HARVARD JOURNAL OF LAW & TECHNOLOGY

VOLUME 31, NUMBER 1 FALL 2017

CONTENTS

ARTICLE

PRIVACY INTERESTS IN PUBLIC RECORDS: AN EMPIRICAL INVESTIGATION

*Kirsten Martin and Helen Nissenbaum**

TABLE OF CONTENTS

---

* Respectively, Associate Professor, Strategic Management and Public Policy, George Washington University School of Business, Professor of Information Science, Cornell Tech, and, Media, Culture and Communication and Computer Science, New York University and Affiliated Professor, NYU School of Law.

The authors would like to thank the participants of the 2016 Privacy Law Scholars Conference for their helpful comments on this paper, in particular Danielle Citron, David Gray, and Felix Wu for their careful reading of an earlier draft. The paper benefited enormously from an outstanding research assistant, Eliana Pfeffer. We were honored to have had our study selected for the *Future of Privacy Forum,* 2016 Privacy Papers for Policy Makers. We are grateful for support from the National Science Foundation under Grant No. 1311823 and No. 1649415. 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.

I. INTRODUCTION

The construct of an information dichotomy has played a defining role in regulating privacy: information deemed private or sensitive typically earns high levels of protection, while lower levels of protection are accorded to information deemed public or non-sensitive. The information dichotomy construct is compelling in a model of privacy regulation where information is accorded differential treatment depending on whether it is deemed private or public. Challenging this approach, the theory of contextual integrity has linked privacy with more complex ontologies of information. Contextual integrity is a normative framework for evaluating the transmission of information between different actors, and it identifies information type as only one of several key variables that both shape people's privacy expectations and underpin privacy's normative foundations. Other contextual variables include key actors — information subjects, sources, and, most importantly, recipients — as well as the circumstances under which information is transmitted, such as "with subjects' consent," "bought and sold," "required by law," "with a warrant," and so forth. Our prior work revealed the systematic impact of these other variables on the privacy assessments of the release of so-called "private information," thereby undercutting the explanatory monopoly of the private-public dichotomy.[1]

1. *See* Kirsten Martin & Helen Nissenbaum, *Measuring Privacy: an Empirical Test Using Context to Expose Confounding Variables*, 18 COLUM. SCI. & TECH. L. REV. 176, 214–15 (2017).

Despite the importance of these factors, countless surveys conducted over approximately the last four decades have ignored their complex interactions, which systematically affect public attitudes toward privacy, resulting in findings that are, at best, skewed, and at worst, misleading. For example, Pew Research found significant consistency across individual respondents in their ratings of the degree of sensitivity of a range of information types presented to them.[2] Findings such as Pew's suggest that privacy protection can be modulated in accordance with the appraised sensitivity of information. However, as we found in previous work, people's judgments about the degree to which their privacy expectations are met depend on much more than the type of information in question; instead, no matter what the information type, the respondents' judgments in our studies were highly sensitive to other contextual parameters such as the recipients of the information, the terms under which the information had been shared, and the uses to which it had been put.[3]

To complement our previous work, this Article shifts away from the class of purportedly sensitive information and considers its opposite — that is, information deemed public. It reports on a second series of studies in which we ask subjects to respond to questions about information deemed public, consequently deserving less privacy protection, or possibly not implicating privacy at all.

The approach we take here is parallel to that taken in our previous work except that this time the focus is on information that is treated as public in that no explicit, legal restraints are placed on its retrieval, dissemination, and subsequent use. For the sake of analytic clarity, we have divided public information into two rough classes: (1) information gathered for and held in public records;[4] and (2) information casually observed in public spaces, for example, that Sally was walking hand-in-hand with Jake in Washington Square Park on October 2, 2015, or that my neighbor's shopping cart contained three boxes of Rice Krispies. The studies reported in this Article focus on class (1), though we note that important questions concerning class (2) have been raised in recent U.S. Supreme Court cases involving warrantless surveillance enabled by location-tracking devices conducted "in public."[5]

---

2. *See Public Perceptions of Privacy and Security in the Post-Snowden Era,* PEW RES. CTR., 6–7 (Nov. 12, 2014), http://www.pewinternet.org/files/2014/11/PI_PublicPerceptionsofPrivacy_111214.pdf [https://perma.cc/H2AB-BECF].

3. *See* Martin & Nissenbaum, *supra* note 1. This first study also showed the limited utility of Westin's privacy categories of "privacy pragmatist," "privacy fundamentalist," and "privacy unconcerned" in explaining privacy expectations of respondents.

4. Our focus is on law and policy in the United States, generally. We have pointed out areas where there is significant variation in local jurisdictions.

5. *See, e.g.*, United States v. Jones, 565 U.S. 400, 419 (2012) (Alito, J., concurring) (permitting short-term monitoring of a person's movements in public, but noting that technolog-

The focus on information deemed public is largely due to the changes in the treatment of information. In a wave of open data initiatives, the federal government as well as state and local municipalities moved to make government records open and machine-readable, thereby rendering the data more accessible to the public.[6] On the face of it, such initiatives should be unassailable as not only are such records generated by public officials and paid for by public funds, but they are already open to the public. Open data initiatives are also seen as furthering economic development and innovation,[7] promoting civic engagement,[8] and creating value for commercial interests.[9]

Despite the enthusiasm for making government records more open, concerns have been raised about their ease of access.[10] For example, in 2015, security researcher Chris Vickery drew surprised reactions when he noticed 191 million voter records with identification of gun ownership available online for anyone to access.[11] Due to these concerns, some scholars have started to ask whether and how to make government data open[12] and what internal resources and factors im-

---

ical advances means that long-term tracking may provide such a comprehensive record that the surveillance would run afoul of the Fourth Amendment's reasonableness clause).

6. *See, e.g.*, *Open Government Initiative | White House,* WHITEHOUSE.GOV (Sep. 16, 2017, 6:39 PM), https://www.whitehouse.gov/embeds/footer [https://perma.cc/UC5L-VWL8]; *The White House Open Data Innovation Summit and Solutions Showcase,* DATA.GOV (Sep. 16, 2017, 6:40 PM), https://www.data.gov/event/white-house-open-data-innovation-summit/ [https://perma.cc/ZPD6-KP7J]; *State of New York | Open Data,* STATE OF NEW YORK (Sep. 16, 2017, 6:41 PM), https://data.ny.gov/ [https://perma.cc/C67J-QMJB]; *City of Chicago | Data Portal,* CITY OF CHICAGO (Sep. 16, 2017, 6:42 PM), https://data.cityofchicago.org/ [https://perma.cc/X4MZ-RM7J]; *Seattle | Open Data,* CITY OF SEATTLE (Sep. 19, 2017, 7:52 PM), https://data.seattle.gov/ [https://perma.cc/2JGM-83VZ]; Jan Whittington et al., *Push, Pull, and Spill: A Transdisciplinary Case Study in Municipal Open Government,* 30 BERKELEY TECH. L.J. 1899, 1899 (2015).

7. *See* Brett Goldstein, *Preface* to BEYOND TRANSPARENCY: OPEN DATA AND THE FUTURE OF CIVIC INNOVATION, at ix (Brett Goldstein & Lauren Dyson eds., 2013).

8. *See* Maxat Kassen*, A Promising Phenomenon of Open Data: A Case Study of the Chicago Open Data Project*, 30 GOV'T INFO. Q. 508, 508 (2013).

9. *See* JOEL GURIN, OPEN DATA NOW: THE SECRET TO HOT STARTUPS, SMART INVESTING, SAVVY MARKETING, AND FAST INNOVATION 10 (2014).

10. *See, e.g.*, ROB KITCHIN, THE DATA REVOLUTION: BIG DATA, OPEN DATA, DATA INFRASTRUCTURES AND THEIR CONSEQUENCES 49 (2014); Amy Harmon, *As Public Records Go Online, Some Say They're Too Public,* N.Y. TIMES (Aug. 24, 2001), http://www.nytimes.com/2001/08/24/nyregion/as-public-records-go-online-some-say-they-re-too-public.html (last visited Nov. 14, 2017). These questions were addressed extensively in two recent conferences: the *Responsible Use of Open Data in Government and the Private Sector* conference at New York University in November 2015 and the *Open Data: Addressing Privacy, Security, and Civil Rights Challenges Symposium* at the University of California, Berkeley in April 2015.

11. *See* Dell Cameron & Kate Conger, *GOP Data Firm Accidentally Leaks Personal Details of Nearly 200 Million American Voters*, GIZMODO (Jun. 19, 2017, 8:00 AM), https://gizmodo.com/gop-data-firm-accidentally-leaks-personal-details-of-ne-1796211612 [https://perma.cc/EMC6-RRQR].

