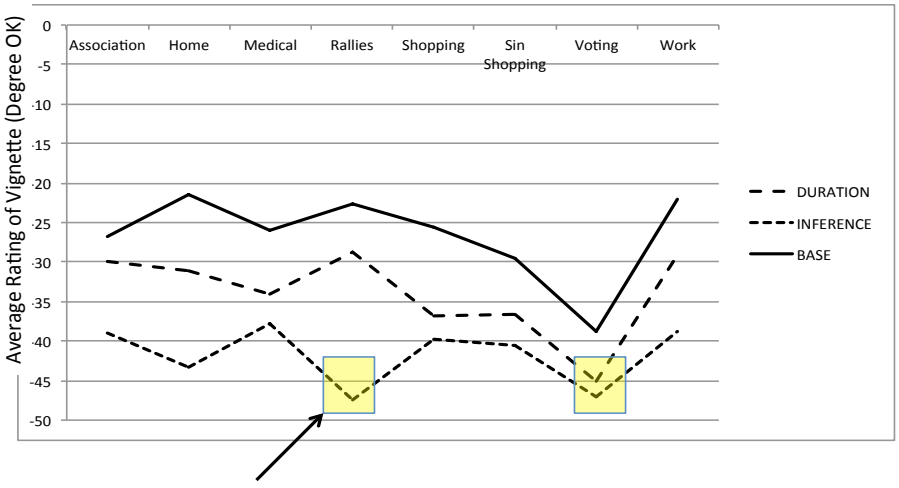
**Figure 3: Average Vignette Rating by Data Collection Medium**




d) Inferred Information

Using location information to identify whether someone voted or attended a rally was rated the lowest among the different inferences to be drawn, with an average rating of about -50. See Figure 4, with voting and attending a rally highlighted with yellow boxes.

**Figure 4: Average Vignette Rating by Inferred Information**




*Note the degree that tracking location about voting and rallies are outside privacy expectations*

*2. Interactions*

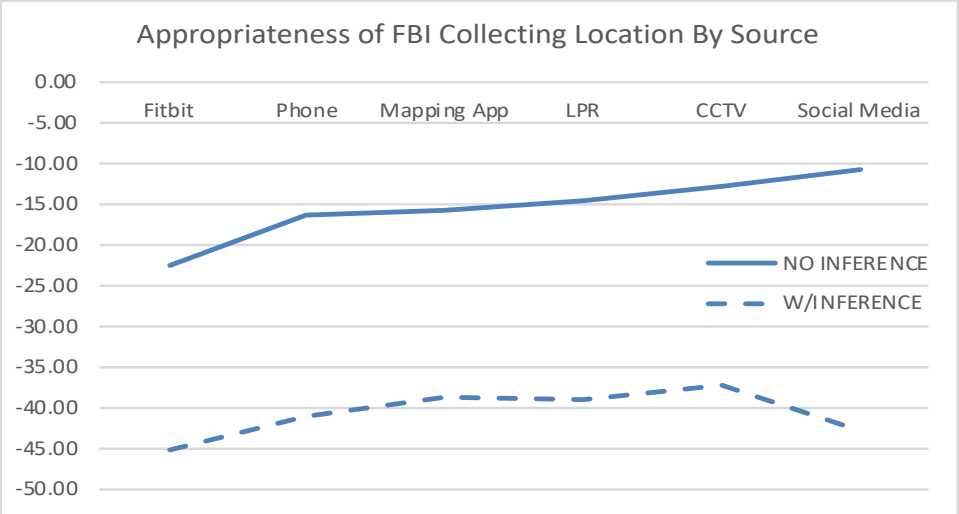
We were interested in whether the source, duration, or inferred information is perceived differently depending on the actor involved. For example, does the importance of the duration of location data collection depend on whether the actor acquiring the data is the FBI versus a family member? Does the importance of how the location data is collected depend on the actor collecting it? In order to identify if the actor modified the importance of the other contextual factors, we calculated the average vignette rating (the degree to which the vignette is rated "okay"). The results are in the Figures below.

a) Appropriateness of Source by Actor

To illustrate this point, we compared the FBI with data aggregators. Figure 5a, below, illustrates that the importance of the source (how the location data was gathered) was relatively stable for each actor aside from when the collecting actor was the FBI. The degree to which the scenario was appropriate was greatest for the FBI acquiring the location data through social media and least for the FBI accessing the location through a FitBit (with no inference explained).

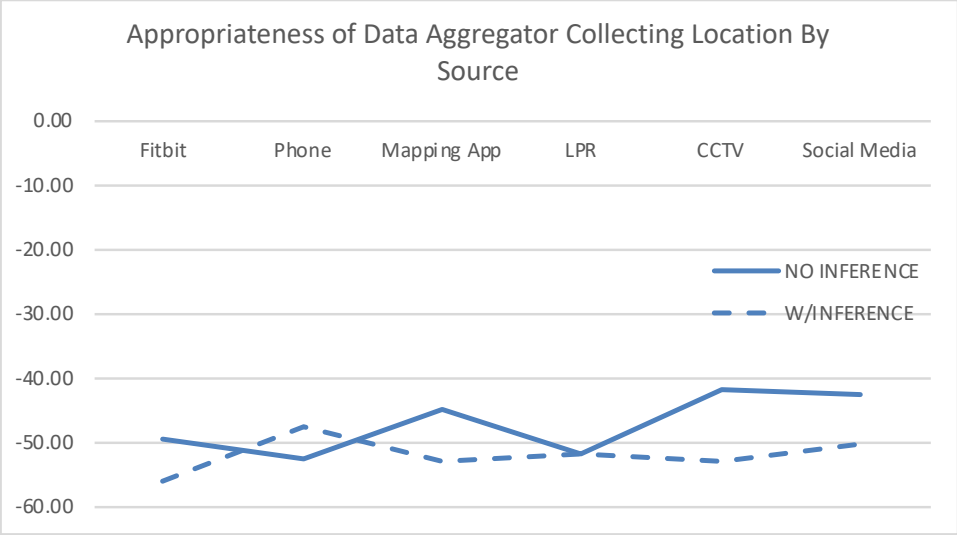
However, when the inference drawn about the individual is added to the vignettes (in Survey 3), the degree the collection of location data is appropriate decreases precipitously, and the manner in which the location data is collected (via phone versus FitBit versus social media) is statistically insignificant.

**Figure 5a: Average Vignette Rating for Each Source for FBI**



For data aggregators, the collection of location data from any source and either with inferences or without inferences included is not appropriate: the average vignette rating across sources being approximately -50.

**Figure 5b: Average Vignette Rating for Each Source for Data Aggregator**




b) Appropriate Duration by Actor

Figure 6a illustrates that duration is significant for particular actors, FBI in particular, compared to data aggregator in Figure 6b. The slope of each line is equal to the relative importance of duration to the rating task: a steeper negative slope is equivalent to the duration being more important to the rating task. For a data aggregator, the rating task is about the same regardless of the duration of the tracking. However, for the FBI, the duration of the surveillance is significant when the inference drawn is not included in the vignette, but disappears when inferences are included. This suggests that the inference drawn about the individual mediates the relationship between duration and the degree to which the location gathering is "okay." In other words, when individuals are concerned about the duration of surveillance, they are actually concerned about what inferences can be drawn from longer-term surveillance.

**Figure 6a: Average Vignette Rating for Each Duration by FBI**

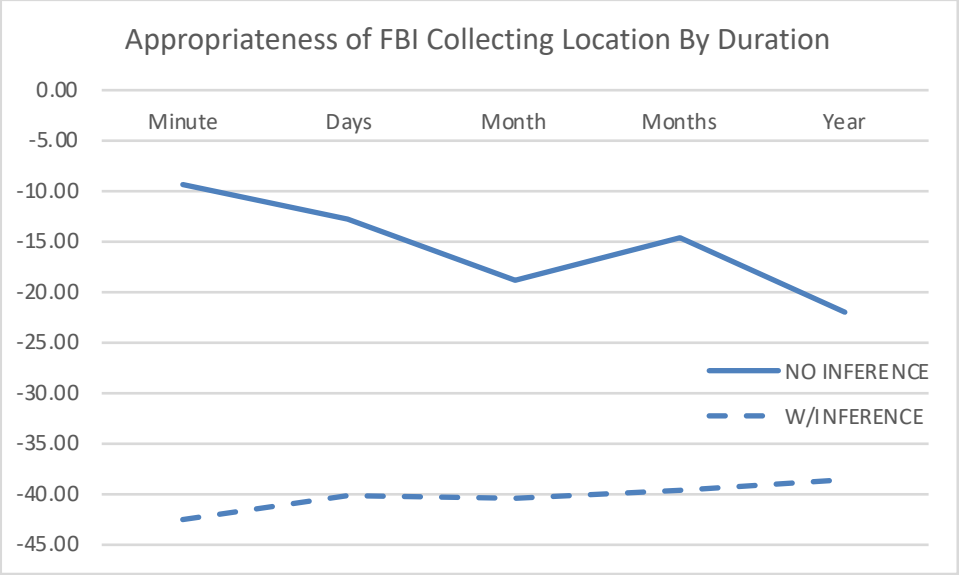
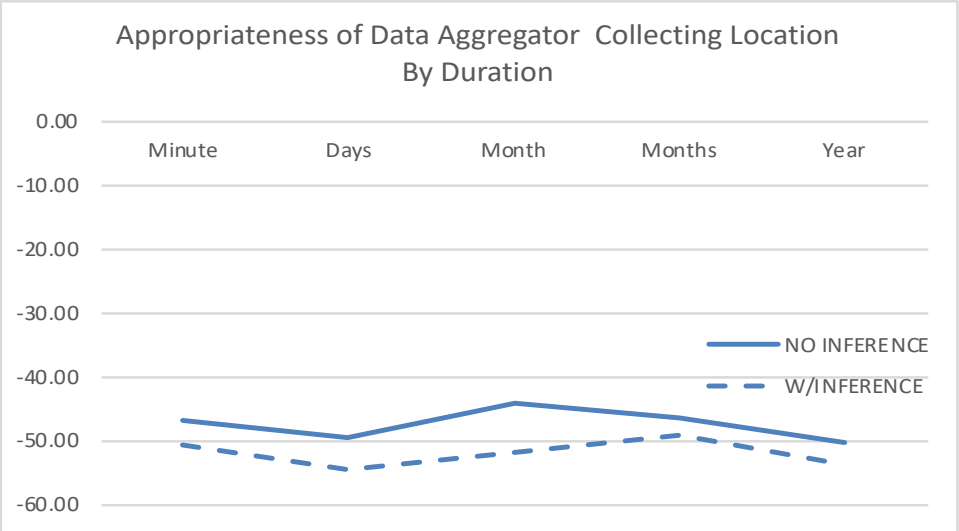



