Autonomy and the Collection of Personal Data: Measuring the Privacy Impact Google's Privacy Policy Change

James C. Cooper*

Draft

September 28, 2016

* Associate Professor of Law and Director, Program on Economics & Privacy, George Mason University Antonin Scalia Law School. I thank Jane Bambauer, Bruce Kobayashi, and Hal Varian for valuable comments, and Ryan Lodata and Travis Royer for outstanding research assistance.

1. Introduction

One of the most vexing problems in privacy policy is identifying consumer harm from unwanted data collection. It's relatively easy to quantify the harm suffered from a data breach, in terms of fraudulent charges or the expense and hassle associated with reestablishing a stolen identity. Unwanted observation, however, is a subjective harm. Yet, these types of situations increasingly are the focal point of privacy policy discussions. Take, for example, the recent kerfuffle over WhatsApp's decision to revise its privacy policy to share user information with its parent, Facebook.[1] Or consider one of the most controversial recent FTC privacy actions, *Nomi Technologies*, which involved in-store tracking of hashed MAC numbers.[2] Further, many of the privacy concerns surrounding the Internet of Things and Big Data involve the collection of data from ubiquitous sensors, which may allow a clearer picture of parts of one's life that were once obscured.[3]

Privacy is a capacious term, capable of changing meaning across people and contexts.[4] Nevertheless, one of the widely recognized components of privacy is autonomy: a sphere in which one can engage in "sheltered experimentation and testing of ideas" free from observation.[5] Autonomy, seen in this way, is a vital component to personal growth. What's more, unwanted observation can reduce incentives to engage in productive activities—a sort of inverse moral hazard that underlies the theory of privileges that attach to conversations between doctors and patients, attorneys and clients, and husbands and wives. Reduced autonomy, moreover, can create external harm in the form of an undesirable homogenization of society or a less vibrant democracy.[6]

Although there is a great deal of support in the privacy law scholarship for the notion that a reduction in autonomy from unwanted observation is harmful, there is no empirical evidence to support this claim. This paper is an attempt to fill this gap in the literature by estimating the impact of Google's 2012 privacy policy change on the volume of sensitive searches—those that involve terms most would like to keep private. The theory is simple. After March 1, 2012, Google combined

---

[1] *See* James C. Cooper, *The WhatsApp Privacy Policy Change: No Cause for Alarm*, FORBES, (Sept. 7, 2016), *at* http://www.forbes.com/sites/jamesccooper1/2016/09/07/the-whatsapp-privacy-policy-change-no-cause-for-alarm/#5b85cc5204db.

[2] *See In re Nomi Technologies*, Dkt No. C-4538 (2015), at https://www.ftc.gov/enforcement/cases-proceedings/132-3251/nomi-technologies-inc-matter.

[3] *See* Scott R. Peppet, *Unraveling Privacy: The Personal Prospectus & The Threat of a Full Disclosure Future*, 105 NW. U. L. REV. 1153 (2011); James C. Cooper, *Separation Anxiety,* VA. J.L. TECH. (*forthcoming* 2017).

[4] *See* Daniel J. Solove, *A Taxonomy of Privacy*, 90 CAL L. REV. 1087 (2002); Daniel J. Solove, *Conceptualizing Privacy*, 154 U. PA. L. REV. 477 (2006).

[5] ALAN WESTIN, PRIVACY AND FREEDOM (1967).

[6] *See*, *e.g.*, Anita L. Allen, *Coercing Privacy*, 40 WM. & MARY L. REV. 723, 746 (1999); Julie E. Cohen, *Examine Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373,1424-25 (2000). *See also* Julie E. Cohen, *What Privacy is For?*, 126 HARV. L. REV. 1904, 1911 (2013).

user information across platforms, meaning that search queries would be matched with YouTube views and Gmail to provide Google with a more comprehensive view of its users. Some may want to avoid this intrusion and forego using Google to perform sensitive searches. In this manner, a reduction in sensitive search is an indirect measure of reduced autonomy.

This study relies on Google Trends as a measure of weekly search volume at the state level. Using a difference-in-difference approach, non-sensitive search volume is employed as a benchmark against which to measure changes in sensitive search volume resulting from the change in Google's privacy policy. The results suggest that there is a short-term (1 month) reduction of sensitive search relative to non-sensitive search volume of about 4 percent, but there is no statistically measurable difference looking at six-month or two-month windows. There also doesn't appear to be any difference between high- and low-privacy demand states—measured by the prevalence of state-level privacy legislation. These results are robust to different samples of sensitive search terms, although I cannot rule out the possibility that seasonality is playing some part in the measured short-term decrease in sensitive search.

The results suggest that consumer choice in privacy works: those who were uncomfortable with the Google's new policy of combining data were able to leave. That the reduction in sensitive search was small and transient indicates that any reduction in autonomy was small and perhaps swamped by customization resulting from the cross-platform data sharing. More generally, the empirical results are also in line with a host of research suggesting that consumers are not terribly concerned with the type of data sharing involved in the day-to-day functioning of the online ecosystem that relies on advertising.

The remainder of the paper is as follows. Section 2 examines the concept of privacy harms in more detail, and describes how the Google privacy policy change can be used as a natural experiment to measure the impact of a decrease in autonomy. Section 3 discusses Google Trends data and the empirical methods used to identify the impact of the privacy policy change on sensitive search. Section 4 presents the main results, along with an examination of how results vary based on state demand for privacy, and robustness checks. Section 5 discusses the policy relevance of the findings and concludes.

2. Privacy Harms & the Google Policy Change

The term "privacy harm" is broad, encompassing both tangible and intangible elements.[7] The tangible part of the set is relatively easy to quantify, or at least to identify. Some invasions of privacy can lead to monetary harm—such as when a

[7] *See, e,g*, M. Ryan Calo, The Boundaries of Privacy Harm, 86 INDIANA L.J. 2, 1213 (2011) (classifying privacy harms as "subjective" and "objective").

credit card number is stolen, or leaked private health information creates a stigma that reduces employment or social opportunities—or physical harm, such as when a crazed man finds his ex-wife's address through a data broker. Slightly less quantifiable are intrusions into seclusion that result from, for example, unwanted telemarketing or door-to-door salesmen. Nonetheless, these interruptions are easily seen as harm, and courts and regulators have treated them as such.[8]

One of the most important domains of privacy harm is the impact of unwanted observation. This type of harm is becoming increasingly more relevant as our online behaviors are relentlessly observed and analyzed. For example, what is the harm if Facebook provides your WhatsApp profile to third party markers,[9] or a retail store tracks your movements by using a hashed MAC address? [10] What about when an email is scanned by a server at Google? [11]

If being observed causes me discomfort, anxiety, or embarrassment, I may decide to alter my behavior to something that may be less embarrassing. For example, if I realize that data brokers can determine my sexual preference from my purchase and browsing habits, I may change them. For this reason, autonomy plays a key role in explaining the benefits privacy: under observation, one will not be able to realize their true self for fear of embarrassment, or even being ostracized.
For example, Julie Cohen observes:

> A realm of autonomous, unmonitored choice . . . promotes a vital diversity of speech and behavior . . . We do not experiment only with beliefs and associations, but also with every other conceivable type of taste and behavior that expresses and defines self. The opportunity to experiment with preferences is a vital part of the process of learning, and learning to choose, that every individual must undergo.[12]