12. *See* Anneke Zuiderwijk & Marijn Janssen, *Towards Decision Support for Disclosing Data: Closed or Open Data?*, 20 INFO. POLITY 103, 104 (2015).

pact the decision to make data open.[13] Others have warned that public acceptance of open data initiatives will hinge on the appropriate use of data and ensuring that privacy interests are respected.[14]

Our work speaks to this last inquiry, assessing what guardrails are needed for open data initiatives that involve records containing information about identifiable individuals.[15] In particular, our work reveals normative judgments on the appropriate use and access of personal data in a broad array of public records, such as those of births, deaths, and marriages,[16] as well as documented transactions with offices and government agencies. These transpire, for example, when obtaining professional, vehicle, and firearm licenses, when acquiring ownership of real property, and when embroiled with various arms of the justice system, such the courts and law enforcement. Depending on the state of one's residency, this personal data may extend to voting records (i.e., whether you are registered or have voted in a given election)[17] and political party registration. According to Daniel Solove, the system of public records has grown rapidly since the mid-twentieth century both in terms of precisely what records are public and the modes of availability of these records to members of the public,[18] though it is important to note that there is no uniformity from jurisdiction to jurisdiction.

The study reported here is the second in a series of empirical challenges we have posed to traditional approaches to privacy, which were based on the idea that information can be placed into two buckets — private versus public — and, further, that people's actual and reasonable privacy expectations map neatly onto this dichotomy. In our previous study — of information deemed "sensitive" — we demonstrated empirically that this was not so.[19] Here, the focus is on data that would be deemed public according to traditional approaches.

---

13. *See* Peter Conradie & Sunil Choenni*, On the Barriers for Local Government Releasing Open Data*, 31 GOV'T INFO. Q. S10, S10 (2014).

14. *See* Teresa Scassa, *Privacy and Open Government*, 6 FUTURE INTERNET 397, 402–05 (2014); Vishanth Weerakkody et al., *Open Data and Its Usability: An Empirical View from the Citizen's Perspective*, 19 INFO. SYS. FRONTIERS 285, 297 (2017) ("The current practice of promoting open data as a means to improve transparency in government seems to be working, especially when it comes to citizens' perception on risk regarding the potential use of open data, as most citizens seem to have no concerns regarding the use of open data").

15. Our inquiry does not extend to records obtainable under the Freedom of Information Act ("FOIA"), which might include personal information, mainly to limit complexity by reducing scope and to focus on records deemed "public" within open data initiatives.

16. Some states allow for marriage records to be confidential.

17. For example, a number of high-profile political figures associated with President Trump were found to be registered in two states through querying voting records, causing embarrassment. *See e.g.,* Erin McCann, *Who Is Registered to Vote in Two States? Some in Trump's Inner Circle*, N.Y. TIMES (Jan. 27, 2017), https://www.nytimes.com/2017/01/27/us/politics/trump-cabinet-family-voter-registration.html (last visited Nov. 14, 2017).

18. Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137, 1142–49 (2002).

19. *See* Martin & Nissenbaum, *supra* note 1, at 39–40.

The salient subcategory that we examine is data held in government public records, by definition deemed public and by parallel assumption deemed not worthy of privacy protection. Against this line of reasoning, our study poses the following questions:

(1) How do type of information, source of information, and context of use affect respondents' judgments as to whether it is appropriate to access and use data from public records in specified ways?[20]

(2) Do differences in age and across gender correlate with respondents' normative judgments as to whether it is appropriate to access and use data from public records?

(3) Do respondents' assessments of how difficult it is to access public records affect their judgments?

In conducting the study, we asked respondents two main types of questions. First, we used a factorial vignette survey, asking respondents to rate (on a range from "Definitely Not OK" to "Definitely OK") the appropriateness of a series of scenarios in which contextual elements were systematically varied. These elements included: data recipient, the type of information held in public records in question, and the immediate sources of the information. The survey was deployed through Amazon's Mechanical Turk, which allowed 992 respondents (47% female and 59% under 35 years old) to rate a total of 39,680 vignettes. Then, we asked respondents to rate how easy/hard they believed it to be to access the four types of information under study: voting, marriage, criminal, and property records.

We found that:

(1) The degree to which information is thought to be accessible does not drive judgments about the appropriateness of accessing that information. In other words, even for information that was deemed easy to access (marital status), respondents still judged it to be inappropriate ("Not OK") to access it under certain circumstances.

(2) The immediate source of information matters to the perceived appropriateness of the data flows, even for information contained in public records. For example, respondents consistently found it inappropriate when data brokers were the immediate sources of information.

---

20. The choice of factors is guided by the theory of contextual integrity, which postulates privacy expectations are formed by the combination of three parameters: actors (subject, sender, recipient), information type, and transmission principles. In this Article we refer to the sender as the source.

(3) All else being equal, respondents were most opposed to accessing voting records across all vignette scenarios (but with the greatest amount of variance across respondents) and least opposed to accessing information about criminal records across all scenarios.

(4) Younger respondents (under 35 years old) were more critical of seeking access to data from data brokers and online government records than of seeking access by asking data subjects directly (the null condition).

(5) Women were more opposed than men to the use of marital status in job applications.

Our findings indicate that the "public records" designation conflates several orthogonal dimensions, such as information type and terms of access, that make a difference in how individuals judge the appropriateness of access to and use of public records data. Although popular opinion is but one determinant of legitimate privacy expectations, consistent findings such as the ones revealed here suggest that there is a need for a careful reexamination of policies surrounding public records and open data initiatives, particularly in light of the recent advancements in digital technologies of aggregation, linkage, and analytics, as well as artificial intelligence.

II. BACKGROUND AND RELATED WORK

A. Public Information

One difficulty in conceptualizing "privacy in public" is the association of the word "privacy" with information that is inaccessible to others.[21] If privacy is that which is not disclosed or utterly obscure, and if public means being accessible, then something is either private or public and cannot be both. The dichotomy that follows from this — of information being secret-or-not or private-or-not[22] — leads to the incorrect conclusion "that there is no claim to privacy when information appears in a public record."[23] Michael Zimmer has observed that this conclusion is pervasive as a defense against troubling practices of information collection and dissemination: "but the data is al-

21. *See* Helen Nissenbaum, *Protecting Privacy in an Information Age: The Problem of Privacy in Public*, 17 LAW AND PHIL. 559, 566–69 (1998).

22. *See* HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 113 (2010); Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CAL. L. REV. 1, 17–20 (2013).

23. Solove, *supra* note 18, at 1140.

ready public."[24] However, critical examinations of how different parties conceive of information "out in public" have focused not only on the privacy expectations of individuals,[25] but also on the possible harm that can come from deployment of seemingly "public" information[26] and the appropriate norms for such information.[27]

Our studies align with the work of others who have drawn attention to privacy interests in public information and public spaces. Those focusing on public spaces question the presumption that information is deemed "up for grabs" simply because the surrounding space in which it has been gathered is "public," open, or readily visible to other people.[28] Consistent with the Fourth Amendment concept of "plain view,"[29] Ryan Calo notes that "[t]he law's approach to privacy in public is monolithic: it generally refuses to see a privacy violation where the observation takes place in public on the theory that people in public have no reasonable expectation of privacy."[30] Public opinion on this matter will not be addressed in the study we discuss below, but will be the subject of a follow-up study, briefly anticipated in the conclusion of this Article.[31]

Others have questioned the status of information gleaned from public records,[32] that is, records collected by government agencies and

---

24. Michael Zimmer, *"But the Data Is Already Public": On the Ethics of Research in Facebook*, 12 ETHICS AND INFO. TECH. 313, 313, 318 (2010).

25. Pew Research Center, *supra* note 2, at 1.

26. *See* M. Ryan Calo, *The Boundaries of Privacy Harm*, 86 IND. L.J. 1132, 1135 (2011); Neil M. Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1962–63 (2013); Julie E. Cohen, *Privacy, Visibility, Transparency, and Exposure*, 75 U. CHI. L. REV. 181, 181 (2008).

27. *See* Nissenbaum, *supra* note 22, at 237; Kirsten Martin, *Understanding Privacy Online: Development of a Social Contract Approach to Privacy*, 137 J. BUS. ETHICS 551, 551 (2016).

28. *See, e.g.*, Philip Brey, *Ethical Aspects of Facial Recognition Systems in Public Places*, 2 J. INFO., COMM. & ETHICS IN SOC'Y 97, 105 (2004); Hartzog & Stutzman, *supra* note 22, at 18–19; Woodrow Hartzog & Evan Selinger, *Surveillance as Loss of Obscurity*, 72 WASH. & LEE L. REV. 1343, 1352–55, 1377 (2015).