**Figure 6b: Average Vignette Rating for Each Duration by Data Aggregator**




## C. DISCUSSION OF MAIN STUDY

Varying the actor in the vignette who gathers the location information affects the degree to which the collection of location data is acceptable. However, the difference between the FBI or city services and a commercial entity (e.g., Yelp) diminishes as duration and inference are added. Importantly,

considering the attention given to the collection of location data from mobile phones, the difference in respondents' ratings of the appropriateness of collecting data across sources is significant but not the central driver of the appropriateness rating. All else being equal, gathering location data via a phone is statistically equivalent to gathering location data from a mapping app or social media but judged more acceptable than license-plate readers, CCTV, and FitBits. This finding may be significant for law and regulation that single out phones for distinctive treatment merely in their capacity to track location; these results suggest that individuals do not differentiate location information gathered via phone versus other mechanisms (CCTV, FitBits, etc.) as having different privacy expectations. The results suggest that the mechanisms for tracking location information, by themselves, do not drive privacy expectations.

The significance of duration disappears when inferences about an individual are also cited. This suggests that it is the potential for drawing inferences that mediates the relationship between duration and the assessments of appropriateness of location tracking. In other words, concerns over surveillance duration are actually concerns over inferences that longer-term surveillance facilitates.

## VI. FOLLOW-UP STUDY

Picking up on an issue we raised in Part I, the findings of our main study revealed one further aspect that needs attention. In particular, we sought greater clarity on how people conceive location in relation to how it is represented in technical systems and the policies that regulate them, either proclaimed by owners or imposed by others. This could tell us something about the match (or mismatch) between what concerns people when they say no to location tracking and the action a company takes to respect this: for example, ceasing to collect GPS data.[104]

Drawing on the finding from the pilot study that people respond to "location" and "GPS" in similar ways, we were interested in the impact of giving meaning or semantics to this numeric value. Pushing a step further toward this Article's driving question—"what is it about location?"—we sought to pinpoint the effects of naming a *place* by comparing it with references to generic *location*. In terms of CI, this follow-up study supplements the main study and pilot study with more specific insights on the parameter of information type and the ontologies that populate its parametric values.

104. Or other numerically represented location markers, such as nearest Wi-Fi coordinates or inferred location based on triangulation with nearby cell tower signals.

### A. FOLLOW-UP STUDY DESIGN

To isolate the importance of adding place to vignettes describing a generic location, we ran two factorial vignette surveys. These allowed us to examine if the meaning of location matters to the respondents by not including the duration or the inference drawn. Otherwise, the same factors and levels were used as in the live survey.

1. Base Survey: Actor-Source

   - {Actor} acquires location data from {Source}

For example, the vignette under the first condition would be:

- The FBI acquires location data from the signal of a mobile phone.
- An employer acquires location data from a mapping app (e.g, Google Maps).
- A city emergency service (ambulance, fire) acquires location data from geo-tagged posts on social media (e.g., Twitter, Facebook, Instagram).

2. Base + Place Survey: Actor-Source-Place

   - {Actor} acquires location data from {Source} and uses this data to figure out if a person was at {Place}.

The vignette under the second condition would be,

- The FBI acquires location data from the signal of a mobile phone and uses this data to figure out if a person was at a liquor store.
- An employer acquires location data from a mapping app (e.g., Google Maps) and uses the data to figure out if a person was at a liquor store.
- A city emergency service (ambulance, fire) acquires location data from geo-tagged posts on social media (e.g., Twitter, Facebook, Instagram) and uses the data to figure out if a person was at a liquor store.

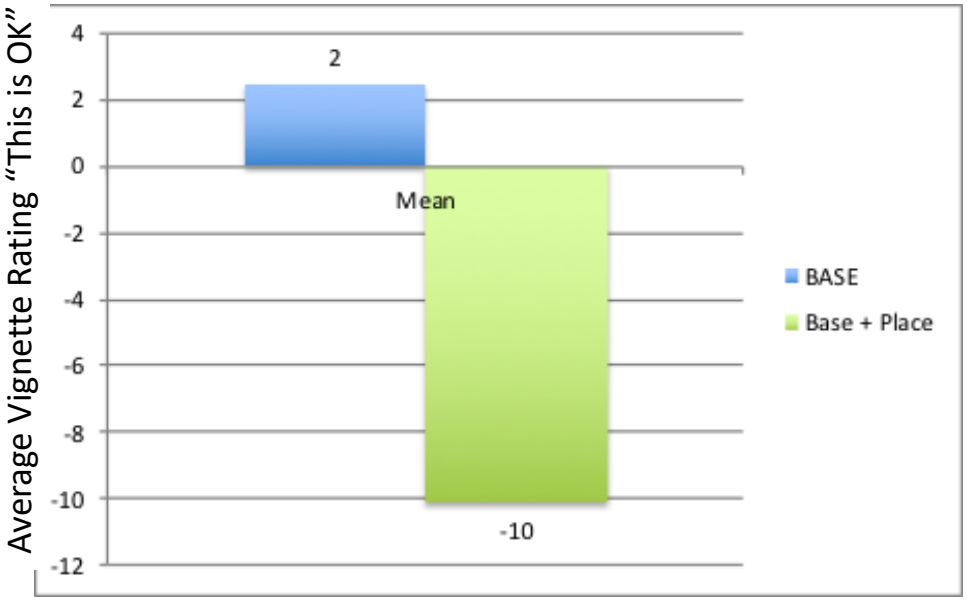
The survey was deployed on approximately 300 Amazon Mechanical Turk respondents, who each rated twenty vignettes.

### B. FOLLOW-UP STUDY RESULTS

#### *1. Average Rating Vignette Is "Okay"*

Results were quite stark: adding meaning to location data significantly drives down the average vignette rating, as shown in Figure 8.

**Figure 8: Average Vignette Rating for Both Conditions**

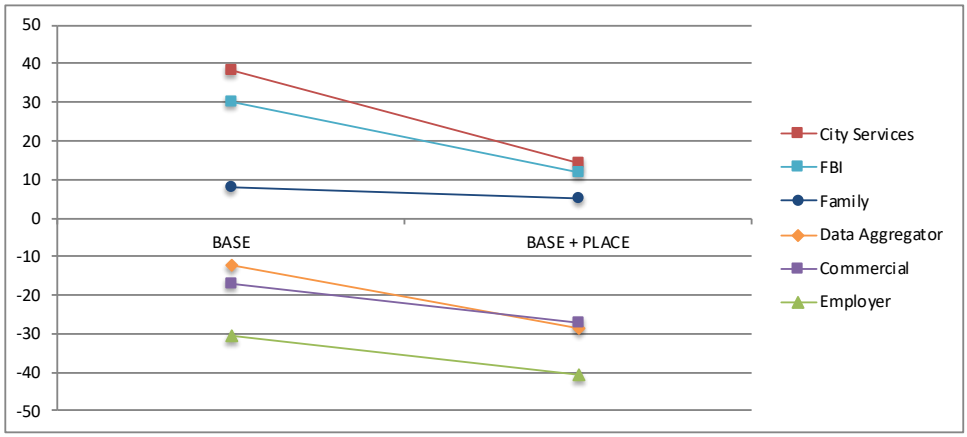



*2. Actors*

In addition, adding place to the vignette affects the collection of location data by the FBI and city services (which were relatively high) disproportionately more than other actors, as shown in Figure 9. The average vignette rating for the FBI drops from +30 to +10.[105]

105. No duration or inferences drawn about the individual were included in this survey. This isolates the impact of adding merely place to location data.