What's more, unwanted observation can reduce incentives to engage in productive activities. It is this realization that motivates the theory of privileges that attach to conversations between doctors and patients, attorneys and clients, and husbands and wives. For instance, revelation of HIV status may dull incentives to become

---

[8] *See Miller v. Nat'l Broad. Co*., 232 Cal. Rep. 668 (1986). The FTC's "Do Not Call" list was grounded in this notion of privacy to protect consumers from unwanted invasions of their homes. *See* J. Howard Beales, III & Timothy J. Muris, Ch*oice of Consequences: Protecting Privacy in Commercial Information*, 75 U. CHI. L. REV. 109, 119 (2008).
[9] *See* James C. Cooper, *The WhatsApp Privacy Policy Change: No Cause for Alarm*, FORBES, (Sept. 7, 2016), *at* http://www.forbes.com/sites/jamesccooper1/2016/09/07/the-whatsapp-privacy-policy-change-no-cause-for-alarm/#5b85cc5204db.
[10] *See In re Nomi Technologies*, Dkt No. C-4538 (2015), at https://www.ftc.gov/enforcement/cases-proceedings/132-3251/nomi-technologies-inc-matter.
[11] *See In re Google, Inc. Gmail Litigation*, 13-MD-02430 (N.D. Cal.).
[12] Julie E. Cohen, *Examine Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373,1424-25 (2000). *See also* Julie E. Cohen, *What Privacy is For*, 126 HARV. L. REV. 1904, 1911 (2013) ("Lack of privacy means reduced scope for self-making . . . . privacy is one of the resources that situated subjects require to flourish").

tested in the first place, although such knowledge clearly is valuable.[13] Just as copyrights and patents are designed to foster incentives to create and invent, moreover, providing exclusive rights in personal information can enhance incentives for self-discovery.[14] Further, some have expressed concern that reduced scope for autonomy may carry with it external effects, reducing diversity and ultimately having a deleterious impact on democracy.[15] This type of privacy harm—a reduction in autonomy that becomes manifest in a censored-self—is what this paper hopes to measure.

To put things more concretely, consider the following. When faced with the undesired gaze, one has two choices: continue with your behavior, and endure the privacy harm; or discontinue the behavior to avoid the anxiety or embarrassment. Importantly, both paths are just different manifestations of the same underlying reduction in privacy. Suppose that the unobserved self would engage in activity X. Under observation, you have the choice to continue engaging in X, with some privacy harm, c, or discontinue X and do Y, a behavior that will not give rise to privacy harms. The choice will depend on whether $U(X) - c \lesseqgtr U(Y)$, where U is utility. $U(X)$ includes the "discovery" value from the activity and is assumed to be greater than $U(Y)$. Analogous to a price increase, the imposition of privacy cost, c, will cause some marginal consumers to leave the market for X, and will reduce (by c) the surplus to the inframarginal consumers who remain. Both $U(X)-U(Y)$ and c are privacy harms, but we will never be able to measure their magnitudes directly.

However, we can indirectly measure the magnitude of c by measuring the number of people who switch from X to Y in response to observation. Given heterogeneous costs and benefits, the choice of X or Y will vary because $U(X)-U(Y) > c$ for some people, and $U(X)-V(Y) < c$ for others. But in general, higher levels of c will lead to more switching because *ceteris paribus*, as c rises, $U(X)-c$ becomes smaller, and thus will be less than Y for a larger proportion of the population. Thus, measuring the magnitude of switching from X to Y will provide some insight into the magnitude of c. Google's 2012 change to its privacy policy provides an opportunity to measure such switching, and thus to indirectly measure autonomy costs.

[13] *See, e.g.*, Benjamin E. Hermalin & Michael L. Katz, Privacy, *Property Rights and Efficiency: The Economics of Privacy as Secrecy*, 4 QUANT. MKT'G & ECON. 209, 212 (2006).

[14] *See* Richard S. Murphy, *Property Rights in Personal Information: An Economic Defense of Privacy*, 84 GEO. L.J. 2381, 2386-87(1996). To the extent that there are positive externalities for society from privacy— for example those that underlie rights-based notions of privacy, which focuses on notions of autonomy that are necessary to spur the type of diversity, creativity, and intellectual development that serves society as a whole—private incentives for self-discovery may not be optimal. *See, e.g.*, Joel Reidenberg, *Privacy Wrongs in Search of Remedies*, 54 HASTINGS L.J. 877 (2003); Daniel Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1892 (2013); Julie Cohen, *What Privacy is For*, 126 HARV. L. REV. 1904, 1911 (2013); Neil Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 407 (2008).

[15] *See* Anita L. Allen, *Coercing Privacy*, 40 WM. & MARY L. REV. 723, 746 (1999) ("It is not simply that people need opportunities for privacy; the point is that their well-being, and the well-being of the liberal way of life, requires that they in fact experience privacy.").

On January 24, 2012, Google announced that it would be introducing "a new main privacy policy that covers the majority of our products and explains what information we collect, and how we use it, in a much more readable way."[16] As Google put it, the combination of policies means that "if you're signed in, we may combine information you've provided from one service with information from other services."[17] Specific to search, Google explained that combining the data will facilitate better results, such as, "figuring out what you really mean when you type in Apple, jaguar or Pink."[18] The new policy would go into effect on March 1, 2012.

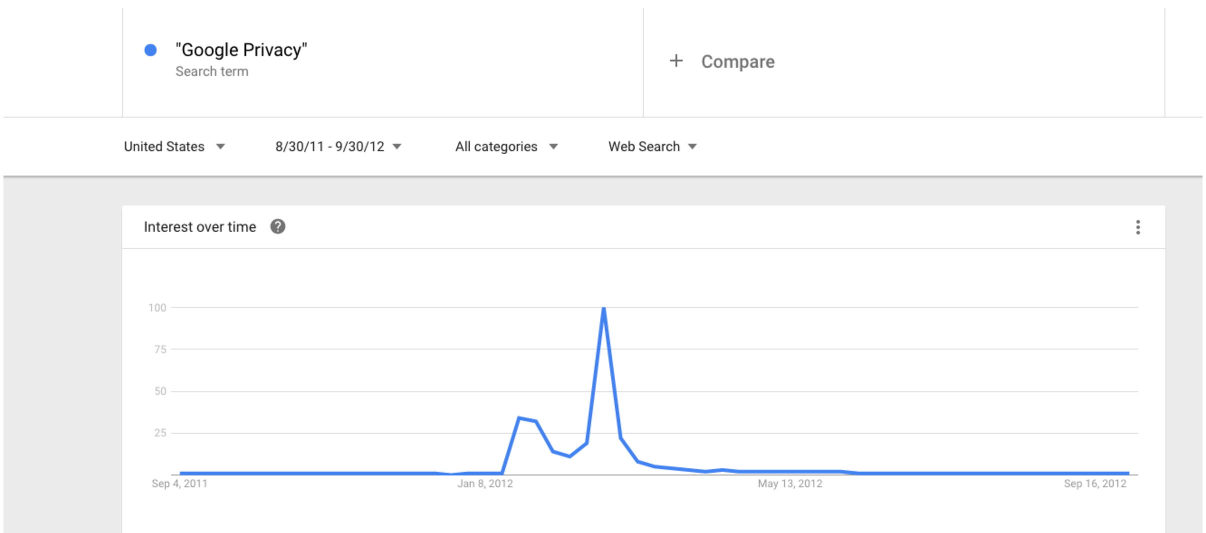
The condemnation of Google was swift in the privacy advocacy community.[19] The Trans Atlantic Consumer Dialogue, a consortium of consumer advocacy groups, described the proposed changes to Google's privacy policy "troubling" and a "mistake."[20] By the end of February 2012, the Electronic Privacy Information Center (EPIC), along with the Center for Digital Democracy, Consumer Watchdog, the Consumer Federation of America, and the US Public Interest Research Groups signed onto a letter submitted to the House and Energy and Commerce Committee calling for an open hearing on "Google['s] plans to go forward with a substantial change in business practices that will affect millions of users of the Internet without any opportunity for users to consent."[21] Further, EPIC sued the FTC to compel it to bock Google's policy change as a violation of Google's 2011 consent decree, which settled privacy charges involving the Google Buzz rollout.[22] EPIC alleged that if the FTC did not act to prevent Google's change, Google users "face an imminent harm that is both certain and great."[23]