29. *See, e.g.*, Coolidge v. New Hampshire, 403 U.S. 443, 465 (1971) (plurality opinion).

30. Calo, *supra* note 26, at 1155.

31. A related, important subset of public data is data made openly available to researchers, commercial interests, and citizens, and hopefully de-identified prior to release. Such data commons are sometimes referred to as open data when the government makes the data open to all, but firms can also make data available, as demonstrated by the cases of Netflix and AOL search query data. Scholarship has also focused on specific misuse of data made public, such as when Acquisti and Gross predict identifiable information from "public data," Social security's Death Master File, as well as information from data brokers or on social networking sites. *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1704, 1705–06 (2009); Alessandro Acquisti & Ralph Gross, *Predicting Social Security Numbers from Public Data*, 106 PROC. NAT'L ACAD. SCI. 10975, 10978–79 (2009). *But see* Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J.L. & TECH. 1, 31–35 (2011); Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 U. COLO. L. REV. 1117, 1129, 1135 (2013).

32. *See* Whittington et al., *Push, Pull, and Spill: A Transdisciplinary Case Study In Municipal Open Government*, 30 BERKELEY TECH. L.J. 1899, 1900 (2015).

either explicitly declared public or implicitly presumed public due to the absence of explicit legal constraints. Robert Gellman[33] and Daniel Solove have written eloquently about these records:

> States maintain records spanning an individual's life from birth to death, including records of births, marriages, divorces, professional licenses, voting information, worker's compensation, personnel files (for public employees), property ownership, arrests, victims of crime, criminal and civil court proceedings, and scores of other information . . . . These records contain personal information including a person's physical description (age, photograph, height, weight, eye color); race, nationality, and gender; family life (children, marital history, divorces, and even intimate details about one's marital relationship); residence, location, and contact information (address, telephone number, value and type of property owned, description of one's home); political activity (political party affiliation, contributions to political groups, frequency of voting); financial condition (bankruptcies, financial information, salary, debts); employment (place of employment, job position, salary, sick leave); criminal history (arrests, convictions, traffic citations); health and medical condition (doctors' reports, psychiatrists' notes, drug prescriptions, diseases and other disorders); and identifying information (mother's maiden name, Social Security number). [34]

Woodrow Hartzog and Frederic Stutzman, in their work on the virtues of obscurity, have highlighted the practical importance of a mere shift in medium and access modality. They observe that public records, for many years, were protected by practical obscurity; individuals had a recognized "privacy interest in information that was technically available to the public, but could only be found by spending a burdensome and unrealistic amount of time and effort in obtaining it."[35] When public records were kept in a paper file within a government building, they were practically inaccessible and could not be linked to information about the individual held by other data

33. *See* Robert Gellman, *Public Records — Access, Privacy, and Public Policy: A Discussion Paper*, 12 GOV'T INFO. Q. 391, 393–95 (1995).

34. *See* Solove, *supra* note 18, at 1139.

35. *See* Hartzog & Stutzman, *supra* note 22, at 21.

sources.[36] The degree of access to these records was vastly increased with their transition to a digital medium and further amplified by their placement online.[37]

In expressing their concern, some commentators have highlighted *government* data mining practices,[38] while others have emphasized public, commercial aggregators. Benefitting from the actions of these parties, commercial stakeholders, such as data brokers, enjoy greater efficiencies in their bulk collection of data from public records, from which they extract knowledge that is attractive to other stakeholders in various sectors (e.g. marketing, finance, etc.). Daniel Solove[39] and Chris Hoofnagle[40] have discussed some of the dangers to individuals caused by increased access, aggregation, and sale of information from public records. Solove writes,

> Consolidating various bits of information, each in itself relatively unrevealing, can, in the aggregate, begin to paint a portrait of a person's life. I refer to this as a "digital biography." A growing number of private sector organizations are using public records to construct digital biographies on millions of individuals. I argue that we should be concerned about the ways in which our digital biographies are being used.[41]

36. Court records constitute a prime example of the difficulty in making public records less obscure. *See, e.g.*, Kristen M. Blankley, *Are Public Records Too Public —Why Personally Identifying Information Should Be Removed from Both Online and Print Versions of Court Documents*, 65 OHIO ST. L.J. 413, 413–16 (2004); Victoria S. Salzmann, *Are Public Records Really Public: The Collision Between the Right to Privacy and the Release of Public Court Records over the Internet*, 52 BAYLOR L. REV. 355, 375–76 (2000); Peter A. Winn, *Online Court Records: Balancing Judicial Accountability and Privacy in an Age of Electronic Information*, 79 WASH. L. REV. 307, 311–14 (2004); David S. Ardia & Anne Klinefelter, *Privacy and Court Records: An Empirical Study*, 30 BERKELEY TECH. L.J. 1807, 1824–28 (2015); Amanda Conley et al., *Sustaining Privacy and Open Justice in the Transition to Online Court Records: A Multidisciplinary Inquiry*, 71 MD. L. REV. 772, 774–75 (2012).

37. *See* Conley, *supra* note 36, at 773–75.

38. *See, e.g.*, Fred H. Cate, *Government Data Mining: The Need for a Legal Framework*, 43 HARV. C.R.-C.L. L. REV. 435, 436 (2008).

39. Solove, *supra* note 18, at 1190, 1193, 1196–97.

40. Chris Jay Hoofnagle, *Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement*, 29 N.C. J. INT'L L. & COM. REG. 595, 595–96, 636–37 (2003).

41. Solove, *supra* note 18, at 1141.

B. Privacy as Contextual Integrity

According to the theory of contextual integrity, protecting privacy means ensuring that information flows appropriately.[42] Whether information flow is appropriate depends on whether it conforms to legitimate and contextual informational norms. These norms prescribe information flows in terms of three parameters: actors (sender, subject, recipient), information types, and transmission principles. When confronted with particular information flows, we judge them as respecting or violating privacy according to whether they conform to expectations of flow within a given context. When this is the case, we can say that contextual integrity has been preserved. When this is not the case, frequently when novel technologies are introduced that disrupt entrenched flows, the prima facie case exists for concluding that contextual integrity has been violated and privacy infringed.

One immediate consequence of defining informational privacy as contextual integrity can be observed in the approach to privacy of public data. Privacy is not lost, traded off, given away, or violated simply because control over information is ceded or because information is shared or disclosed, only if ceded or disclosed *inappropriately*. Releasing information is not the same as giving up privacy if the flow is appropriate. Buying a house or filing a tax return with the Internal Revenue Service does not amount to giving up privacy, only to sharing information. Privacy as contextual integrity, therefore, would imply that individuals will have normative judgments as to the appropriateness of the information flows of information contained in public records.

III. METHODS

A. General Methods

We used factorial vignette survey methodology to investigate what is deemed to be the appropriate use of information contained in public records.[43] Factorial vignette surveys present respondents with a series of vignettes comprised of several sentences that contain factors

---

42. *See* Helen Nissenbaum, *A Contextual Approach to Privacy Online*, 140 DAEDALUS 32, 37–39 (2011), http://www.mitpressjournals.org/doi/pdf/10.1162/DAED_a_00113 [https://perma.cc/7DF5-6N9Z]; *see also* Helen Nissenbaum, *Respecting Context to Protect Privacy: Why Meaning Matters*, SCI. & ENGINEERING ETHICS, July 15, 2015, at 4–10, http://www.nyu.edu/projects/nissenbaum/papers/Respecting%20Context%20to%20Protect%20Privacy%20Why%20Meaning%20Matters.pdf [https://perma.cc/Q4FX-MADE].

43. Guillermina Jasso, *Factorial Survey Methods for Studying Beliefs and Judgments*, 34 SOC. METHODS & RES. 334, 340–41 (2006); *see also* Steven L. Nock & Thomas M. Guterbock, *Survey Experiments*, *in* HANDBOOK OF SURVEY RESEARCH 837 (Peter V. Marsden & James D. Wright eds., 2d ed. 2010).

relevant to the judgment; these vignette factors are the independent variables and are systematically varied. Unlike survey research which presents a respondent with a single vignette and poses a number of questions about it, our study presents all respondents with forty vignettes and asks respondents to complete the same rating task for each vignette. This method allows the researcher to vary contextual factors simultaneously within each vignette while relying on a single judgment for the question or rating task — namely, the degree to which a scenario is appropriate, or "OK." In our study, the factors are information type, source, and recipient. The responses to each vignette permit researchers to measure how these contextual factors affect respondents' judgments.

We deployed the factorial vignette survey methodology to address our three research questions:

(1) How do the factors, type of information, source[44] of information, and context of use, affect respondents' judgments as to whether it is appropriate to access and use data from public records in specified ways?

(2) Do differences in age and across gender correlate with respondents' normative judgments about public records information?