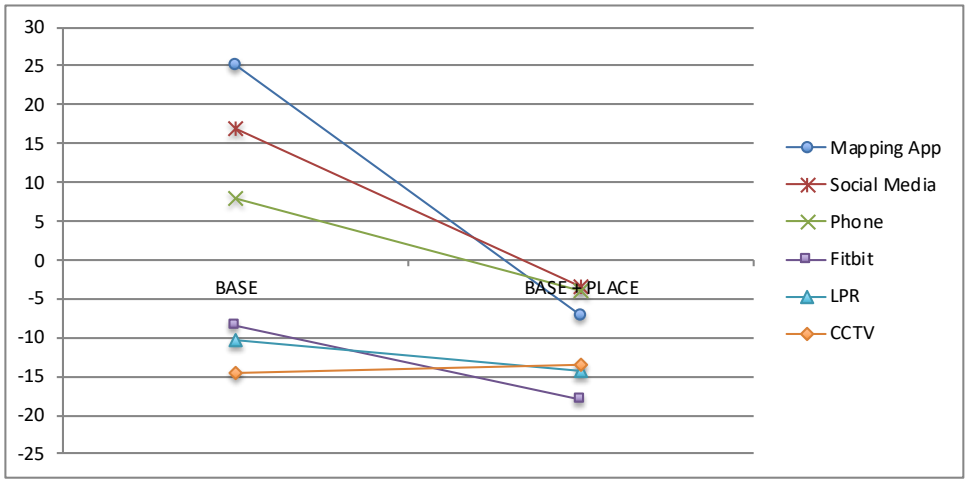
**Figure 9: Average Vignette Rating for Each Actor by Condition**




### *3. Source*

Finally, three sources are disproportionately impacted when the place is given meaning in the vignette: collecting location data from a phone, social media, and mapping app is positive when no meaning for place is provided but negative once the vignette included the place inferred from the location data. This suggests that asking respondents about data collection via these sources normally does not evoke particular places, and it would need to be made explicit in any survey.

**Figure 10: Average Vignette Rating by Data Collection Method (Source)**



### C. FOLLOW-UP STUDY DISCUSSION

In sum, we found the following.

- Adding place to a generic location negatively affects the degree to which the scenario is "okay" overall, and particularly for the FBI (from +30 to +12) and city services (from +38 to +14) as actors.
- Adding place significantly decreases the degree to which the scenario is "okay" for three sources: Mapping App: from +25 to -7; Social Media: from +17 to -4; and Phone: from +8 to -4.

## VII. SIGNIFICANCE FOR TECHNOLOGY, REGULATION, AND LAW

Amidst growing concerns about the steep rise of location tracking technologies and the widespread infiltration of location into data analytics, this Article asked, "what is it about location?" that worries us, the subjects of tracking. Our results shed light on how people understand location data, and how contextual factors affect people's reactions to others' knowing their whereabouts. Among many interesting and actionable findings, the results once and for all debunk the fiction that no expectations of privacy apply in public locations. To the contrary, we found not only that people have definite expectations, but that these expectations are nuanced and are systematically linked to the contextual factors for which we tested. Further, it is also clear from our findings that many common practices in which government and commercial entities engage are at odds with the expectations and attitudes that our studies reveal. Some of those findings are listed below:

- The collection of location data across actors and sources was judged "Not Okay" by respondents. The average ratings for each survey ranged from -29.7 (with place included) to -46.3 (when inferences were included);
- The results suggest that the word "location" is synonymous with "GPS" in judging the scenario as appropriate with no significant difference between the two levels ($p=0.95$). Further, the less precise measurement of locating someone at a street address or within a city was only a small improvement in the appropriateness of collecting location data.
- Duration was significant to the appropriateness of collecting location data when the inference drawn about the individual was not included. Interestingly, neither the length of storage nor the frequency of collection was similarly significant.

- As CI predicts, we found the actor collecting data to be a significant factor affecting people's attitudes. While respondents judged the collection of location data by all actors as "Not Okay" (with a negative rating), they did differentiate between actors. The relatively high approval of the FBI and city services as recipients of location diminished when duration is added as a factor, as well as inferences drawn.
- We anticipated differences, but the results showed that the source of location data (phone, Fitbit, social media) was not a significant predictor of respondents' judgments.
- The simple act of including place (at home, at work, etc.) had an outsize influence on responses. Adding place to a generic location negatively affects the degree to which the scenario is "okay" overall, and particularly for the FBI (from +30 to +12) and city services (from +38 to +14) as actors. Also, adding place significantly decreases the degree to which the scenario is "okay" for three sources: Mapping App: from +25 to -7; Social Media: from +17 to -4; and Phone: from +8 to -4.
- Across all variants, third-party data aggregators were among the most reviled among actors. With or without inferences drawn across sources, the average vignette rating was approximately -50. The juxtaposition of these findings with a hyperactive marketplace of third-party location data brokers siphoning up—buying and selling—is unsettling.[106]
- Finally, it is worth noting the degree of resentment people express about employers collecting location data, except when the location in question happens to be the workplace. Our findings are compatible with important work on employee surveillance by Professor Karen Levy as well as Professors Ifeoma Ajunwa, Kate Crawford, and Jason Schultz.[107] Our findings reinforce their arguments that even lawful employer surveillance of employees contravenes robust expectations. Our study shows clearly that there is a serious need to calibrate the existing letter of law with reasonable expectations of privacy.

106. *See* Jennifer Valentino-De Vries et al., *supra* note 27.

107. Karen E.C. Levy, *The Contexts of Control: Information, Power, and Truck-Driving Work*, 31 INFO. SOC'Y 160 (2015) (examining the impact of monitoring employees); Ifeoma Ajunwa et al., *Limitless Worker Surveillance*, 105 CALIF. L. REV. 735 (2017) (discussing the implications of employee surveillance by an employer in the context of the U.S. trucking industry).

### A. TECHNOLOGY

Our results are directly relevant to privacy-by-design. Here, we return to the intriguing question of how well people's understanding and preference for location privacy is represented (or modeled) in technology architecture and technical mechanisms. Our results flatly contradict the proposition that location privacy can be achieved by simply not collecting one of the technical markers, such as GPS coordinates. Technical research has shown that a smartphone user can be located using publicly available information, even when their location services are turned off.[108] Indeed, Google has admitted to tracking individuals with location services turned off (i.e., no GPS coordinates tracked), by triangulating an individual's whereabouts via nearest cell towers.[109] A further challenge comes from the ability to infer location from ostensibly non-location sensor data, collected by mobile devices where users' permission is not even needed.[110]

These results show that how location data is collected is not important to the privacy expectations. Further, the format of the data, whether as GPS coordinates or a street address, is not as important as locating someone at a "place." For companies, identifying individuals' location via other means, such as a data aggregator, a Wi-Fi sniffer, or a social media post is still not considered "okay." Further, asking individuals about the collection of GPS coordinates or even "location" data may not be precise enough for individuals to make a judgment—for, as noted, the simple inclusion of place (at home, at work, etc.) had an outsized influence on judgments of appropriateness. Furthermore, unless users are informed about the types of inferences that may be drawn, general questions about location privacy are ambiguous.

---

108. *See, e.g.*, Arsalan Mosenia et al., *PinMe: Tracking a Smartphone User around the World*, 4 IEEE TRANSACTIONS ON MULTI-SCALE COMPUTING SYS. 420 (2017) (demonstrating that minimal information is required to track a smartphone user's location even when GPS is turned off); Lindsey Barrett, *Model(ing) Privacy: Empirical Approaches to Privacy Law & Governance*, 35 SANTA CLARA HIGH TECH. L.J. 1, 20 (2018) (noting the different studies which have found that, despite privacy policies, applications on phones can still access the information even when not in use).

109. Shannon Liao, *Google Admits it Tracked User Location Data Even When the Wetting was Turned Off*, VERGE (Nov. 21, 2017), https://www.theverge.com/2017/11/21/16684818/google-location-tracking-cell-tower-data-android-os-firebase-privacy [https://perma.cc/K98P-ZMTF]; Keith Collins, *Google Collects Android Users' Locations Even When Location Services are Disabled*, QUARTZ (Nov. 21, 2017), https://qz.com/1131515/google-collects-android-users-locations-even-when-location-services-are-disabled/ [https://perma.cc/A9ZL-FLX9].

110. Sashank Narain et al., *Inferring User Routes and Locations Using Zero-Permission Mobile Sensors*, 2016 IEEE SYMP. ON SECURITY & PRIVACY 397 (explaining how a user's travel route can be inferred with high accuracy from gyroscope, accelerometer, and magnetometer information).

Beyond the challenge of stopping end runs around technical location markers, our results raise the question of how to represent location semantics through technical variables, and whether it is even possible. Protecting against inferences drawn from an individual's whereabouts (particularly patterns of movement over time) may be beyond purely technical means; similarly, limiting access to an individual's "place" may be challenging. As noted, place data (home, work, shopping, etc.) may be available outside of GPS and GIS systems via natural language communication on social media.