There is reason to believe that a substantial number of consumers, and certainly those who are the most privacy sensitive, would have been aware of this change. Google emailed the change to all of its Gmail users, and National media and tech-centric news outlets quickly picked up news of Google's changes.[24] One online

---

[16] Alma Whitten, *Updating Our Privacy Policies and Terms of Service*, Google Official Blog (Jan. 24, 2012), https://googleblog.blogspot.com/2012/01/updating-our-privacy-policies-and-terms.html.
[17] *Id*.
[18] *Id*.
[19] *See*, *e.g.,* David DiSalvo, *Google Says Bye Bye to User Privacy*, FORBES, Jan. 24, 2012, available at http://www.forbes.com/sites/daviddisalvo/2012/01/24/google-says-bye-bye-to-user-privacy/#164e3de37b0a
[20] John Mello, *Multinational Consumer Group Asks Google to Delay Privacy Changes*, PCWORLD, Feb. 29, 2012, *at* http://www.pcworld.com/article/251058/multinational_consumer_group_asks_google_to_delay_privacy_changes.html.
[21] Letter from Marc Rotenberg, President and Exec. Director, Electronic Privacy Information Center, et al to Members of the House Energy and Commerce Committee, (Feb. 24, 2012), *at* https://epic.org/privacy/ftc/google/Privacy-Groups-ltr-to-Bono-Mack.pdf.
[22] *See Electronic Privacy Information Center v. Federal Trade Commission* (D.D.C. Feb. 8, 2012).
[23] *Id*. at 20.
[24] *See* Hayley Tsukayama, *FAQ: Google's New Privacy Policy*, WASHINGTON POST, Jan. 24, 2012, *available at* https://www.washingtonpost.com/business/technology/faq-googles-new-privacy-policy/2012/01/24/gIQArw8GOQ_story.html. *See also* Claire Miller, *Google to Update Privacy Policy to Cover Wider Data Use*, N.Y. TIMES, Jan. 24, 2012, available at

publication even went so far as to create a page that was dedicated to providing updates on the ensuing media firestorm and governmental action.[25] In its filings against the FTC, moreover, EPIC described Google's policy changes as so widely known among the public that "[i]f the government is unaware that Google plans to make a substantial change in its business practices on March 1, 2012, it should turn on a computer connected to the Internet."[26] In a February 1, 2012 blog post, the Electronic Frontier Foundation stated that Google has done "a great job of informing users that the privacy policy has been changed through emails and notifications."[27] Consumer search behavior also suggests an interest in Google's privacy policy change. As the screenshot from Google Trends below shows, searches for "Google and Privacy" peak around the announcement in January 2012, and again when the policy went into effect in March.



Given the widespread knowledge of Google's privacy policy change, and its clear implications for privacy, I use it as a natural experiment to identify autonomy harms. To some, the new policy means unwanted observation—the ability to link Gmail and YouTube with search queries may provide Google with too accurate a view of oneself. Some may want to avoid this intrusion by no longer using Google to search for sensitive topics, such as those dealing with sexuality, embarrassing health

---

http://bits.blogs.nytimes.com/2012/01/24/google-to-update-its-privacy-policies-and-terms-of-service/; Leena Rao, *Google Consolidates Privacy Policy; Will Combine User Data Across Services*, TECHCRUNCH, Jan. 24, 2012, available at https://techcrunch.com/2012/01/24/google-consolidates-privacy-policy-will-combine-user-data-across-services/; Tim Carmody, *Google Streamlines Privacy Policy to Integrate Its Products*, WIRED, Jan. 24, 2012, available at https://www.wired.com/2012/01/google-streamlines-privacy/.

[25] Dante D'Orazio, *Google's 2012 Privacy Policy Changes: The Backlash and Response*, THE VERGE, Feb. 1, 2012, available at http://www.theverge.com/2012/2/1/2763898/google-privacy-policy-changes-terms-of-service-2012.

[26] *EPIC v. FTC (Enforcement of the Google Consent Order)*, ELECTRONIC PRIVACY INFORMATION CENTER, https://epic.org/privacy/ftc/google/consent-order.html.

[27] Rainey Reitman, *What Actually Changed in Google's Privacy Policy*, Electronic Frontier Foundation (Feb. 1, 2012), https://www.eff.org/deeplinks/2012/02/what-actually-changed-google's-privacy-policy.

conditions, or controversial political views. At the margin, therefore, we can expect privacy sensitive consumers to be deterred from conducting such sensitive searches. Simply put, by increasing the scope of observation, the privacy policy change may have increased the price of sensitive search sufficiently to prompt some to exit the market entirely. In this manner, a reduction in sensitive search is an indirect measure of reduced autonomy.

.

3. Data & Estimation Strategy

3.1 Data

The goal of this paper is to measure the privacy impact of Google's decision to combine data from all of its platforms. One way to go about this would be through surveys, to see if people's intentions to use Google search to perform sensitive queries has changed due to privacy concerns. There are several problems with this method, however. First, and most problematic, surveys are stated preferences, rather than a measure of actual consumer behavior given a set of real-world choices.[28] Second is the problem of social desirability bias in surveys: respondents may answer in a way that is seen as socially acceptable rather than give a true answer.[29] Third, a survey response has a lag—it could only be operationalized after the research question is formalized and the survey instrument is constructed. Finally, because surveys are expensive and time consuming, they will at most provide only a couple snapshots in time.

In an effort to avoid these problems, this paper uses Google Trends (GT) data as a measure of revealed preference for privacy. GT is historical search volume data (dating back to 2004) at various geographic levels (world, country, state, city) that Google makes freely available.[30] I collected weekly GT data on twenty sensitive search terms and twenty non-sensitive search terms to serve as a control group. I chose a window of 6 months before and after the privacy policy change (September 1, 2011 – September 1, 2012) to measure reaction, although smaller windows are also examined.