(3) Do respondents' assessments of how difficult it is to access public records affect their judgments?

The factorial vignette methodology has proven effective for addressing normative research questions which are notoriously difficult to study.[45] 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 that appear in traditional surveys. The first type of bias occurs when respondents adjust their answers in order to appear more ethical or concerned. However, because many factors are changing simultaneously in the factorial vignette survey, respondents are less likely to make such adjustments. Further, respondents may have difficulty identifying and articulating the reasons behind their judgments.[46] This too is alleviated with the factorial vignette methodology, as the results themselves show the researcher which factors moved the respondent's rating of a given vignette, so it

---

44. The choice of factors is guided by the theory of contextual integrity, which postulates that privacy expectations are formed by the combination of three parameters: actors (subject, sender, recipient), information type, and transmission principles. In this Article we refer to the sender as the source.

45. Jasso, *supra* note 43, at 410–11.

46. *See* Jonathan Haidt, *The Moral Emotions*, *in* HANDBOOK OF AFFECTIVE SCIENCES 852, 865–66 (Richard J. Davidson et al. eds., 2003); Chen-Bo Zhong, *The Ethical Dangers of Deliberative Decision Making*, 56 ADMIN. SCI. Q. 1, 4, 7 (2011).

is not necessary to directly ask the respondent for a prioritized list of factors considered.

Figure 1 provides a screenshot of what the respondents rated. Each of the vignette factors, which are underlined in Figure 1, was assigned a value. For example, the information type factor was assigned one of four values: marriage, voter, court, or property. The vignette factors, values, and sample vignettes are described below.

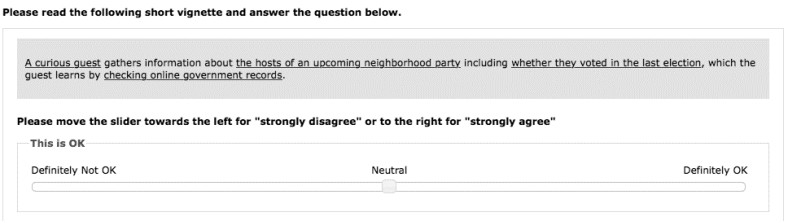

Figure 1: Vignette as Seen by Respondents

B. Vignettes

The vignettes each describe a realistic scenario that involves the collection and use of data found in public records. Each respondent was presented with forty vignettes, in which values for each of the three factors were systematically varied. The survey instrument varies the factors by randomly selecting values for each, thus generating vignettes in real time. For each generated vignette, the respondents were then asked to rate the degree to which the vignette was "OK."

1. Vignette Factors

We selected three factors: information type, recipient, and source. Based on the parameters that contextual integrity postulates are critical to the definition of information privacy norms, a complete operationalization of contextual integrity would have required five vignette factors, specific to a given context, namely: sender, subject, recipient, information type, and transmission principle. However, factorial vignette methodology requires balancing the need for realism, statistical validity, and maintaining a reasonable number of variables. Accordingly, we took deliberate steps to simplify the task by holding fixed certain variables, thereby limiting the survey to three factors, and by significantly reducing the scope of these variables. These modifications made the vignette task more accessible to our subjects while generating numbers that were sufficiently high to conduct the analysis.

As mentioned above, values were randomly assigned to the three factors (information type, recipient, and source) within each vignette. Figure 2 depicts the assignment process that generates the vignette, such as that seen in Figure 1, above.

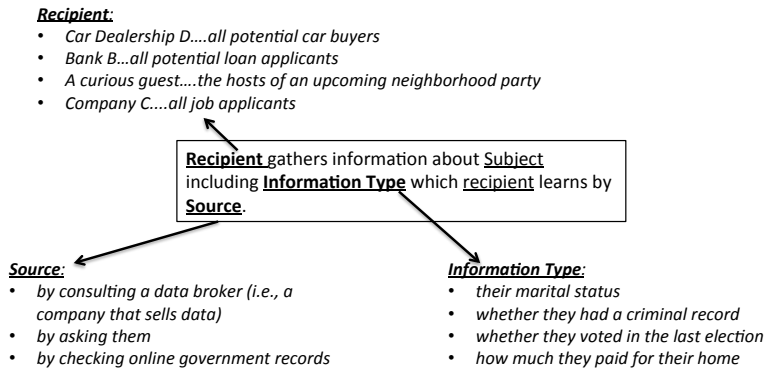

Figure 2: Sample Vignette and Vignette Factors

a. Information Type Factor

For the information-type factor, we selected marriage,[47] voter,[48] court,[49] and property records as values.[50]

47. *Cf.* Hoofnagle, *supra* note 40, at 601, 635; Solove, *supra* note 18, at 1139, 1143; Grayson Barber, *Personal Information in Government Records: Protecting the Public Interest in Privacy*, 25 ST. LOUIS U. PUB. L. REV. 63, 64 n.2 (2006).

48. *Cf. Statement on the Constitutionality of the Disclosure of Name and Address Information From Public Records Before the New Jersey Privacy Study Commission* (November 12, 2003) (statement of Fred H. Cate, Distinguished Professor, Indiana University School of Law-Bloomington), http://www.cspra.us/yahoo_site_admin/assets/docs/CateestimonyNJ.31984315.pdf [https://perma.cc/DW6T-JFMY]; Barber, *supra* note 47, at 83; Kwame N. Akosah, *Cracking the One-Way Mirror: How Computational Politics Harms Voter Privacy, and Proposed Regulatory Solutions*, 25 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1007, 1018 (2015); Daniel Kreiss, *Yes We Can (Profile You): A Brief Primer on Campaigns and Political Data*, 64 STAN. L. REV. ONLINE 70, 71 (2012).

49. *Cf.* Ardia & Klinefelter, *supra* note 36, at 1818–23; Blankley, *supra* note 36, at 413–16; Salzmann, *supra* note 36, at 359–61; Peter A. Winn, *supra* note 36, at 308, 310; Solove, *supra* note 18, at 1145–49; Conley et al., *supra* note 36, at 773–77.

50. *Cf.* Manya Sleeper et al., *I Know Where You Live: Analyzing Privacy Protection in Public Databases*, CARNEGIE MELLON UNIVERSITY 1, 4–7 (2011), https://www.cylab.cmu.edu/files/pdfs/tech_reports/CMUCyLab11015.pdf [https://perma.cc/BZY7-X6VH]; Solove, *supra* note 18, at 1139, 1145.

b. Recipient Factor

For the recipient factor, we selected actors whose roles clearly situated them within the respective social context, such as a bank gathering information on loan applicants, a car dealer gathering information on a potential customer, a company gathering information on a job applicant, and a guest gathering information on a party host.

Table 1: Vignette Factors and Levels

Factor	**Operationalized in Vignette**		
Information	Marriage Records	their marital status	
	Court Records	whether they had a criminal record	
	Voter Records	whether they voted in the last election	
	Property Records	how much they paid for their home	
Source	Data Broker	by consulting a data broker (i.e., a company that sells data)	
	Subject	by asking them	
	Online Records	by checking online government records	
		Subject	**Recipient**
Context	Retail	all potential car buyers	Car Dealership D
	Bank	all potential loan applicants	Bank B
	Social	the hosts of an upcoming neighborhood party	a curious guest
	Employment	all job applicants	Company C

c. Source Factor

For the source factor, we selected an online government record, a data broker, or the data subject him/herself.

2. Sample Vignette

The general template used and a sample vignette are shown below. Each respondent received forty vignettes and each vignette was created by assigning a value to each factor.

a. General Template

> *Recipient* gathers information about *Subject* including *Information Type* which *Recipient* learns by *Source*.

b. Examples

> *Bank B's loan officer* gathers information about *all loan applicants* including *whether they voted in the last election*, which *Bank B* learns by *checking online government records*.
>
> *Company C's recruiting manager* gathers information about *all job applicants* including *how much they paid for their home*, which *Company C* learns by *asking them*.

3. Vignette Rating Task

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

C. Respondent-Level Measures

1. Standard Controls

We captured the respondents' baseline disposition to trust by asking them to rate, on a scale of "strongly disagree" to "strongly agree," their level of agreement with the statement "[i]n general, I trust people until proven otherwise." We also asked them to evaluate the statement "[i]n general, I believe privacy is important."

2. Accessibility of Public Records

In order to measure how knowledgeable the respondents were about the accessibility of public records, we asked the following four questions:

(1) How easy is it to find out if someone is married, divorced, or single without asking them?

(2) How easy is it to find out someone's house value without asking them?

(3) How easy is it to find out if someone has a criminal record without asking them?

(4) How easy is it to find out when someone last voted without asking them?

The respondents were asked to rate the ease of accessing this information on a scale ranging from "Very Hard" to "Very Easy" or (1–5).