Finally, the collection of location data by third-party data aggregators was consistently judged inappropriately, no matter the duration or inferences drawn. This finding is consistent with our previous work showing strong disapproval of third-party brokers and aggregators, even when collecting data from public records.[111] This means that the common practice of integrating code from external libraries that either shares or integrates location data from third-party aggregators flies in the face of privacy expectations and significantly undermines trust.[112]

### B. SIGNIFICANCE FOR REGULATION

The GDPR has introduced new privacy requirements for the data processing practices of firms doing business with European individuals. The processing of identifiable information, including location data, is limited: data processors must meet one of a few criteria, including that the processing of location data is necessary to complete a contract obligation, to protect vital interests of data subject or other person, to perform a task in the public interest, to comply with a law or regulation, or "for the legitimate interests" pursued by the data controller or a third party.[113]

Our findings are helpful for defining a company's legitimate interests. According to GDPR Article 6, there is a three-part test for identifying exceptions to a data processor's legitimate interests.[114] First, is there a legitimate interest behind the processing of location data? Second, is that processing necessary for that purpose? Finally, is the legitimate interest overridden by the individual's interests, rights, or freedoms? According to the GDPR, individuals' interests are defined in terms of reasonable expectations

111. *See* Martin & Nissenbaum, *Privacy Interests in Public Records*, *supra* note 7.

112. Martin, *Breaking the Privacy Paradox*, *supra* note 90; Martin, *The Penalty for Privacy Violations*, *supra* note 90.

113. *Lawful Basis for Processing*, *supra* note 40.

114. *What Is the 'Legitimate Interests' Basis?*, INFO. COMM'R OFF., https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/ [https://perma.cc/8TDY-5UF9] (last visited Dec. 30, 2019).

around the collection and use of location data.[115] In other words, if individuals have strong expectations that location data will not be processed for that purpose, the interests of the individual are superseded by the 'legitimate' interests of the data processer.

The findings here illustrate under what conditions individuals find the gathering of location data to be appropriate and the appropriate inferences to be drawn about them. While using location data to identify if an individual is at home is deemed appropriate for a family member, the majority of scenarios were deemed inappropriate, particularly for commercial actors such as an employer, a data aggregator, or a commercial service (such as Yelp). Taking reasonable expectations seriously, our results suggest that the number of uses for location data aligning with a company's purposes are considerably fewer than companies may seek to claim.

Outside of GDPR, the study suggests that regulations should focus on the type of information, rather than how the information is collected, to protect the interests of consumers and users. Location data was deemed inappropriate to collect regardless of how the information was gathered. Therefore, regulations that mistakenly narrowly focus on types of collection mechanisms (e.g., only based on trackers or Wi-Fi sniffers) would allow companies to collect location data that people deem inappropriate; such regulations would do little to actually protect the privacy interests of individuals. Our findings reinforce the idea that regulations should not follow specific technologies, but instead map onto values for the nature of the information, the recipients, and the flow constraints (collection, use, sharing, etc).

Instead, privacy regulations should look to limit who has easy access to location data after the initial collection. The results here show that particular actors, such as data aggregators, were consistently deemed not appropriate to collect location data. The current focus in the United States is to heavily regulate the handoff or initial disclosure of information through adequate notification and user consent. After disclosure, regulations are silent. Regulations should instead shift to focus on the sharing, aggregation, and use of information, including location data, after initial disclosure.

115. *Id.* ("The GDPR is clear that the interests of the individual could in particular override your legitimate interests if you intend to process personal data in ways the individual does not reasonably expect."); Regulation 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) (GDPR) art. 6, 2016 O.J. (L 119) 1.

### C. SIGNIFICANCE FOR LEGAL DECISIONS

This study fills the need to better understand societal privacy expectations as a means to ascertain whether actual expectations are reasonable. Professors Matthew Kugler and Lior Strahilevitz clearly summarize why actual beliefs (as measured in surveys) are relevant to court opinions. They state, "[w]e show how scientific polling can alleviate concerns that, in undertaking such an inquiry, judges will place undue weight on their own beliefs or on the beliefs of people in their social orbits."[116] Around location data specifically, Kugler and Strahilevitz note that reasonable expectations of privacy are "the average person's expectations" or "popular expectations."[117] These studies empirically examine the privacy expectations of individuals around the collection of location data in public.

Many of our findings could be useful in the courts. For example, the precision of location data (GPS coordinates, street address, city block, etc.) is less important than *who* is collecting the information. Further, the level of precision was far less important to respondents than whether location was identified in terms of a meaningful place (work, rally, home, etc.). Finally, the precision was less important than the type of inferences drawn or type of knowledge created by the breadth of location data collected.

In the past, courts have focused on GPS data gathered from a phone or GPS device in a car. Going forward, they would do well to highlight the nature of the collecting actor ("recipient" in CI terms; boss, the FBI, parents, etc.), rather than the source alone (phone, CCTV, mapping app, etc.). An exception we found was FitBit, which provoked greater disapprobation.

The appropriate duration of collecting location data before a warrant is needed is in flux. Previous work by Kugler and Strahilevitz found duration is not significant in affecting judgments concerning when a warrant is necessary.[118] In our study, duration seems to matter only insofar as it mediates inferences and ceases to play an explanatory role once an inference or place has been declared. In other words, respondents cared about duration only when no meaning was given to the location data. This conforms to what others have dubbed "mosaic theory,"[119] which is an awareness that insignificant bits of information, aggregated, may create a fine-grained picture that can threaten privacy. Interestingly, in the cases of inferring voting and attendance at a

116. Kugler & Strahilevitz, *supra* note 1, at 220.
117. *Id.* at 207.
118. *Id.* at 245, 248.
119. Ohm, *supra* note 43, at 373.

political rally from location, respondents disapproved across the board for all the recipients we listed in our study.

### D. SIGNIFICANCE FOR HOW LOCATION IS LABELED IN SURVEYS AND LAW

How we ask about location in surveys matters to the normative judgments of individuals. As noted above, previous empirical work has asked about the collection of GPS coordinates. However, adding details such as the duration of collection, the place, and the inferences drawn about the individual decreases the degree to which the vignettes are "okay" and can change the relative importance of the actors, source, and place in determining whether the scenario is acceptable. This means that surveys about GPS location data will not capture privacy expectations regarding location unless surveys also include the type of knowledge created by the aggregated data and the purpose of the collection of the data in terms that are meaningful.

Based on the impacts on judgments when place is specified, researchers should start with the assumption that *place* is independent of the numerical GPS measure of latitude-longitude. Thus, it warrants independent study, in addition to interactions with the identities of collecting actors (recipients).

Three sources are disproportionately affected when place is given meaning in the vignette: collecting location data from a phone, social media, and mapping app is positive when no meaning for place is provided but negative once the place inferred from the location data is included in the vignette. This suggests that asking respondents about data collection via these sources would need to be made explicit in any survey.

Further, respondents appear to assume a short duration of collecting data when no duration of collection is mentioned. The vignette with no duration included was rated the same on average as a vignette with location data stored for one minute only. This is important, as respondents in a survey make assumptions about the given scenario when the researcher is silent on the matter. From our research here, respondents assume no particular "place" as being inferred from location data and assume the duration is short. By remaining silent on those factors, surveys might mistakenly be thought to support the collection of location data under a variety of conditions, when, in fact, this is an artifact of respondents assuming a best-case scenario.

Finally, more research is needed in order to explain consistently negative reactions to data aggregators or data traffickers.[120]

120. *See* Lauren Henry Scholz, *Privacy Remedies*, 94 IND. L.J. (forthcoming 2019) (manuscript at 12–13), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3159746.

## VIII. CONCLUSION

It is important to clarify the scope of our research and highlight its key contributions while acknowledging its limits. Our work set out to address gaps and shortcomings in how location privacy is conceived, which has led, in our view, to flawed technical, regulatory, and legal responses. Because *reasonable expectation of privacy* has served as a critical linchpin in all three of these domains, we approached these gaps and shortcomings through large-scale empirical studies structured around the theory of CI to provide a solid basis for revising these responses. The results of these studies are most striking in roundly debunking assumptions that have impeded privacy policy and practice generally, and in this instance, for location privacy.

What are some of these debunked truisms? First is a misplaced faith in the decisive influence of the public-private dichotomy. The studies reported in this Article (as with those in the preceding two) confirm that "public is public" is plain wrong; our respondents revealed strong, nuanced, and systematic privacy expectations in spaces and places typically considered public. Although we have been early proponents of this view and are no longer alone in holding it,[121] our studies offer compelling empirical backing. Second, our studies reveal that purely mathematical, non-semantic location markers (e.g., GPS coordinates) do not adequately model location privacy expectations. Misleading labels in device and system interfaces may, therefore, deceive users about underlying data practices. Finally, like it or not, respondents were highly discriminating on the question of recipients. From the perspective of CI, this is unsurprising, but, once again, this finding exposes how poorly the public-private dichotomy models expectations of privacy. Respondents were varied in their judgments of appropriateness for family, FBI, etc., though were consistently and deeply negative about third-party location aggregators and brokers.