It is important to note that GT data are not raw volume numbers, but rather an index ranging from 0 to 100. As explained on the Google Trends site, a GT score of 50 means that the search volume for that week was half as large as at its peak.[31]

---

[28] *Cf.* Jerry Hausman, *Contingent Valuation: From Dubious to Hopeless*, 26 J. ECON. PERSPECTIVES 43 (2012).
[29] *See* Seth Stephens_Davidowtiz & Hal Varian, *A Hands-on Guide to Google Data* 16 (Mar. 7, 2015), *at* http://people.ischool.berkeley.edu/~hal/Papers/2015/primer.pdf.
[30] https://www.google.com/trends/.
[31] As noted in the help portion of the page:

Although the exact methodology by which the index is calculated is not public, a recent paper by two Google economists describes it in the following manner: "The index measures the fraction of queries that include the term in question in the chosen geography at a particular time relative to the total number of queries at that time."[32] This means that lower values do not necessarily imply lower volume, but rather that "there are fewer searches, as a percent of all searches, than there were previously."[33] Because GT scores do not measure volume directly, they do not allow direct comparison of magnitudes across terms. Thus, if Taylor Swift and Saran Wrap have average GT scores of 40 and 25, respectively, it does not necessarily mean that Taylor Swift's search volume is 60 percent higher than Saran Wrap. GT scores, however, allow *directional* comparisons. For example, suppose that two search terms have a value of 50 at a given time, then the following period, term 1 has a value of 40 and term 2 has a value of 60. Although we cannot compare magnitudes, we can say that *as a proportion of all searches in that region and time frame*, term 1 volume decreased by 20 percent, whereas term 2 volume increased by 20 percent. This facet is important for the empirical analysis, because being able to identify relative trends in search volume change will allow a difference-in-difference approach to identify the impact of the Google policy change on search behavior.

Despite these potential shortcomings, GT data increasingly has been used in academic studies as a measure of real time consumer interests. For example, Choi and Varian,[34] and Brynjolfsson and Wu[35] show how GT data can be used as an indicator of consumer demands to improve short-term or near-real time of sales in real estate and automobile markets, as well as a predictor of unemployment claims and travel. Perhaps the most famous uses of GT data for so-called "nowcasting" is Flu Trends; data mining discovered patterns of searches that were statistically associated with near-real time flu outbreaks.[36] GT data also has been used as a proxy for unobservable consumer sentiment. For example, Stephens-Davidowitz uses GT data on searches for racial slurs to detect unobservable racial animus.[37] He finds this racially charged search rate correlated with other measures of racial views, and a negative predictor of Barak Obama's vote share. Similarly, Baker and

---

> Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. Likewise a score of 0 means the term was less than 1% as popular as the peak.

[32] Stephens-Davidowitz & Varian, *supra* note 31, at 12.
[33] *Id*.
[34] Hyunyoung Choi & Hal Varian, *Predicting the Present with Google Trends* (2011), *at* http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf.
[35] Lynn Wu & Erik Brynjolfsson, *The Future of Prediction: How Google Searches Foreshadow Housing Prices & Sales* (2009), at http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf.
[36] Jeremy Ginsberg *et al*., *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012 (Nov. 2008).
[37] Seth Stephens-Davidowtiz, *The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data*, 118 J. PUB. ECON. 26 (2014).

Fradkin use GT to construct an index of job search intensity.[38] They show this index to be correlated with existing job search data, and use the GT index to estimate the impact of certain changes in unemployment insurance programs on job search activity.

In the study that is closest to the present work, Marthews and Tucker examine the impact of the Snowden revelations on search behavior that may trigger government interest.[39] Using GT data for 2013, they find that after the Snowden revelations, searches for words flagged as potentially related to terrorism by the Department of Homeland Security (DHS) fell relative to searches for neutral terms. As discussed in more detail below, I use an approach similar to Marthews and Tucker in testing the impact of Google's policy change on sensitive versus non-sensitive search.

For this study, GT data from January 1, 2011 to December 31, 2013 from all fifty states plus Washington, DC were collected. The non-sensitive terms are Google's top twenty searches over the time period. The sensitive terms for this study were taken from the Marthews and Tucker study, in which they crowd-sourced a group of search terms that were "likely to be embarrassing to a friend." We chose the twenty terms with the highest embarrassment scores.[40] Appendix Table 1 shows the average Google Trends score for each term. The average trends score for sensitive search is 52.8 compared to 58.1 for non-sensitive search. "Porn," "Depression," and "Acne" have the highest sensitive scores, while "Yahoo," "E-bay" and "Mail" have the highest non-sensitive scores.

3.2 Estimation Strategy

The primary approach I use to identify the impact of the Google privacy policy change is difference-in-difference estimation. This approach measures the difference in the outcome variable of interest between a treatment and a control group both before and after the relevant policy change. Unlike the typical difference-in-difference estimation strategy, in this setting everyone in the US was treated simultaneously, which means that I do not have separate treatment and control groups—e.g., a group of states or people who were impacted by the Google change, and another group that were not. Instead, identification is based on the difference between a set of behaviors that would be impacted—sensitive search—

[38] Scott R. Baker & Andrew Frandkin, *The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data* (July, 2016), *at* http://andreyfradkin.com/assets/FullTexasJobSearch.pdf.
[39] Alex Marthews & Catherine Tucker, *Government Surveillance and Internet Search Behavior* (Apr. 2015), at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564.
[40] Tucker & Marthews crowd sourced a preliminary set of sensitive terms, and then used Mechanical Turkers to rate the sensitivity of these terms on a scale of 1-5.

and another set that would not—non-sensitive search. Thus, the basic empirical strategy is to measure the following:

$$\widehat{\Delta\Delta} = \left(GT_{NS,B} - GT_{S,B}\right) - \left(GT_{NS,A} - GT_{S,A}\right), \qquad (1)$$

where the subscripts *NS* and *S* are for non-sensitive and sensitive terms, respectively, and the subscripts *B* and *A* are for before and after Google's policy change, respectively. In this manner, the relationship between sensitive and non-sensitive GT scores in the pre-policy change period is used to construct a counterfactual world in which Google never changed their privacy policy, and the actual measured difference is compared to what the difference would have been in the counterfactual world. An estimate of $\widehat{\Delta\Delta}$ less than zero would be evidence that Google's policy change had a greater impact on sensitive than non-sensitive search behavior.

Because the difference-in-difference approach relies on measuring changes in relative trends, the lack of relative comparability in GT data should not hinder identification. Although I will be unable to measure changes in actual volume, I should be able to identify changes in sensitive search terms relative to non-sensitive search terms. For example if, as is hypothesized, Google's policy change impacts sensitive but not non-sensitive search, GT data for sensitive search should exhibit a downward trend relative to sensitive search.[41]