D. Sample

The survey was deployed through Amazon's Mechanical Turk where 992 respondents rated a total of 39,680 vignettes. Of the 992 respondents, 47% were female and 59% of all respondents were under the age of 35. The sample was US-only and each respondent was paid $2 for taking the survey. The survey took approximately 10–12 minutes to complete. Although the use of Mechanical Turk for survey deployment can be controversial,[51] studies have shown that mTurk workers are more representative of the United States population than other samples often used in social science research.[52] In fact, in a sep-

51. *See* Matthew Lease et al., *Mechanical Turk Is Not Anonymous*, SSRN 3 (2013), http://papers.ssrn.com/abstract=2228728 [https://perma.cc/UVU6-KWEZ]; *see also* Joel Ross et al., *Who are the Crowdworkers? Shifting Demographics in Mechanical Turk*, 28 ACM CONF. HUM. FACTORS COMPUTING SYS. 2863, 2864–65 (2010).

52. mTurk has been used for consumer perceptions in marketing. *See* Sybil Yang & Michael Lynn, *More Evidence Challenging the Robustness and Usefulness of the Attraction Effect*, 51 J. MARKETING RES. 508, 508–10 (2014); Daniel G. Goldstein et al., *The Economic and Cognitive Costs of Annoying Display Advertisements*, 51 J. MARKETING RES. 742, 748 (2014). In addition, a recent survey replicates (and extends) a Pew Research Study. *See* Mary Madden et al., *Public Perceptions of Privacy and Security in the Post-Snowden Era,* PEW RES. CTR. (Nov. 12, 2014), http://assets.pewresearch.org/wp-content/uploads/sites/14/2014/11/PI_PublicPerceptionsofPrivacy_111214.pdf [https://perma.cc/65D7-X8RL]. *See also* Martin & Nissenbaum, *supra* note 1, at 200, 202 (exploring privacy expectations around sensitive information on mTurk); Catherine E. Tucker, Social Networks, Personalized Advertising, and Privacy Controls, 51 J. MARKETING RES. 546 app. at 2 (2014) (highlighting that mTurk is particularly useful for representing consumers likely to be online).

arate survey on privacy expectations for websites, Kirsten Martin has compared results from Amazon Mechanical Turk with results from a nationally representative sample from KnowledgeNetworks (GfK). The survey results from the mTurk sample produced the same theoretical generalizations as did the survey from the KnowledgeNetworks (GfK) sample, illustrating the ability to build a generalizable theory from Mechanical Turk samples in online privacy studies.[53]

IV. RESULTS

A. Respondents' Assessments of the Difficulty of Accessing Information

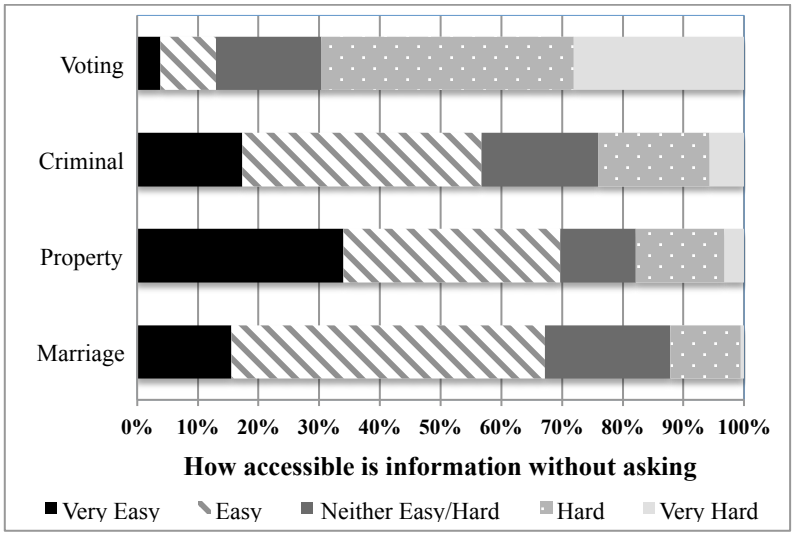

Figure 3: Frequency of Each Response —
Ease of Accessing Public Records as Perceived by Respondents

To measure the respondents' assessments of the perceived difficulty of accessing information normally found in public records without asking the subject of the record directly, respondents were asked to judge the accessibility of four types of public records. The results in Figure 3 illustrate great variability across the different information types. Voting records were judged to be difficult to access whereas a person's marital status was considered to be easy to access. Interestingly, while a majority found property value and marital status

53. 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, 216–17, 219–20 (2015).

easy/very easy to access, fewer than 15% of respondents judged it to be easy/very easy to access information regarding the subject's voting history.

To analyze the impact of the respondents' assessment of how difficult a record is to access on how they judged the appropriateness of gathering and using that data in the vignette, we considered two measures: (1) whether the information that was deemed "hard to access" in the initial set of questions was also judged to be "not appropriate to gather" in the contextualized vignettes; and (2) whether respondents who rated the public records as "easy" or "hard to access," respectively, differed in their judgments of the vignettes.

In general, data that is "hard to access" does not have a clear relationship with being "OK" or appropriate to gather across all scenarios. Table 2 has the average rating of appropriateness for each information type as well as how hard respondents felt it would be to access.

Table 2: Appropriateness Rating and Degree Respondent Perceived as Accessible

	Vignette	**Respondent Judgment**
	Average "OK"	% Hard to Access (Hard/Very Hard)
Voting	-29.81	69.65%
Criminal	+13.82	24.1%
Property	-5.61	17.9%
Marriage	+2.15	12.2%

Although the information judged most hard to access (voting) is also the information found in the vignettes to be least appropriate to gather and use, this relationship does not hold for criminal records, which are perceived by respondents as the second most difficult to access (24.1% hard/very hard), but also the most appropriate to gather (+13.82).

In general, we found that the relationship between difficulty to access and appropriateness is not linear. The degree of accessibility therefore does not explain the degree to which respondents judge the gathering and usage of information to be appropriate. These findings alone are sufficiently interesting to warrant future study.

B. Whether It Is Appropriate to Access Public Records

1. Factors Impacting Respondents' Judgment on Whether it is Appropriate to Access Information Contained in Public Records

In studying the perceived appropriateness of accessing public records, we examined how each receiver of information — *a bank* receiving a loan application, *a guest* inquiring about a party host, *a car dealer* gathering information about a potential customer, and *a company* with a job applicant — was judged based on the type of public record information accessed and the source of that information. For each receiver, the average appropriateness rating is graphed in Figure 4,

Figure 5, and Figure 6 based on the type of information and the source (sender) of the information. These graphs depict the degree to which a given scenario is deemed "Definitely OK" depending on the respective values assigned to each factor. The coefficients of the regression analysis of the degree to which a given scenario is judged to be "OK" on the vignette and respondent factors are also depicted in Table 3.

Table 3: Regression of Appropriateness of Scenario on Vignette and Respondent Factors

	Entire Sample	
	Regression of Degree Scenario is "OK" on Vignette and Respondent Factor	
	coef	**P**
Public Record Information		
Criminal Status	19.15	0.00
Marital Status	7.84	0.00
Voter Record	-24.73	0.00
(null = Property Value)		
Recipient		
Party Guest	-1.96	0.01
Bank	23.27	0.00
Employer	5.51	0.00
(null = Car Dealer)		
Source of Information		
DataBroker	-65.40	0.00
OnlineGovtRecords	-37.67	0.00

(null = Subject)		
Respondent Controls		
AgeUnder35	5.30	0.00
Female	-8.52	0.00
PrivacyImport	-0.31	0.00
TrustDisposition	0.04	0.01
_cons	55.64	0.00
Sample Statistics		
Ave	-4.80	
N	39,680	
ICC	20.3%	
sd(_cons)	26.59	

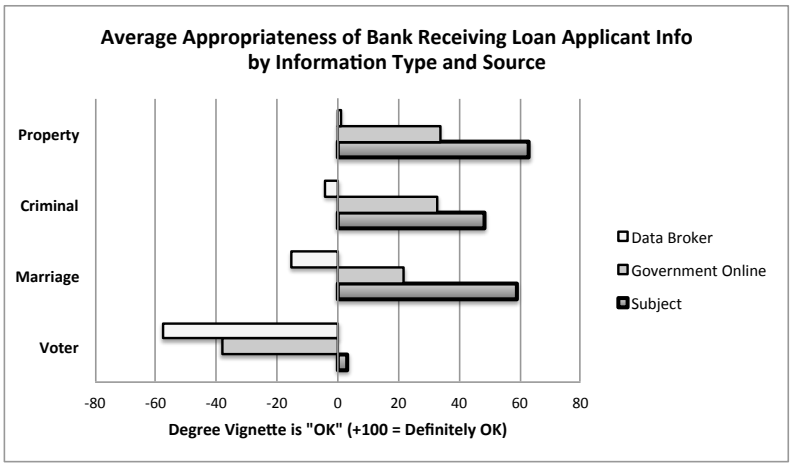

Figure 4: Appropriateness of Scenarios Depicting a Bank Receiving Loan Applications by Source and Information Type

Figure 4 shows the respondents' perceptions of appropriateness of different scenarios in which a bank receives information about a loan applicant. The graph shows that respondents differentiate between the different types of information and how the information is sourced. Whereas marriage records, criminal records, and property values are deemed appropriate given the situation, voter information is deemed the least appropriate for loan applications. In addition, respondents

consistently penalized those accessing the data through a data broker across information types.[54]

The same general trend regarding the source of information holds for the scenario of a company with a job applicant, where querying information from a data broker is judged to be less appropriate than accessing government records online. This appears to be the case for all information types. However, the type of information deemed appropriate differs from the bank scenario described above, as only criminal information is deemed appropriate for job applications, as seen in Figure 5.