With these and other general findings, our studies demonstrate the need for further and more detailed investigations. Our findings were obviously rooted in our own intuitions, particular interests, and controversies reported in mainstream media. Clearly, they do not offer anything close to a complete picture of location privacy expectations, particularly under the assumption that five parameters, at least, are simultaneously relevant to these expectations. With some of the gaps filled and a few key misconceptions debunked, this Article is an argument for more detailed and better studies of location privacy that will serve sounder court decisions, regulation, and technology design.

121. *See* Woodrow Hartzog, *The Public Information Fallacy*, 99 B.U. L. REV. 459 (2019).

## APPENDIX A – PILOT STUDY FOR SURVEY DESIGN

Important findings from the pilot studies:

- While the frequency of tracking did not significantly affect the degree to which a vignette was rated "okay," the duration of tracking did. In particular, duration of tracking was important when the FBI was the actor acquiring the data. These results are suggestive of further lines of inquiry into variation across individuals—for example, whether individuals with high privacy despair (low trust, high privacy is important) are particularly sensitive to the duration factor in relation to the FBI.
- We found, inadvertently, two potential framing effects: a) the order in which subjects were presented control questions, whether before or after vignettes, affected judgments; and b) whether vignettes expressed in second or third person, also had an impact. These findings suggest a need for future work to refine privacy survey methodology.

We focused on the collection of location data due to how easily the data can be used to identify other information (e.g., who you are with) and because new technology supports tracking location data in new ways. Even narrowing the focus to location, grappling with the complexity of studying privacy in public called for a pilot study. In particular, the pilot study helped us to settle preliminary choices along three dimensions: first was the selection and articulation of factors of the factorial vignette survey; second was to guide our choices of respondent control questions; and third was to guide our choices of survey features, including the order of presentation of the parts, and survey voice ("you" or "a person").

In the pilot survey, we experimented on these three elements. For example, in order to simplify the vignette, we tested a few of the factors (location, frequency, storage) included in the respondent controls both before and after the vignettes, and tested three different 'voices' for the vignette. In the end, we were able to simplify by using the word "location" only, dropping frequency, using the third-party vignette voice, and including control questions after the vignettes.

A respondent's degree of extroversion[122] also impacted the privacy expectations in public in the study conducted by Matthew Kugler and Lior

122. The format of the surveys and controls are the same as the live survey in the Article above. We added one control (personality), explained here.

Strahilevitz.[123] Based on the five personality factors used across disciplines,[124] we used the scale for extroversion with the degree to which the respondent saw themselves as "Extroverted, enthusiastic (that is, sociable, assertive, talkative, active, NOT reserved, or shy)."

### A. PILOT STUDY SURVEY DESIGN

Two facets of the survey design initially tested: (1) the ordering of the control questions; (2) the voice of the vignettes, and two parameters of the factorial vignette; (3) the precision of the location data described; and (4) significance of frequency of tracking. The results of this pilot study were used to design the main surveys described later.

#### *1. Features Tested*

a. **Ordering of Controls and Vignettes**. Did placing the controls before or after the vignettes matter to (i) the rating of the vignette or (ii) the respondents' ratings of the controls? To ensure the ordering did not impact the vignette ratings, we ran the pilot survey with the respondent controls both before and after asking the respondents to rate the vignettes. Table 1 illustrates the respondent controls being asked after the vignettes.

b. **Vignette Voice** ("you" versus "a person"). We tested if the 'voice' of the vignette (second person, third person, or third person plural) mattered to the judgment of whether the information flow was appropriate as has been suggested before.[125] The survey was run three times with each type of voice and as depicted in Table 2.

c. **Location**. Did the operationalization of "location" as GPS, location, street address, or city matter to the appropriateness of the scenario offered?

d. **Storage versus Frequency of Data Collection**. In order to test if the frequency of the data collection or the time the data was stored impacted appropriateness of the information flow, we included both factors in the vignette.

123. Kugler & Strahilevitz, *supra* note 1, at 251–55.

124. Robert R. McCrae & Paul T. Costa, Jr., *Validation of the Five-Factor Model of Personality Across Instruments and Observers.*, 52(1) J. PERSONALITY & SOC. PSYCHOL. 81, 83 (1987).

125. Slobogin & Schumacher, *supra* note 10, at 759.

*2. Vignette Factors in Pilot Study*

Table A1 below includes the vignette factors included in the pilot study to identify the importance of voice, storage, frequency, and precision of location. These factors are later 'fixed' in the subsequent study.

**Table A1: Factors for Pilot Survey**

| **Factor** | Levels | Operationalized in Vignette |
|---|---|---|
| **Frequency** of data gathering | Continuous | Continuously, every hour, every day. |
| **Storage** How long the information is retained | Continuous | Indefinitely, 1 year, 1 month, 1 day, 1 hour, 10 minutes and then discarded. |
| **Actor** | Government | The local police |
| | Federal Government | FBI |
| | Phone | The operating system of a phone/device (e.g., Google Android or Apple iOS) |
| | Commercial | Companies offering a location-based service (local reviews or recommendations) |
| | Family | Family members (e.g., parents, spouse, or sibling) |
| **Voice** | | 1st person, **3rd Person Singular,** *3rd Person Plural* |
| **Precision** How specific is the location data | Location | Your location, **a person's location,** *individuals' location* |
| | City | Which city you are in, **which city a person is in,** *which city individuals are in* |
| | Street Address | Your nearest street address, **a person's nearest street address,** *individuals' nearest street address* |
| | GPS | Your GPS coordinates, **a person's GPS coordinates,** *individuals' GPS coordinates* |

*3. Vignette Template for Pilot Study*

[Actor] collects [Precision] [Frequency] and stores that data [Storage]. E.g.,

- Second person: Companies offering a location-based service (local reviews or recommendations) collects your location every 15 minutes and stores that data for 1 year**.**
- Third person: The FBI collects a person's nearest street address continuously and stores that data for 1 hour.
- Third person plural: The operating system of a phone/device (e.g., Google Android or Apple iOS) collects which city individuals are in continuously and stores that data for 1 hour.

*4. Vignette Rating Task*

For each vignette, respondents were instructed to indicate the degree to which they agreed with the question "Is this okay?" with a slider. The left side of the slider indicated "Definitely Not Okay" and the right side of the slider indicated "Definitely Okay." The slider was on a scale of -100 to +100 with the number suppressed so the respondents saw only the labels "Okay" and "Not Okay."

B. PILOT RESULTS

*1. Ordering of Controls and Vignettes*

In order to test if the order in which the respondents were asked to rate the vignettes and control questions mattered to the results, the surveys were run first with the vignettes after the control questions and a second time with the controls asked after the vignettes. Table A2 includes the sample statistics of both surveys run.

The average vignette rating did not change when the control questions were asked first versus after the vignettes. The average rating remained about -36 ("Not Okay"). Interestingly, the ratings for certain control variables did change when the controls were asked after the vignettes, as shown in Table A2.

Specifically,

- The authoritarian score decreased from -13.32 to -20.58 when the question is asked after the vignettes. In other words, the respondents are less authoritarian after rating scenarios about commercial and governmental tracking.

- The average trust in business rating also decreases from -12.12 to -25.95 when the question is asked after the vignettes are rated. This is consistent with previous work on trust and privacy.[126]

**Table A2: Sample Statistics for Surveys with Control Questions Before and After Vignettes**

| | Average Sample Statistics | |
|---|---|---|
| | Controls 1st | Controls 2nd |
| N Respondents | 444 | **406** |
| Authoritarian Scale | -13.32 | **-20.48** |
| Trust Scale | 0.42 | **-8.80** |
| Female | 1.53 | 1.41 |
| AgeOver35 | 0.55 | 0.37 |
| Privacy Important | 72.56 | 71.85 |
| Trust Government | -23.14 | -29.63 |
| Trust Business | -12.12 | **-25.95** |
| _eq2_R2 | 0.77 | 0.77 |
| DV Mean | -36.72 | -35.82 |

*2. Vignette Voice ("You" Versus "A Person")*

To test the importance of the vignette voice, the survey was run three times. Voice did make a difference. When the vignettes included a reference to the respondent ("you"), the vignettes were rated less "okay" (-35.32) compared to a third-person voice (-27.05) or a third-person plural voice (-30.45).

*3. Location*

In order to examine how respondents' make sense of the precision of the location data collected, the rating task was regressed on the vignette factors and the results are in Table A3 below. The results suggest that the word "location" is synonymous with "GPS" in judging the scenario as appropriate, with no significant difference between the two levels (p=0.95). Precision does matter to the vignette rating with a reference to only the street address (+5.69) and city (+8.19), as both considered improvements for the respondents over collecting GPS-level data.

126. *See* Martin, *supra* note 95 (examining the impact of the introduction of privacy notices on consumer's trust).