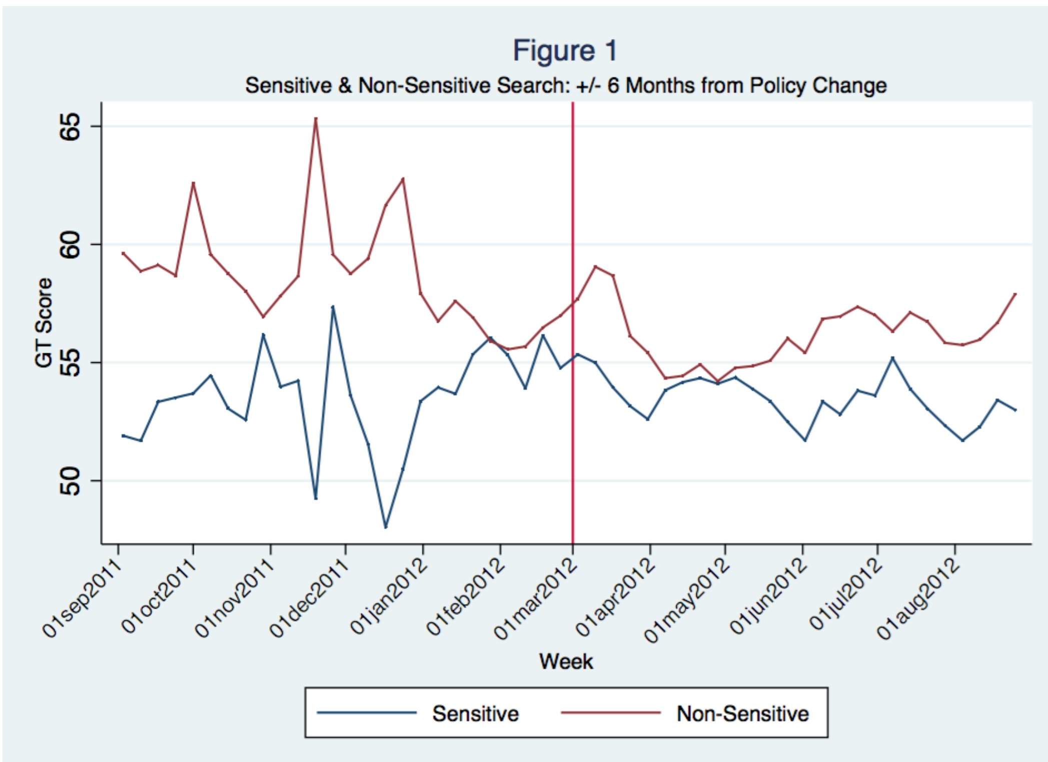
4. Empirical Results

4.1 Overview of Trends

Before turning to the main regression results, I present some simple graphical representations to give a flavor of the difference-in-difference estimation

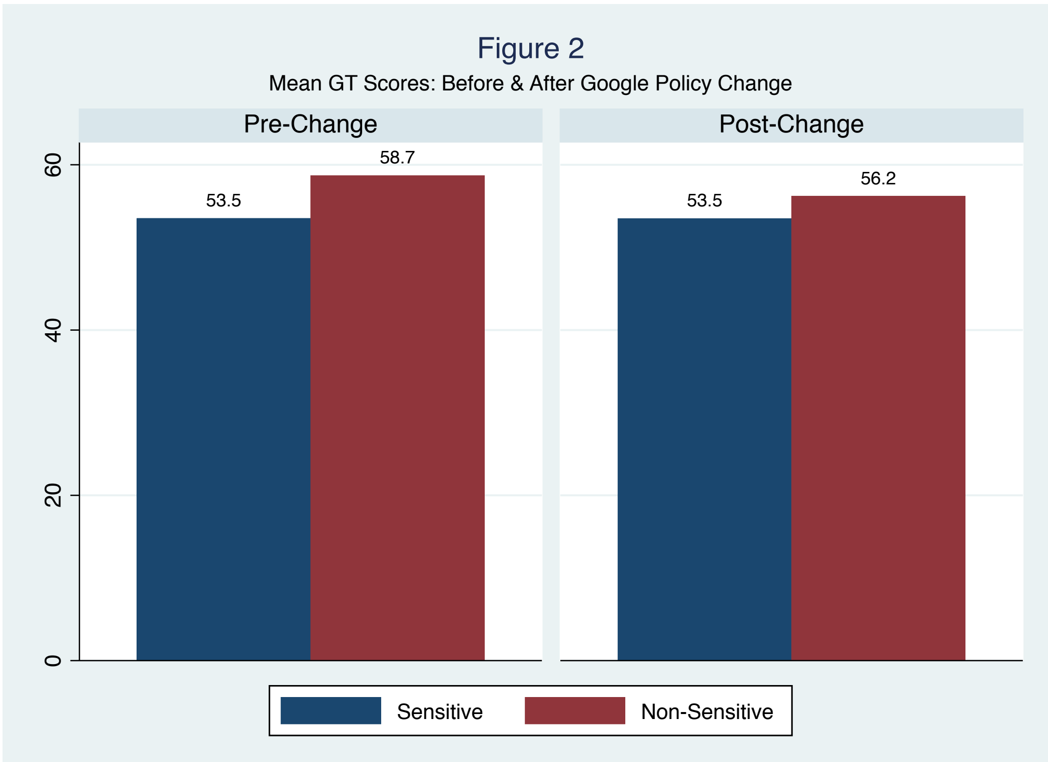
[41] Users also could avoid Google surveillance of their sensitive searching by logging out of their Google accounts. GT data include searches from both logged in and non-logged-in users, so using GT data to measure the impact of Google's policy change will underestimate the total number of people who changed their behavior to avoid observation to the extent that some users logged out to perform sensitive search. Android users must be logged onto their Google accounts for most functionality, reducing this as an option for mobile search from the leading mobile operating system. It is also possible that logged-on users changed to private search sessions (e.g., "incognito mode" in Chrome or "private window" in Safari). These modes remove local traces of the browsing session through eliminating search history and not accepting cookies, but do not prevent Google from logging search queries. Thus, there is no technical reason that these users' search queries could not be included in GT data. However, it is unclear whether Google excludes private browsing sessions from GT calculations. If it does exclude private browsing sessions, then GT data accurately capture those who are seeking to keep sensitive searching private. If GT data do not exclude private browsing sessions, then it will underestimate those who are *seeking* to avoid detection by Google, but will accurately measure those who are not *actually* hiding their search behavior.

strategy. First, Figure 1 shows weekly search volume for sensitive and non-sensitive search for six months prior to and after the Google policy change. The vertical line at March 1, 2012 divides the sample between pre- and post-Google policy change. As a whole, Figure 1 does not provide much evidence that the difference between sensitive and non-sensitive search changed dramatically. Before the policy change, the two types of searches appear to have a weak inverse relationship—at several points sensitive search volume appears to fall as non-sensitive search volume rises, and vice-versa. Although there is wide variation in GT scores, the trends appear level. After the policy change, the inverse relationship



appears to continue; there appears to be a slight downward trend in sensitive searches immediately following the policy change, through June, and an upward trend in non-sensitive search volume, which ends in April. A striking feature of the time series is the large reduction in variance that occurs around the beginning of 2012, as both series tend to fluctuate in closer proximity to their long run means.