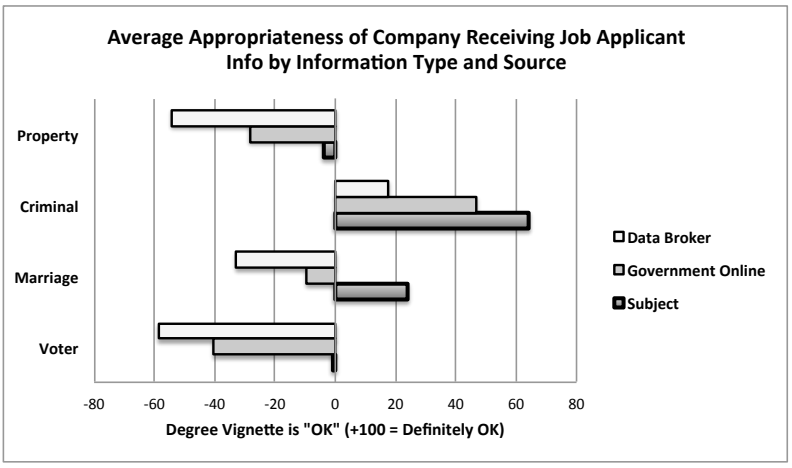

Figure 5: Appropriateness of Scenarios Depicting a Company Receiving Job Applications by Source and Information Type

54. This is also seen in Table 1for all information types: in general, respondents find scenarios less appropriate by -65.40 points when a data broker is the source. *See supra* Table 1.

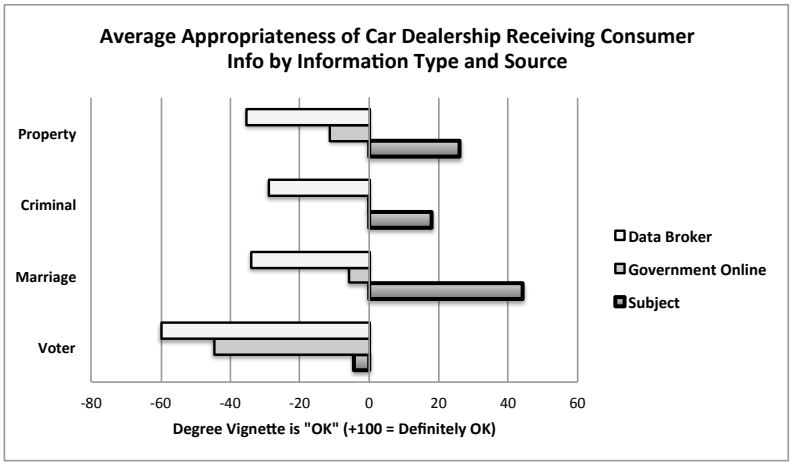

Figure 6: Appropriateness of Scenarios Depicting a Car Dealership Gathering Information on Potential Customers by Source and Information Type

Figure 6 shows that the trends for appropriate flow of information — including the source of the information and the information type — for car dealerships gathering information about a potential customer parallel the appropriateness of information flow for the bank and the employer. The similarities could suggest a common perception of the appropriateness of the use of information in the commercial space.

Finally, the situation in which a guest attempts to gather information about the party host differs slightly. While asking the party host questions about any information type is judged appropriate, accessing the information through an online government record or a data broker is deemed inappropriate across all information types for a guest, as seen in Figure 7.

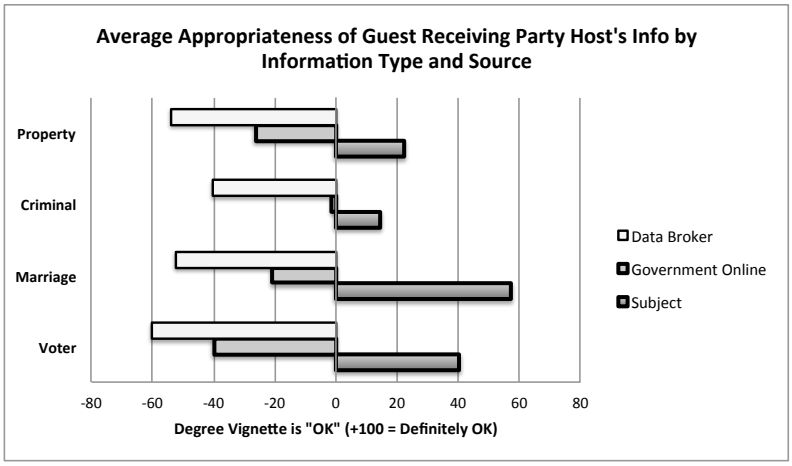

Figure 7: Appropriateness of Scenarios Depicting a Guest Gathering Information on Party Host by Source and Information Type

2. Respondents' Judgment by the Type of Public Record

Table 4 and Table 5 include the results of regressing the appropriateness of the scenario (degree to which the scenario in the vignette is deemed "OK") on the vignette factors and respondent attributes for each type of information (criminal, voter, marriage, and property record) resulting in the four conditions (A–D). By comparing the regression results across each condition, we can see how the relative importance of the vignette factors differs for each information type. The penalty of accessing information through a data broker versus querying government records online or asking the subject directly can be seen across information types. In fact, the largest impact on the respondents' perceptions of the appropriateness of information access is the source of the information. There is some variation in the level of importance, as the relative importance of data brokers being the source varies from -50.89 (criminal records) to -80.47 (marriage records) ($p < 0.01$).

There are several other noteworthy findings from the respondent controls shown in Table 4 and Table 5. The age of the respondent matters when examining the perceived appropriateness of accessing voting records (Condition B: whether they voted in the last election), with younger respondents judging accessing voting records to be more "OK" than older respondents. However, age was not a significant factor in judging other information types. In addition, women found accessing voting, marriage, and property records less appropriate on average. Additional analysis shows that women judge using marital

status in job application to be more problematic ($\beta = -18.50$, $p < 0.01$) than men ($\beta = -8.71$, $p < 0.01$).

Table 4: Regression Results of the Appropriateness of the Vignette ("Is this OK?") on the Vignette Factors and Respondent Controls: Conditions A and B

	Condition A		**Condition B**	
	Criminal Record — Whether they committed a crime		Voter Record — Whether they voted in last election	
	coef	**p**	**coef**	**p**
Recipient				
Party Guest	-6.03	0.00	16.38	0.00
Bank	29.33	0.00	3.63	0.00
Employment	47.07	0.00	2.21	0.08
(null = Car Dealer)				
Source				
DataBroker	**-50.89**	**0.00**	**-68.87**	**0.00**
OnlineGovtRecords	-17.29	0.00	-49.26	0.00
(null = Subject)				
Controls				
AgeUnder35	3.48	0.15	**13.84**	**0.00**
Female	-3.70	0.12	-9.94	0.00
PrivacyImport	-0.17	0.00	-0.48	0.00
TrustDisposition	0.04	0.06	0.02	0.41
Sample Statistics				
Ave	13.82		-29.81	
N	10,117		9,880	
ICC	26.51%		36.58%	
sd(_cons)	34.09		39.31	

Table 5: Regression Results of the Appropriateness of the Vignette ("Is this OK?") on the Vignette Factors and Respondent Controls: Conditions C and D

	Condition C		**Condition D**	
	Marriage Record — Marital status		Property Record — Value of house	
	coef	**p**	**coef**	**p**
Recipient				
Party Guest	-6.25	0.00	-13.20	0.00
Bank	19.63	0.00	37.68	0.00
Employment	-5.27	0.00	-23.34	0.00
(null = Car Dealer)				
Source				
DataBroker	**-80.47**	**0.00**	**-62.79**	**0.00**
OnlineGovtRecords	-50.98	0.00	-34.42	0.00
(null = Subject)				
Controls				
AgeUnder35	1.89	0.41	1.64	0.45
Female	**-10.11**	**0.00**	-11.91	0.00
PrivacyImport	-0.25	0.00	-0.32	0.00
TrustDisposition	0.07	0.00	0.04	0.04
Sample Statistics				
Ave	2.15		-5.61	
N	9,848		9,835	
ICC	25.12%		22.28%	
sd(_cons)	33.48		31.62	

The relative importance of a set of vignette factors and respondent controls can also be quantified by the explained variance. When a new "block" of factors is added, the amount of variance explained by the additional factors is calculated. For marriage records (Condition C), the dominant set of factors explaining variance is the source of the information (explaining 25.2% of the variance). For criminal records, the source matters, but it is less important to the respondents and explains less of the variance (11.04%) than it does for the other information types. For voter information (whether the subject voted in the last election), the source of the information explains a significant portion of the variance (20.13%), but the recipient of the information

does not (0.86%). These findings suggest that the source of the information is one of the more important factors to the respondents' judgments regarding the appropriateness of gathering information contained in public records.