**Table A3: Regression of Vignette Rating Task on Vignette Factors for Pilot Design Survey**

| | Coef. | p |
|---|---|---|
| **Location** | | |
| Street | 5.69 | 0.00 |
| City | 8.19 | 0.00 |
| GPS | 0.09 | 0.95 |
| (null = location) | | |
| **Actor** | | |
| CommercialActor | -14.13 | 0.00 |
| FBIActor | -30.14 | 0.00 |
| PhoneOSActor | -19.78 | 0.00 |
| PoliceActor | -34.15 | 0.00 |
| (null = family) | | |
| FrequencyScale | 0.86 | 0.26 |
| StorageScore | -8.80 | 0.00 |
| **Controls:** | | |
| HighExtroversion | 0.70 | 0.64 |
| HighAuthoritarian | 12.09 | 0.00 |
| HighTrustDisposition | 5.48 | 0.00 |
| HighPrivacyImport | -20.90 | 0.00 |
| HighTrustBusiness | 20.32 | 0.00 |
| _cons | 9.88 | 0.00 |

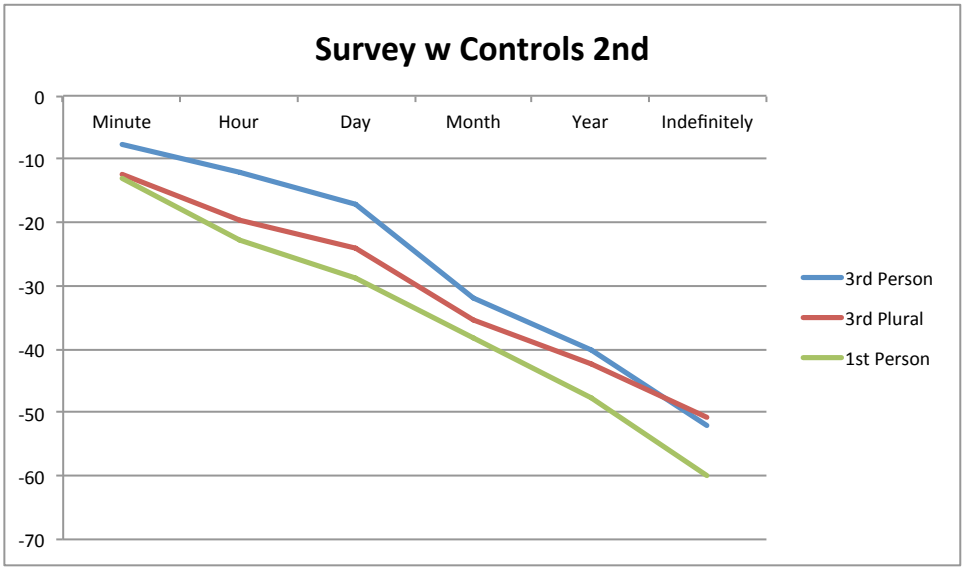
*4. Storage Versus Frequency of Data Collection*

In order to focus on whether the storage of data or the frequency with which the location data is collected impacted the outcome, we examined the relative importance of both factors in the regression results in Table A1 and the average rating task (the degree to which the scenario is "okay") for each amount of storage in Figure A1.

The length of storage time is inversely related to the rating of the vignette as "okay," indicated by the steep negative slope in Figure A1. Frequency, by

contrast, was not significant; respondents did not rate the vignette any differently as the frequency levels changed.[127]

**Figure A1: Average Vignette Rating for Each Level of Storage by Survey Voice**




*5. Discussion of Pilot Survey*

In sum:

- We used the term "location" in the later studies, knowing that the term is equivalent to "GPS" for the respondent;
- We dropped the use of frequency;
- We shifted to the term "duration" for the duration of tracked location information;
- We used the third-person plural in the later vignettes and asked the control questions after the vignettes.

127. Because this result was somewhat surprising, we ran another vignette survey without storage included as a factor to allow the respondent to focus on frequency (from every five seconds to once per day). However, frequency was still not significant; the only difference was the average vignette rating decreased from -35.52 to -31.57 when storage was removed as a factor.

## APPENDIX B – FOLLOW-ON STUDY

We tested the impact of adding place to a vignette with just actors and source. This would allow us to see the effect of explaining the place that the location data provides. In other words, does it matter to respondents if we describe location tracking as gathering location data versus gathering location data to figure out someone is at a particular place?

### A. FOLLOW-ON STUDY #1: ADDING PLACE TO A SURVEY ABOUT LOCATION

To isolate the importance of adding place to vignettes describing a generic location, we ran two factorial vignette surveys. This allowed us to pilot if giving meaning to the location would matter to respondents. The same factors and levels were used in the live survey. The table is provided below as Table B1.

1. Base Survey: Actor-Source

    - {Actor} acquires location data from {Source}.

2. Base + Place Survey: Actor-Source Place

    - {Actor} acquires location data from {Source} and uses this data to figure out if a person was at {Place}.

For example, the vignette under the first condition would be,

- The FBI acquires location data from the signal from a mobile phone.
- An employer acquires location data from a mapping app (e.g., Google Maps).
- A city emergency service (ambulance, fire) acquires location data from geo-tagged posts on social media (e.g., Twitter, Facebook, Instagram).

Whereas the vignette under the second condition would be,

- The FBI acquires location data from the signal from a mobile phone and uses this data to figure out if a person was at a liquor store.
- An employer acquires location data from a mapping app (e.g., Google Maps) and uses the data to figure out if a person was at a liquor store.
- A city emergency service (ambulance, fire) acquires location data from geo-tagged posts on social media (e.g., Twitter, Facebook, Instagram) and uses the data to figure out if a person was at a liquor store.

**Table B1: Factors Used in Pilot**

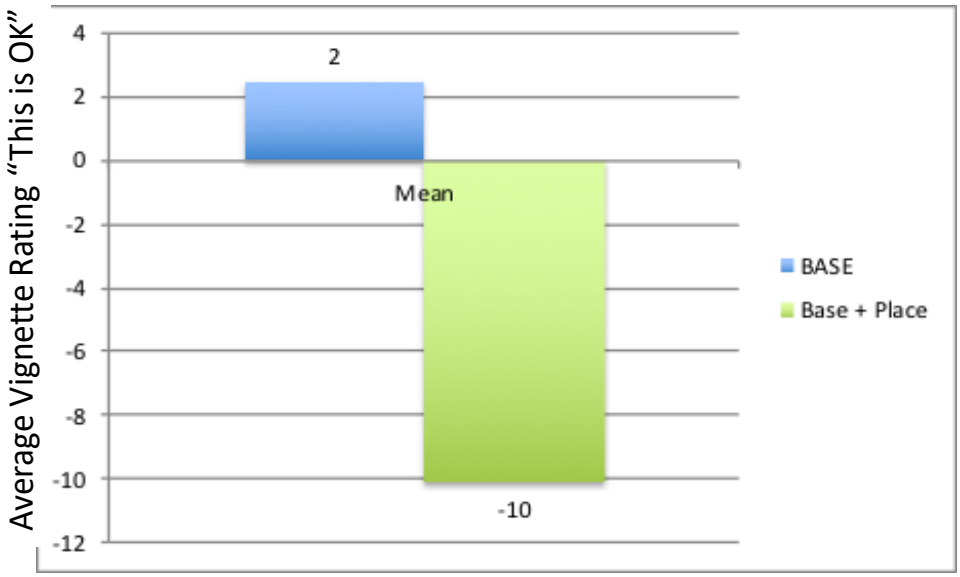
| **Concept** | Description | As operationalized in Vignette |
|---|---|---|
| **Duration** | Length of tracking | A year, about 6 months, a month, a few days, a few minutes |
| **Actor** | Government | A city emergency service (e.g., ambulance or fire) |
| | Federal Government | The FBI |
| | Employer | An employer |
| | Commercial data aggregator | A commercial data broker |
| | Commercial | A commercial location-based service (e.g., Yelp) |
| | Family | A family member (e.g., parents, spouse, or sibling) |
| **Source** | License-plate reader | License-plate readers |
| | CCTV | CCTV cameras with facial recognition |
| | Phone | The signal from a mobile phone |
| | Fit Bit | A fitness app (e.g., FitBit or Stava). |
| | Social media | From geo-tagged posts on social media (e.g., Twitter, Facebook, or Instagram) |
| | Mapping app | A mapping app (e.g., Google Maps) |
| **Place** | Association | A restaurant or cafe |
| | Protests/rallies | The National Mall |
| | Sin Shopping | A liquor store |
| | Shopping\| | A shoe store |
| | Home | Home |
| | Work | Work |
| | Medical | A medical clinic |
| | Voting | A voting site |

The survey was deployed using Amazon Mechanical Turk. Approximately 150 respondents each rated twenty vignettes.

*1. Average Rating Vignette Is "Okay"*

Adding place to the vignette and giving meaning to what location data could mean drives down the average vignette rating, as shown in Figure B1.