Figure 2 shows the average GT scores for sensitive and non-sensitive searches six months before and after that privacy policy change. Again, it's not the magnitude of the difference, but the directional change with which we're concerned. The data show that the average difference in GT scores actually *shrunk* after the policy change, from 5.2 to 2.7, with the change coming from a reduction in the average non-sensitive GT score; the average sensitive GT score remained unchanged



4.2 Regressions

Although visual inspection of the data and a simple comparison of means provides little evidence that the policy change had an impact on consumers' willingness to engage in sensitive searches, there are myriad factors that impact search behavior. Simple averages may mask underlying trends that are correlated with the policy change. To control for these factors, I estimate the following difference-in-difference model:

$$GT_{ijt} = b_1{*}(Sensitive \ x \ GooglePolicy) + \theta_i + \Phi_j + \alpha_t + e_{ijt} \qquad (2)$$

In equation (2), GT_{ijt} is the GT value for term *i*, in state *j*, during week *t*. *Sensitive* is a dummy variable equal to 1 if the search term is sensitive, and *GooglePolicy* is a dummy variable equal to 1 after March 1, 2012. Thus, b_1 is the difference-in-difference estimator ($\widehat{\Delta\Delta}$), measuring the differential impact on sensitive and non-sensitive search due to the policy change. A negative sign on b_1 would support the hypothesis that Google's policy change impacted consumer's privacy. The terms θ_i, Φ_j, and α_t are search term, state, and week fixed effects. These controls are designed to account for idiosyncratic state factors that don't vary over time or with search type; idiosyncratic search factors that don't vary over time or across states; and events that may be specific to a certain week across all states and search terms. All specifications are estimated with robust standard errors clustered at the search term level.

The main results are reported in Table 1. The first specification uses pooled data. The estimated main effect of *Sensitive* shows that over the year measurement period, the GT scores for sensitive searches are an average of 5.4 points lower than non-sensitive scores, although this is measured with imprecision and statistically insignificant. The results suggest that following the Google policy change, average GT scores fell for both sensitive and non-sensitive search by 2.5 points. The estimate of *Sensitive*GooglePolicy*, is positive, which would suggest that the Google policy change actually had a smaller impact on sensitive than non-sensitive search. That this estimate is insignificant even with over 100,000 observations, however, strongly suggests that the policy had no differential impact.

The next column includes term, state, and year effects. Because the main effects of *Sensitive* and *GooglePolicy* are perfectly collinear with time and state effects, they are left out of the equation. The estimated coefficient on *Sensitive*GooglePolicy* is almost identical to the pooled specification, but the R^2 increases dramatically (.01-.65), suggesting that most of the variation in GT scores is unrelated to the type of search. [42]

TABLE 1
IMPACT OF GOOGLE POLICY CHANGE

	Window			
	+/- 6 Months	+/- 6 Months	+/- 2 Months	+/- 1 Month
*Sensitive * GooglePolicy*	1.95 (1.54)	1.96 (1.53)	-.666 (1.07)	-2.41** (1.21)
Sensitive	-5.44 (5.05)			
Google Policy	-2.50** (1.15)			
Fixed Effects	N	Y	Y	Y
R^2	.01	.649	.700	.708
N	108,056	108,056	35,326	18,702

Dependent variable is Google Trends index score for term *i*, in state *j*, during week *t*. Fixed Effects include search term, state, and week effects. 95% confidence interval reported in brackets; robust standard errors clustered on search term are reported in parentheses.

[42] The nearly identical parameter estimates is due to the fact that there is no state-level variance in policy change.

It is possible that a +/- 6-month widow is too large to detect an impact that may have occurred very close to the policy change. Further, as can be seen in Figure 1, the relationship between sensitive and non-sensitive search is somewhat erratic from September – December 2011, calling into question the extent to which non-sensitive search can act as a valid control for this time span.[43] Accordingly, columns 3 and 4 examine windows of +/-2 months (December 31 – May1) and +/-1 month (January 31 – April 1), respectively. The point estimates for *Sensitive*GooglePolicy* are negative in both of these specifications, but it is significant only for the +/-1-month window, suggesting that the average GT score gap between sensitive and non-sensitive search increased by 2.4 points after the policy change.

Overall, the main results suggest that if there were any impact, it was brief and small—around a 5 percent average reduction that's not detectable beyond one month after the policy change.

4.2 Privacy Demand

In this section, I investigate the extent to which demand for privacy may impact consumer reaction to the privacy policy change. Although it is impossible to know the preferences of each individual searcher, I proxy demand for privacy based on privacy protections the searcher's state has in place. More specifically, using data from the National Association of State Legislatures, I constructed a privacy index based on the presence of law related to online privacy in 2016. It is reasonable to assume that states with more privacy laws may have populations that on average have higher demand for privacy.

Table 2 shows states with privacy laws in seven categories: Explicit constitutional provisions; employee email and Internet usage monitoring; requiring privacy policies; online privacy for children; e-Reader privacy; social media monitoring; and License plate readers.[44] A majority of states (34) have at least one of these provisions. California, Delaware, and Connecticut have the largest number of privacy provisions, with 6, 5, and 4 respectively. Of the states with privacy laws, the average number of provisions is just under two (1.76), and the most common

[43] In regressions (not reported), *Week* was interacted with *Sensitive* for each week prior to the policy change back to September 1, 2011. An *F-test* that the estimated coefficients on the *Week-Sensitive* interactions are jointly equal to zero is rejected ($F_{25,41} = 5.31$, $p < .001$), suggesting that sensitive and non-sensitive search moved differently through time during this period. The same analysis with a pre-policy period dating back to December 31, 2011 does not reject the hypothesis that the *Week-Sensitive* interactions are jointly equal to zero ($F_{7,41} = -1.29$, $p = .282$), suggesting that non-sensitive search moves with sensitive search for the period December 31, 2011-February 28, 2012, and hence is a suitable control.

[44] *See* NCSL, State Laws Related to Internet Privacy, at http://www.ncsl.org/research/telecommunications-and-information-technology/telecom-it-privacy-security.aspx.

provision is one controlling the ability of employers or educators to examine applicants' social media accounts (74 percent).

TABLE 2
STATE PRIVACY LAWS

State	**Total**	Employee Email	Privacy Policy	e-Reader	Children	License Plate	Social Media	Constitution
CA	6	0	1	1	1	1	1	1
DE	5	1	1	1	1	0	1	0
CO	3	1	0	0	0	1	1	0
CT	3	1	1	0	0	0	1	0
TN	3	1	0	0	0	1	1	0
AR	2	0	0	0	0	1	1	0
AZ	2	0	0	1	0	0	0	1
FL	2	0	0	0	0	1	0	1
IL	2	0	0	0	0	0	1	1
LA	2	0	0	0	0	0	1	1
MD	2	0	0	0	0	1	1	0
ME	2	0	0	0	0	1	1	0
MT	2	0	0	0	0	0	1	1
NH	2	0	0	0	0	1	1	0
UT	2	0	0	0	0	1	1	0
WA	2	0	0	0	0	0	1	1
AK	1	0	0	0	0	0	0	1
HI	1	0	0	0	0	0	0	1
MI	1	0	0	0	0	0	1	0
MN	1	0	0	0	0	1	0	0
MO	1	0	0	1	0	0	0	0
NC	1	0	0	0	0	1	0	0
NE	1	0	0	0	0	0	1	0
NJ	1	0	0	0	0	0	1	0
NM	1	0	0	0	0	0	1	0
NV	1	0	0	0	0	0	1	0
OK	1	0	0	0	0	0	1	0
OR	1	0	0	0	0	0	1	0
RI	1	0	0	0	0	0	1	0
SC	1	0	0	0	0	0	0	1
VA	1	0	0	0	0	0	1	0
VT	1	0	0	0	0	1	0	0
WI	1	0	0	0	0	0	1	0
WV	1	0	0	0	0	0	1	0
Mean	1.76	.12	.08	.12	.06	.35	.74	.29

I use two approaches to examine whether more privacy-sensitive states would experience a greater relative reduction in sensitive searching from Google's policy change. First, I split the sample based on the presence of any privacy law, or whether the state had a high demand for privacy (defined as having three or more privacy laws). If reactions in these states to the policy change is larger due to a more privacy-sensitive population, we would expect the estimated coefficients on *Sensitive*GooglePolicy* to be larger than those reported for the full sample in Table 1. Results are reported in Table 3, and are almost identical to those in the full sample. The estimated impact for the +/6-month widow is positive and insignificant. Similarly, the estimated impact for the +/-2-month window is small, negative, and statistically insignificant like the main results. Surprisingly, the estimated impact in states with any privacy law is marginally lower (-2.34 vs. -2.41), and the +/-1-month window for the highest privacy states is statistically indistinguishable from zero.