C. The Effects of Demographic Differences on Respondents' Judgments

Even with all the vignette factors, interactions, and respondent controls included, we still found a significant portion of the variance in the rating task to be unexplained by the vignette factors. We therefore examined individual differences across respondents to investigate whether these differences may have had an effect on how they judged the vignettes. To do so, we regressed the dependent variable (acceptability, or "this is OK") on the vignette factors for each respondent. This produced a new data set with the respondent-level equations for which N ~ 992. This new data set included the relative importance of each vignette factor to the appropriateness of the vignette (e.g., the importance of online government records and data broker for each respondent (_b_OnlineGovtRecords and _b_DataBroker) are the respondent-specific coefficients in Figure 8 and Figure 9 below). We then regressed the respondent-level coefficients for the vignette factors of source and receiver from this new data set on the respondent controls. The results are shown below in Table 6.

Table 6: Regression of Respondent-Level Coefficient (e.g., Importance of Data Broker) on Respondent Controls

	Source Vignette Factors			
	DataBroker		Online Records	
	coef	p	coef	p
AgeUnder35	**-14.03**	**0.00**	**-7.64**	**0.01**
Female	-4.33	0.19	-5.00	0.07
PrivacyImp	-0.48	0.00	-0.37	0.00
Trust	-0.07	0.03	-0.04	0.18
_cons	-9.50	0.24	5.00	0.47

	Receiver Vignette Factors					
	Bank Receiving		Employer Receiving		Party Guest Receiving	
	coef	p	coef	p	coef	P
AgeUnder35	**-7.75**	**0.00**	**-4.76**	**0.01**	**-6.36**	**0.01**
Female	**5.51**	**0.00**	1.58	0.40	2.64	0.24
PrivacyImp	**0.14**	**0.00**	**0.12**	**0.00**	0.01	0.87
Trust	-0.01	0.50	-0.03	0.09	-0.05	0.01
_cons	6.83	0.14	-3.10	0.50	-1.72	0.75

1. Age

The results on this metric were surprising. Younger respondents (under 35 years old) were critical of using data brokers and online government records when compared to the null condition of asking the subject directly. Similarly, younger respondents were less approving of banks, employers, and a potential guest accessing public records compared to the null condition of a car dealership accessing the information. With voting records, respondents under 35 years old placed greater positive importance on using voting records. In other words, respondents under 35 were more approving to the use of voting records in general. This can be seen in Figure 8 and Figure 9.

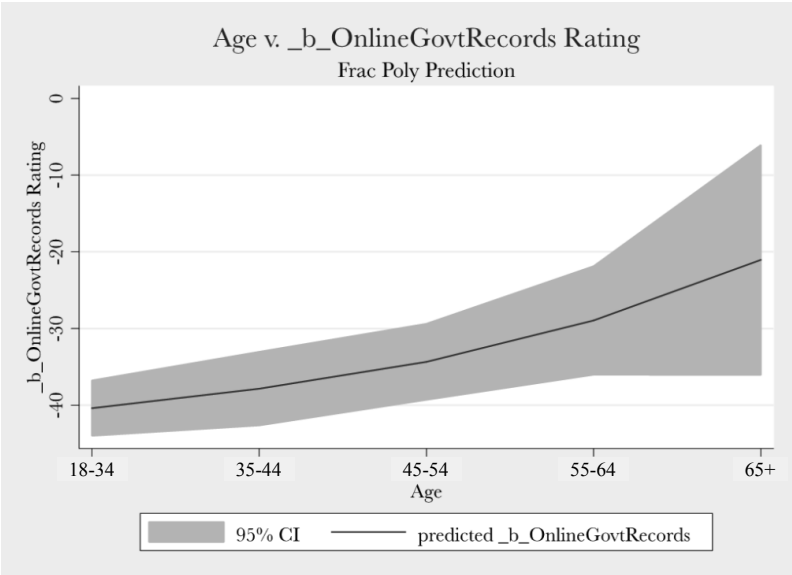

Figure 8: Respondent-Level Importance of Online Government Records over Null (Subject) versus Age

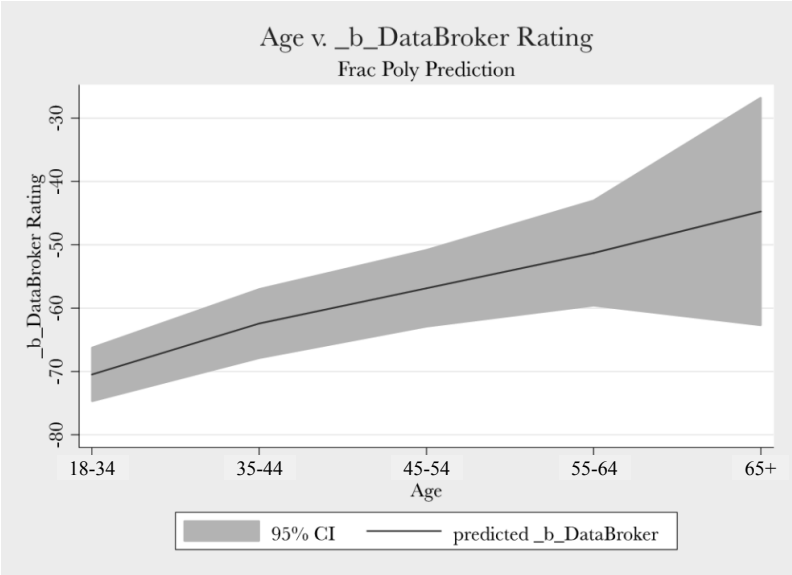

Figure 9: Respondent-Level Importance of Using Data Broker over null (Subject) versus Age

2. Gender

We next examined the role of gender in the perceived level of appropriateness accorded to the gathering and using of public records information in the vignettes. The average vignette rating is lower for female respondents (mean = -9.6) compared to male respondents (mean = -0.5; $t = 4.85$, $p < 0.01$) meaning that women generally found the accessing and use of public records information to be less appropriate. The difference between male and female respondents appears to be most pronounced in the use of marriage records in the employment context. For example, female respondents' judgments that the vignette is appropriate were on average 20 points lower than those of male respondents for the use of marriage information in an employment context when the source was either the individual or online government records, and 15 points lower when sourcing the information from a data broker. Despite these differences, the overall analysis reveals consistent trends across gender in respondents' judgments as to the appropriateness of gathering and using public records information.

V. DISCUSSION AND CONCLUSION

A. Discussion of Current Study

The inspiration for this study comes from a presumption that we believe to be false, but which nonetheless has shaped data and information practices for several decades. This presumption is that public records define a class of information that, by virtue of its presence within a public record, defies all privacy claims. The belief is that for this class of information, "anything goes." Although the presumption may always have deserved greater scrutiny than it has received, the need for such scrutiny has grown exponentially as technology for gathering, disseminating, and manipulating data has vastly amplified the potential "anything" that this presumption allows. Actors from public, governmental and private commercial sectors are now able to sweep the contents of these records, in bulk, into aggregated datasets, from which they can form profiles of individuals, link them to data acquired from other, proprietary sources, and perform sophisticated analyses on this information.

There is a growing clamor for open access to bulk government data in machine-readable form, which includes the contents of these public records. Nevertheless, it is impossible to accept that those who held that some information collected by the government should be made available to all citizens, under no constraints of accountability, had in mind the circumstances we are now experiencing. If they had in mind the benefits of public education and political accountability, which they believed would outweigh the privacy costs to individuals, surely they would realize that given the technological changes, a new calibration is now needed.