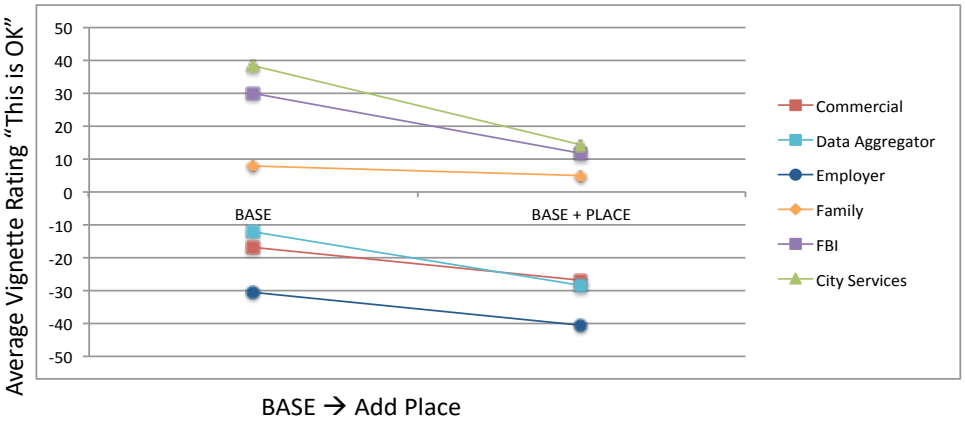
**Figure B1: Average Vignette Rating for Both Conditions**




*2. Actors*

In addition, adding place in condition 2 impacts the collection of location data by the FBI and city services more than other actors (although all were impacted), as shown in Figure B2. The average vignette rating for the FBI drops from +30 to +10.

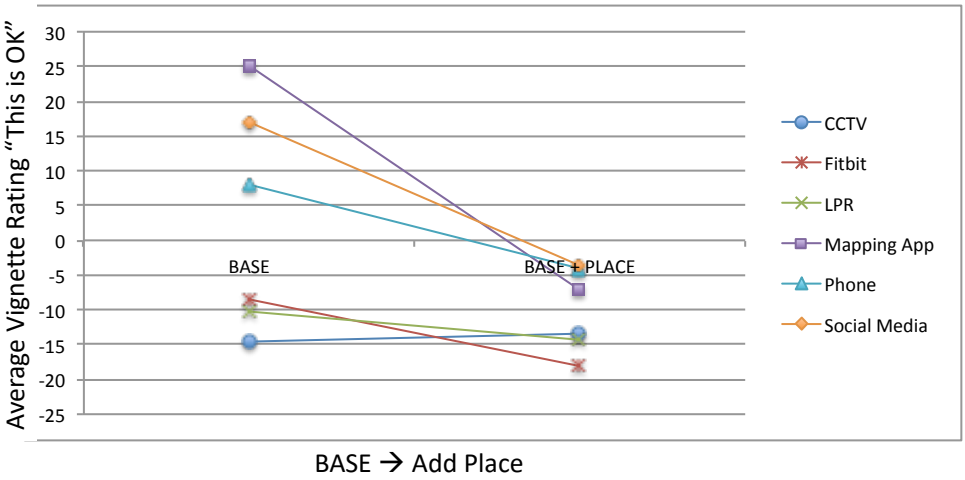
**Figure B2: Average Vignette Rating for Each Actor by Condition**




*3. Source*

Finally, three sources are disproportionately impacted when the place is given meaning in the vignette: collecting location data from a phone, social media, and a mapping app is positive when no meaning for place is provided but negative once the place inferred from the location data is included in the vignette. This suggests that asking respondents about data collection via these sources normally does evoke particular places. This would need to be made explicit in any surveys.

**Figure B3: Average Vignette Rating for Each Source by Condition**



## B. FOLLOW-ON STUDY #2: ADDING PLACE TO SURVEY WITH DURATION INCLUDED

We then ran the survey two more times, with the duration factor added to both a base survey (actor, source, duration) and the survey with place included (actor, source, duration, and place). This allowed us to isolate the importance of place when duration is also included. In addition, this allowed us to measure how important duration is when the meaning of the location is also included.

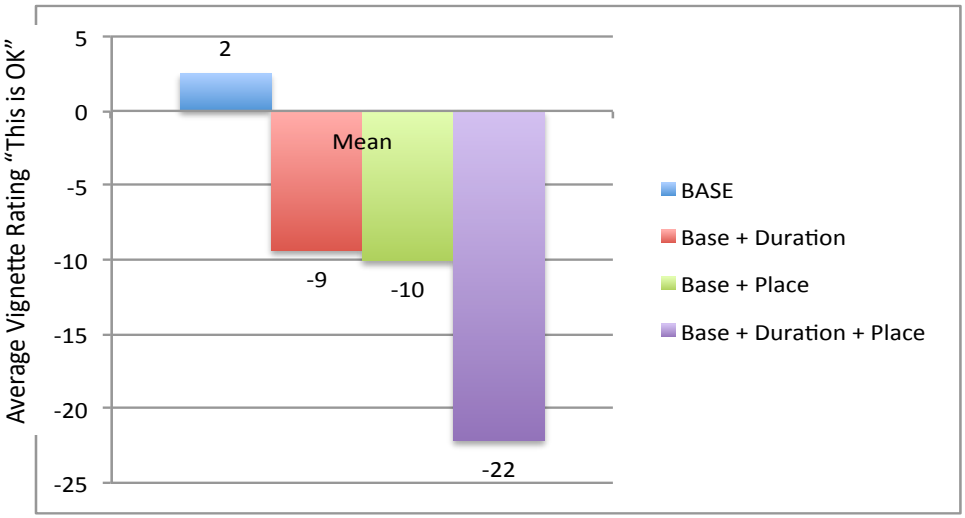
In other words, it is possible that when people are concerned about the duration of data collection, they are actually worried about what someone could find out about them. This would suggest that duration could be mediated by place.

The surveys were run again on Amazon Mechanical Turk with approximately 150 respondents for each condition.

### *1. Average Rating Vignette Is "Okay"*

Adding place to the survey with duration already included did impact the average vignette rating.

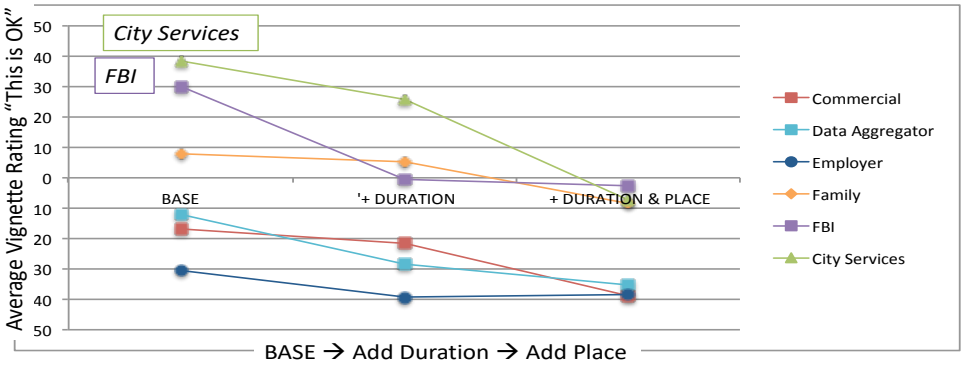
**Figure B4: Average Vignette Rating for All 4 Conditions**



*2. Actor*

To see how actor is judged when the place is added, we can track how the average rating ("This is Okay") changes across surveys. The FBI and city services are impacted the most by including duration and place in the vignette.

**Figure B5: Average Vignette Rating by Actor for Place Condition**

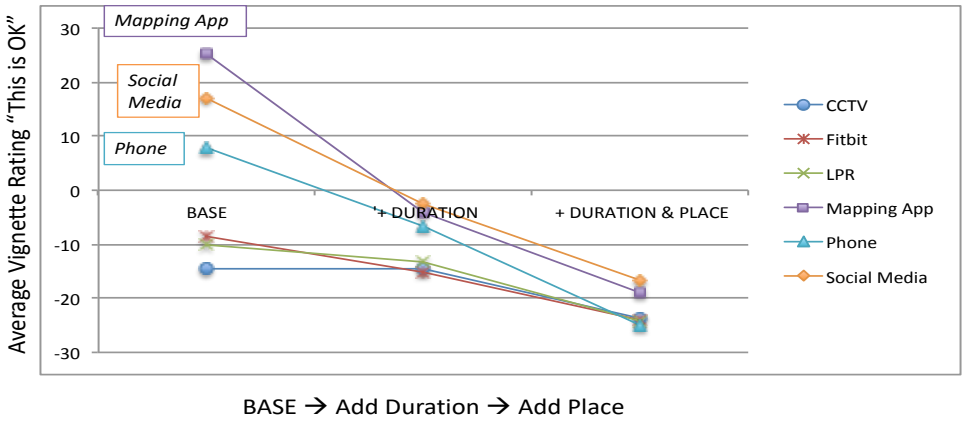



*3. Source*

Mapping App, Social Media, and Phone are impacted the most by including duration and place in the vignette.