TABLE 3
PRIVACY-SENSITIVE STATE SAMPLE

	Any Privacy Law			High Privacy Law		
	+/- 6 Months	+/- 2 Month	+/- 1 Month	+/- 6 Months	+/- 2 Months	+/- 1 Month
*Sensitive * GooglePolicy*	2.29 (1.52)	-.617 (1.11)	-2.34* (1.27)	2.14 (1.46)	-.691 (1.33)	-2.51 (1.63)
Fixed Effects	Y	Y	Y	Y	Y	Y
R^2	.653	.581	.712	.687	.607	.742
N	73,580	24,055	12,735	10,868	3,553	1,881

Dependent variable is Google Trends index score for term *i*, in state *j*, during week *t*. Fixed Effects include search term, state, and week effects. Robust standard errors clustered on search term are reported in parentheses.

A second approach to examining whether privacy demand impacts the results is to include an interaction between a measure of state privacy laws and *Sensitive*GooglePolicy*. If consumers from states with higher privacy demand are more privacy sensitive, the sign on the estimated coefficient of this triple-difference (*PrivacyLaw*Sensitive*GooglePolicy*) should be negative. That is, any measured increased between sensitive and non-sensitive GT scores after the Google policy change should be larger (*i.e*, more negative) for states with a higher demand for privacy. Three different measures of state privacy laws were interacted with *Sensitive*GooglePolicy*: the raw number of privacy laws (*PTOTAL*); an indicator variable equal to one if the state had any privacy laws (*PLAW*); and an indicator variable equal to one for "high privacy" states, defined has having three or more privacy laws (*HI_PLAW*). Results are reported in Table 4.[45]

[45] The main effects of PTOTAL, PLAW, and HI_PLAW are excluded because of collinearity with state fixed effects.

TABLE 4
PRIVACY LAW INTERACTIONS

	Window								
	+/- 6 Months			+/- 2 Months			+/- 1 Month		
*Sensitive * GooglePolicy*	1.68 (1.59)	2.25 (1.62)	1.84 (1.53)	-.808 (1.06)	-.776 (1.06)	-.249 (1.05)	-2.67** (1.19)	-2.21* (1.17)	-2.53** (1.20)
PTOTAL Sensitive * GooglePolicy*	.235 (.159)			.118 (.172)			.218 (.221)		
PLAW Sensitive * GooglePolicy*		-.421 (.381)			1.09 (.724)			-.284 (.571)	
HI-PLAW Sensitive * GooglePolicy*			1.18* (.597)			-.609 (.409)			1.19 (.162)
Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	.649	.649	.649	.700	.700	.700	.708	.708	.708
N	108,056	108,056	108,056	35,326	35,326	35,326	18,702	18,702	18,702

Dependent variable is Google Trends index score for term *i*, in state *j*, during week *t*. Fixed Effects include search term, state, and week effects. Robust standard errors clustered on search term are reported in parentheses.

Consistent with the results from the sample split approach, there does not appear to be any statistically measurable difference in reaction to the Google policy change between states with and without privacy laws. The estimated coefficient on the state law interaction term is positive for the +/-6 and +/-1 month windows specifications with *PTOTAL* and *HI_PLAW* interactions, and marginally significant for one specification. The interaction on *PLAW* is negative, but also statistically insignificant for both six-month and one-month windows. Further, the estimated coefficients on the main interaction term, *Sensitive*GooglePolicy*, are consistent with the main results presented in Table 1: statistically indistinguishable from zero for the six-month and two-month windows, and negative and significant for the one-month window.

Both the sample splitting and interactions suggest that consumer demands for privacy, as represented through privacy laws, are not a related to search behavior in the wake of the Google policy change. This suggests that privacy laws do a poor job at capturing citizen's demands for privacy or that demand for privacy—or at least the type of privacy represented in the state laws—is unrelated to perceptions of privacy invasions resulting from Google's policy change.

4.3 Robustness Checks

4.3.1 Random Sampling of Sensitive Terms

The main regressions are based on a subset of the search terms crowd-sourced by Marthews and Tucker that had the highest scores for their potential embarrassment with friends. Because the sensitivity of a search term is subjective, it is likely that the terms used in the main analysis are both under inclusive and over inclusive. That is, some terms on the list many not be sensitive to some, and there are likely to be search terms outside of the list that are considered highly sensitive to others. To explore whether the selection of terms is driving the results, I ran the regressions reported in Table 1 with random draws of twenty terms from the full set of 97 sensitive terms that Marthews and Tucker crowd sourced.[46] This procedure was performed one hundred times for each window to generate a sample of parameter estimates for *Sensitive*GooglePolicy*. Observing a large variance in the empirical distribution of this parameter, or a wide disparity between the parameter estimates presented in Table 1 and the empirical distribution of *Sensitive*GooglePolicy*, would call into question the robustness of the main results.[47] Figures 3-5 show the distributions of estimates of *Sensitive*GooglePolicy* for +/-6, +/-2, and +/-1 month windows.[48] These figures also include a scatterplot of the significance levels (versus a null of zero) of these parameter estimates.

The average (median) parameter estimate for the +/-6-month window is .318 (.173), which is significantly smaller than the main estimates in Table 1 (1.96). Estimates range from -4.34 to 6.86, with distribution split nearly evenly between positive and negative estimates. As the significance plot shows, almost all (94%) of these estimates are statistically indistinguishable from zero ($p < .10$). The average (median) parameter estimate for the +/-2-month window is 1.51 (1.49), which is opposite sign of that reported in Table 1 (-.666). Further, 80 percent of these estimates are greater than zero, although only 6 percent are statistically different than zero. The average (median) parameter estimate for the +/-1-month window is -1.72 (-1.70), which is slightly smaller (in absolute value) than the main estimates (-2.41). 89 percent of the parameter estimates are negative, and 22 percent of these estimates are statistically significant.

---

[46] *See* Marthews & Tucker, *supra* note 41, Appendix A, Tables 14-15.
[47] National level data are used due to resource constraints. Because there the parameter estimates of *Sensitive*GooglePolicy* in pooled and fixed effects models are nearly identical, using national data should lead to very similar results as a state-based sample.
[48] An empirical distribution was also generated using a bootstrap method, with estimates the model based on random samples (with replacement), stratified based on search terms. This meant that each model is estimated repeatedly with a different mix of sensitive *and* non-sensitive terms. Results are very similar to the random draw method.