Our study does not take up many of the grand policy issues surrounding public records, such as what records should be deemed public, the bases for making such determinations, and how rules of access to and the use of public records should be adjusted in light of the powers of information technologies and data science. Our study is but a small step toward these larger questions. We set out to carefully study how closely aligned people's judgments regarding flows of information held in public records are with the interpretations of policies currently in effect. As we expected, we found that almost all of the people surveyed aligned more with our own intuitions as opposed to current policy interpretations. They expressed disapproval for certain modes of acquiring information and certain modes of promulgating such information despite the fact that all of the information presented is available in public records. Moreover, by drawing on the parameters of contextual informational norms to select the factors we presented in the factorial-vignette surveys, we were able to uncover

useful trends across cases and respondents. Certainly, these findings are merely a beginning, but they clearly indicate that, at least from the perspective of public expectation and opinion, the current policy approach to public records requires a thoughtful and sophisticated overhaul.

Some key findings of the study include:

(1) The degree to which the perceived accessibility of information does not drive judgments about the appropriateness of accessing that information. In other words, even for information that was deemed easy to access (marital status), respondents still judged it to be inappropriate ("Not OK") to access it under certain circumstances.

(2) Appropriateness, as judged by the respondents, depends on the immediate source of information, even for information that could be found in public records. For example, while asking a party host questions about any information type is considered appropriate, accessing the same information through an online government record or a data broker was considered inappropriate. Respondents were particularly consistent, in finding access to be inappropriate in all circumstances where data brokers were the immediate sources of information.

(3) Respondents were most opposed, on average, to accessing voting records across all vignette scenarios (but within this group there was the greatest variance across respondents), and least opposed to accessing information about criminal records, across all scenarios.

(4) Younger respondents (under age 35) were more critical of accessing information by consulting data brokers and online government records than they were of accessing information by asking data subjects directly (the null condition).

(5) Women were more opposed than men to the use of marital status in job applications.

The question then becomes: what is important in judging the appropriateness of accessing each type of public record information? The short answer is: the source of the information. Accessing information through the use of a data broker is consistently perceived as inappropriate, even when the type of information accessed and the receiver of the information are judged to be appropriate. In fact, the relative importance of data brokers as a source varies from -50.89 (criminal records) to -80.47 (marriage records) ($p < 0.01$). In other words, respondents agreed that the use of data brokers was inappro-

priate and the only differences were a matter of the degree to which data brokers are seen as inappropriate sources of public records data. This reaffirms previous findings that individuals agree about privacy norms, expectations, and appropriate uses of information — contrary to labels such as "privacy unconcerned" or "privacy pragmatists."[55]

Interestingly, younger respondents (under the age of 35 years) were more critical of using data brokers and online government records as compared to the null condition of asking the subject directly — contradicting the shorthand and oft-repeated maxim that young adults do not care about privacy. These findings suggest that dismissing or ignoring the concerns of young adults as a policy is shortsighted; in fact, a recent study found that young adults are more aware of privacy issues online and more likely to utilize privacy protecting measures.[56]

B. Implications for Public Policy

1. Public Data versus Available Data

Considering the respondents' strong judgments about the appropriate uses of information, the term "public data" may be not only inaccurate, but also misleading. The term "public" is often conflated with "not private" thereby leading policy makers to believe that individuals have no privacy concerns or expectations around the access and use of these public records. However, our study suggests the opposite. The data presented shows that individuals have deep concerns about who should have access to public records data and how it should be used. While technology has revolutionized the meaning of "public" in public records, it can also be used to refine the systems in place for providing individuals with access to information in a way that promotes the public interest. This will require revisiting the principles behind the creation of public records and considering what types of access and use support these principles.[57]

---

55. *See, e.g.*, Martin & Nissenbaum, *supra* note 1, at 177, 216; Joseph Turow, Michael Hennessy & Nora Draper, *The Tradeoff Fallacy: How Marketers Are Misrepresenting American Consumers and Opening Them Up to Exploitation*, ANNENBERG SCH. COMM., UNIV. PA. 19 (Jun. 2015), https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf [https://perma.cc/U2RF-BDT3]; Chris Jay Hoofnagle & Jennifer M. Urban, *Alan Westin's Privacy Homo Economicus*, 49 WAKE FOREST L. REV. 261, 298–99 (2014).

56. *See* Lee Rainie, *The State of Privacy in Post-Snowden America*, PEW RES. CTR. (Sept. 21, 2016), http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america [https://perma.cc/U2TW-ZRJE].

57. *See* Conley et. al., *supra* note 36, at 778–803.

2. Ex Ante Policies Around Open Data

Perhaps even more pressing, government offices preparing to make data open and machine readable should give careful ex ante consideration to privacy, ensuring that it is an explicit factor in the design of open data systems. Seattle's work with academic researchers at the University of Washington provides an interesting case study.[58] By enlisting the advice of researchers in urban planning, information studies, and privacy, the City of Seattle thoughtfully planned the execution of its open data initiative to reflect an access policy more nuanced than "anything goes."

3. Specific Use of Available Records Data

The findings here would suggest that limiting who has access to the data and how the data will be used should be specifically considered in designing open data initiatives. For example, some states have implemented a multi-step process for accessing voting records and have imposed limits on their use.[59] In general, careful, systematic consideration should be given to data about individuals. It should be noted that the same sophisticated technologies that have caused threats to privacy also hold promise for its protection, for example, for expressing complex rules providing differential access to different parties for different types of data.

4. Review Boards and Open Data

Before making data more accessible, review boards — akin to Institutional Review Boards (IRBs) — could aid both private firms[60] and municipal governments[61] who are seeking thoughtful consideration of data practices. Review boards would be instituted to examine open data initiatives and could include outside experts — from public policy, law, economics, privacy, and business — to ensure policies meet the requirements of all stakeholders.[62]

---

58. *See* Whittington et al., *supra* note 6, at 1903.

59. *See e.g.*, Paul Grabowicz, *Tutorial: Voter Registration Records*, UC BERKELEY GRADUATE SCH. JOURNALISM (2014), https://multimedia.journalism.berkeley.edu/tutorials/voter-registration-records/ [https://perma.cc/5VA2-CVS3].

60. *See* Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 STAN. L. REV. ONLINE 97, 101–02 (2013).

61. *See* Whittington et al., *supra* note 6, at 1958.

62. For example, New Jersey created a review board to assess whether and how court records could be made more "open" while respecting privacy. *See* Conley et al., *supra* note 36, at 790.

5. Private Firms and Open Data Initiatives

Finally, private firms may be enlisted to carry out the open data initiatives.[63] Firms may gather and store the data or even perform some of the analysis. Careful crafting of the contracts would be needed to ensure that policies based on sections 1 through 4, *supra*, are used to govern the practices of firms acting on behalf of the government.

C. Future Research

In the introduction to this Article, we mentioned two challenges: one concerning data that is stored in public records, which this paper addressed, and the second concerning data that is gathered in public spaces, which we are planning to address in future research. Whether such data implicates privacy interests — and if it does, to what degree — has puzzled the courts, vexed the propriety of social practice, and attracted scholarly interest in information, such as location data.[64] In thinking about future research, we plan to investigate how people think about information regarding actions such as purchases, commutes, and attendance at political gatherings. In particular, we are interested in the public's perception of having this information be gathered by a diverse range of parties (other individuals and public and private institutions), and via a diverse range of media (e.g., video, sensors, etc.). Considering the volume of information that the government gathers from individuals in public spaces, the privacy expectations and concerns of individuals ought to be considered a critical input in making determinations of public policy. Furthermore, instead of allowing judges to determine as a matter of law whether individuals have manifested a subjective "reasonable expectation of privacy" for Fourth Amendment purposes, this research might permit reliance on empirics of how privacy manifests in a technological era.[65]

---

63. Seattle provides a good example of holding accountable private firms contracted to carry out data acquisition and retention. *See* CITY OF SEATTLE, OPEN DATA POLICY 3 (2016), https://www.seattle.gov/Documents/Departments/SeattleGovPortals/CityServices/OpenDataPolicyV1.pdf [https://perma.cc/K9UY-HYJU].

64. *See, e.g.*, Stephanie K. Pell & Christopher Soghoian, *Can You See Me Now?: Toward Reasonable Standards for Law Enforcement Access to Location Data That Congress Could Enact*, 27 BERKELEY TECH. L.J. 117 (2014) (discussing the proper law enforcement access standard for prospective location data).

65. *See, e.g.*, Henry F. Fradella et al., *Quantifying Katz: Empirically Measuring "Reasonable Expectations of Privacy" in the Fourth Amendment Context*, 38 AM. J. CRIM. L. 289, 362–65 (2010).