- Mapping App: +25 → -19
- Social Media: +17 → -16
- Phone: +8 → -25

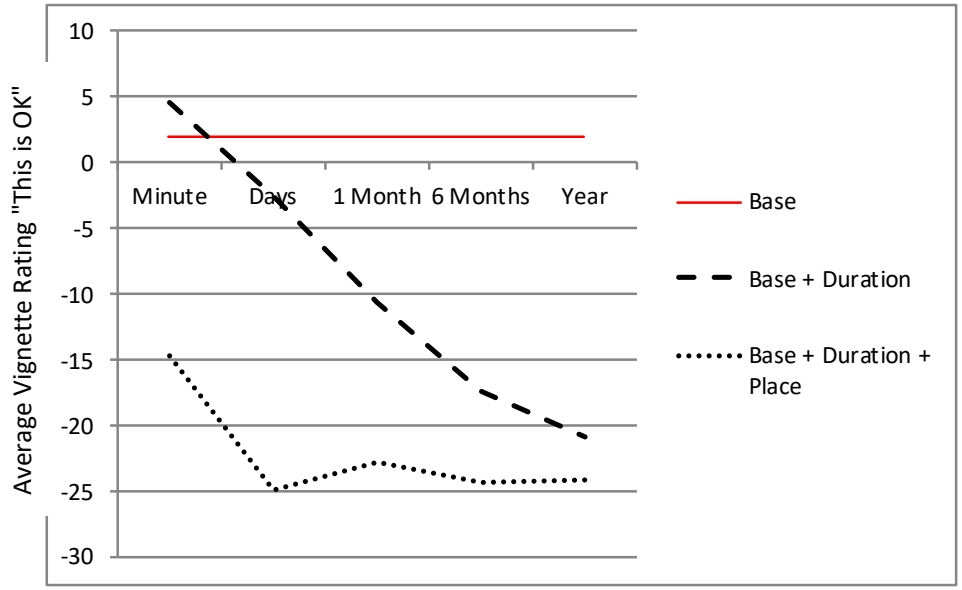
**Figure B6: Average Vignette Rating by Source for Place Condition**



*4. Duration*

Finally, we can isolate the importance of duration when place is included in the vignette.

**Figure B7: Importance of Duration when Adding Place**




In sum, we found the following:

- Adding place to a generic location negatively impacts the degree the scenario is "okay" overall, particularly for the FBI and the city services as actors.
- In addition, the importance of duration is diminished if the place inferred is included. This suggests that "place" explains what respondents were worried about.
- Adding duration and place significantly decreases the degree the scenario is "okay" for three sources:
    - Mapping App: +25 → -19
    - Social Media: +17 → -16
    - Phone: +8 → -25

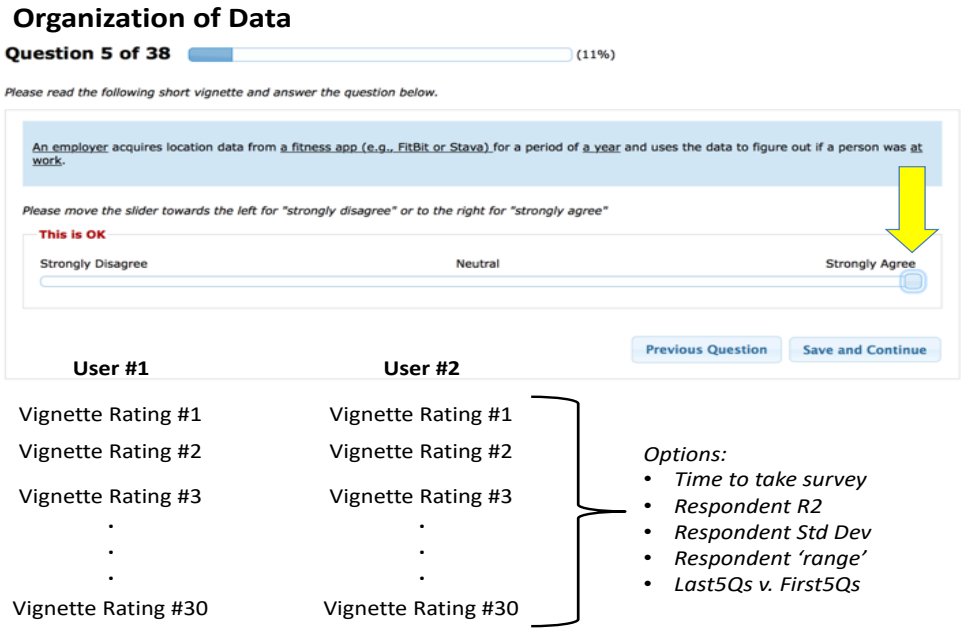
## APPENDIX C – QUALITY OF SAMPLES

The main survey was deployed through KnowledgeNetworks for a nationally-representative sample. Approximately 1,500 respondents took one of three possible vignette surveys. KnowledgeNetworks is an online research panel representative of the entire U.S. population. KnowledgeNetworks panel members are randomly recruited through probability-based sampling. Households are provided with access to the internet and hardware if needed.

At the same time, the survey was deployed through Amazon's Mechanical Turk where 1,200 respondents rated a total of 12,600 vignettes; 43% were female and 39% were over thirty-five years old. The sample was United States-only and each respondent was paid $1.70 for taking the survey.

In a separate survey on privacy expectations for websites, Kirsten Martin has compared results from Amazon Mechanical Turk with results from a nationally representative sample from KnowledgeNetworks. The survey from the Amazon Mechanical Turk sample produces the same theoretical generalizations as the survey from the KnowledgeNetworks survey, illustrating the ability to build generalizable theory from Amazon Mechanical Turk samples in online privacy studies.[128]

**Figure C1**




128. Martin, *Privacy Notices as Tabula Rasa*, *supra* note 90, at 16; Martin, *The Penalty for Privacy Violations*, *supra* note 90, at 108.

The sample was analyzed for nonresponsive respondents. Since the respondents each rated thirty independently-generated vignettes, the pattern of their rating on a sliding scale of -100 to +100 for each vignette could be analyzed. We marked two types of surveys as unresponsive: those that rated over twenty of the thirty vignettes as "0" (never moved the slider) and those that rated over twenty-five vignettes at one of the end points (moved the slider to one end almost every time). For the KnowledgeNetworks sample, this resulted in 10% of Survey 1 respondents, 13% of Survey 2 respondents, and 16% of Survey 3 respondents being removed from the pool. The number of respondents discarded from non-responsive ratings. For the Amazon Mechanical Turk sample, 2% of Survey 1 respondents, 5% of Survey 2 respondents, and 11% of Survey 3 respondents were found to be unresponsive. The Amazon Mechanical Turk sample was higher quality than the KnowledgeNetworks sample with the same theoretically generalizable findings.

**Table C1**

| | | | | EndPts > 20 | EndPts > 25 | 0s > 15 | 0s > 20 |
|---|---|---|---|---|---|---|---|
| | **N** | **Bad Resp** | **Very Bad Resp** | **Bad EndPts** | **Very Bad EndPts** | **Bad 0s** | **VeryBad 0s** |
| **Mechanical Turk** | | | | | | | |
| Base | 396 | 5.5% | 2.3% | 5.5% | 2.3% | 0.0% | 0.0% |
| Duration | 407 | 5.6% | 5.4% | 10.2% | 5.4% | 0.2% | 0.2% |
| Inference | 400 | 21.6% | 11.3% | 21.1% | 11.3% | 0.5% | 0.0% |
| **Knowledge Networks** | | | | | | | |
| Base | 502 | 19% | 10% | 11.0% | 3.7% | 10.8% | 7.6% |
| Duration | 524 | 22% | 13% | 13.7% | 6.4% | 12.1% | 8.8% |
| Inference | 509 | 30% | 16% | 23.1% | 11.5% | 9.1% | 6.7% |

**Table C1 (continued)**

| M Turk | N | Average OKDV | Female | Age Over 35 | Trust Business | Privacy Important | Trust Gov't | Trust Disposition | Authoritarian Scale | Trust Scale |
|---|---|---|---|---|---|---|---|---|---|---|
| Base | 396 | -23.25 | 46% | 43% | 3.58 | 74.78 | -16.49 | 33.82 | -16.42 | 6.97 |
| Duration | 407 | -27.40 | 38% | 47% | 3.73 | 73.30 | -18.21 | 31.22 | -15.38 | 5.58 |
| Inferen | 400 | -51.18 | 47% | 46% | 5.23 | 78.78 | -26.63 | 34.73 | -22.26 | 4.45 |

| KN w/o Bad | N | Average OKDV | Female | Age Over 35 | Trust Business | Privacy Important | Trust Gov't | Trust Disposition | Authoritarian Scale | Trust Scale |
|---|---|---|---|---|---|---|---|---|---|---|
| Base | 407 | -26.64 | 49% | 75% | 3.22 | 71.94 | -20.93 | 37.77 | 7.42 | 6.54 |
| Duration | 408 | -33.95 | 54% | 74% | 4.02 | 69.62 | -20.51 | 36.22 | 7.60 | 7.81 |
| Inference | 356 | -41.83 | 50% | 73% | 4.41 | 69.75 | -23.88 | 35.99 | 2.09 | 5.73 |

| KN w/o Very Bad | N | Average OKDV | Female | Age Over 35 | Trust Business | Privacy Important | Trust Gov't | Trust Disposition | Authoritarian Scale | Trust Scale |
|---|---|---|---|---|---|---|---|---|---|---|
| Base | 453 | -29.70 | 49% | 75% | 2.85 | 72.32 | -23.08 | 36.40 | 5.57 | 5.44 |
| Duration | 455 | -36.60 | 53% | 75% | 2.53 | 70.97 | -22.93 | 35.52 | 6.83 | 5.94 |
| Inference | 427 | -46.30 | 50% | 75% | 4.10 | 72.14 | -27.67 | 35.88 | 2.76 | 4.57 |

Figure C2: Theoretical Generalizations