Figure 3

Random Sample Estimates: +/-6 Month Window

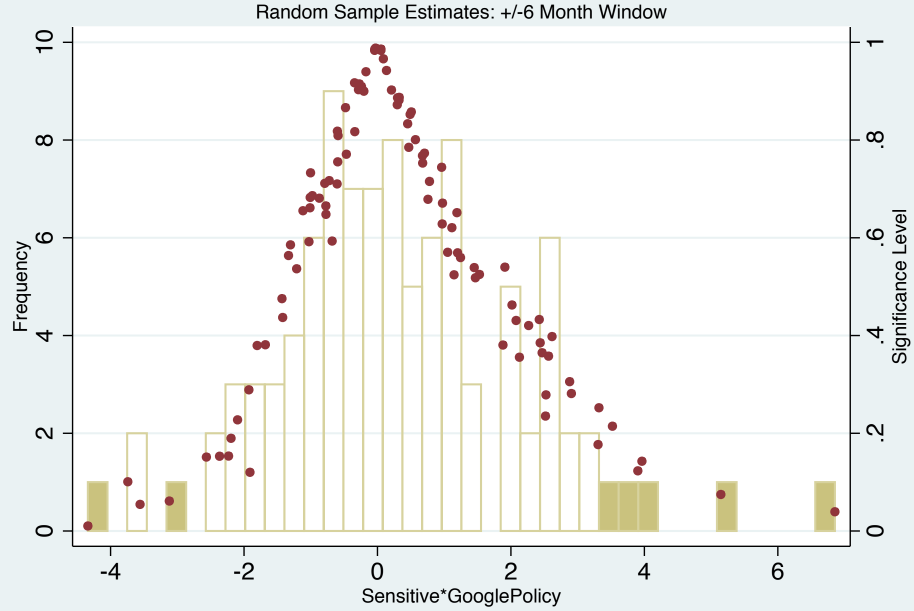
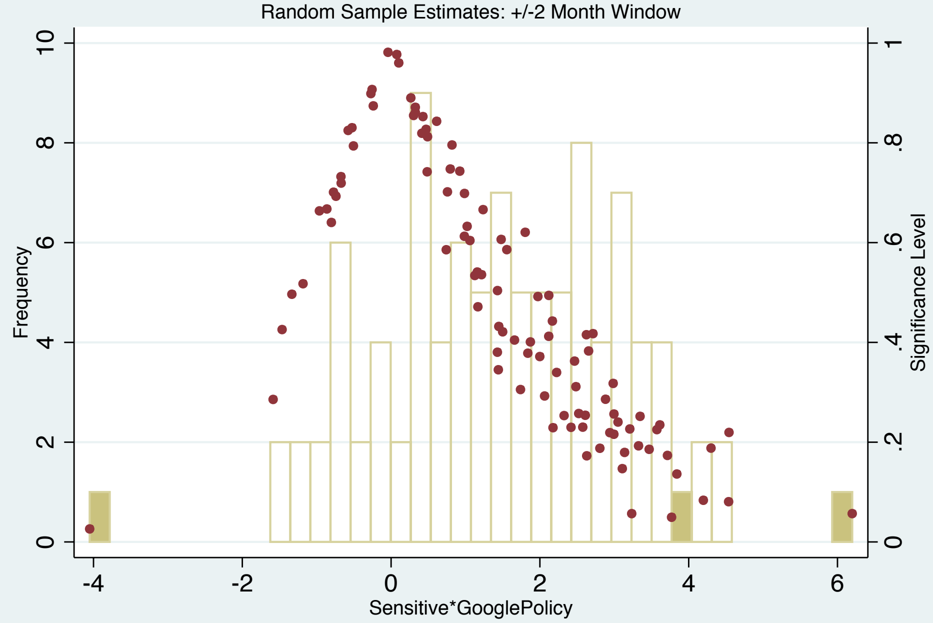
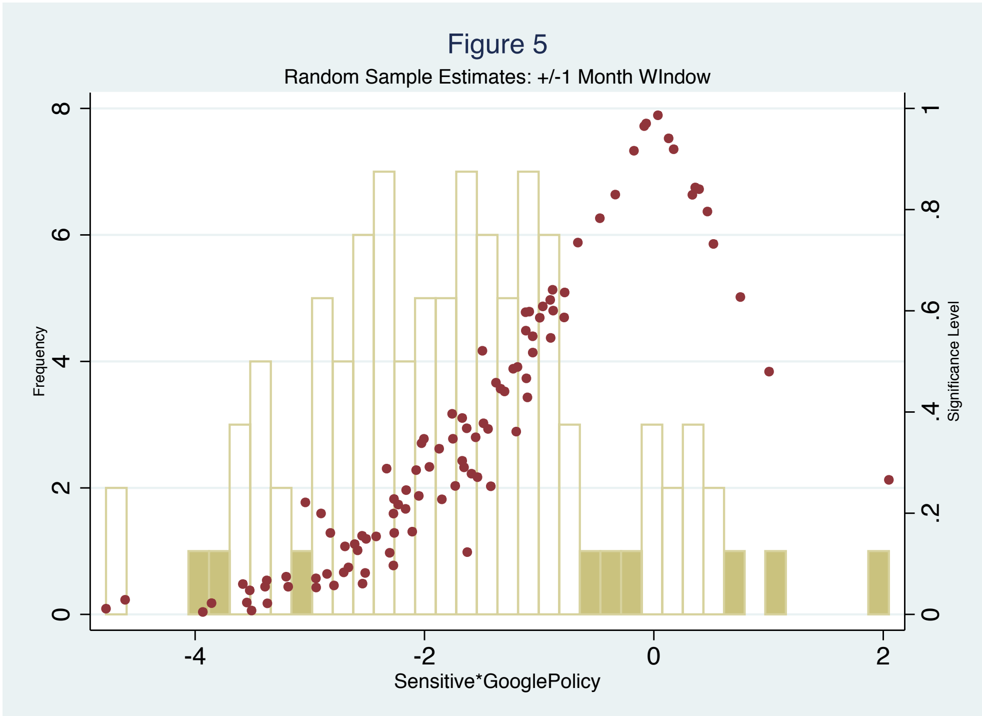


Figure 4

Random Sample Estimates: +/-2 Month Window





Overall these results lend confidence to the main estimates. Although the parameter estimates do appear sensitive to the mix of sensitive terms chosen, like the main estimates, the average random sample estimates for the +/-6-month are positive and insignificant. Although the main +/- 2-month estimate is signed differently than the average parameter estimate from the random sample, they are both statistically indistinguishable from zero, and could have come from the same distribution.[49] And in any event, the random sample results strongly reject a negative impact on sensitive search from Google's policy change for the six- and two-month windows. The consistent negative sign on the +/-1 month window estimates provides confidence that there was a reduction in sensitive search relative to non-sensitive search during over the +/-1-month window, even if the point estimates are dependent on the exact mix of search terms.

4.3.2 Placebo Estimation

To examine the possibility that the measured impact from the one-month window was an artifact of seasonality, specifications were run that falsely define the Google policy change data as March 1, 2013 and March 1, 2011, respectively. Regressions were rerun on the full sample, as well as subsamples for privacy demanding states. Results are reported in Table 5

[49] With 100 replications, at a 95% confidence level, one expects 5 estimates to be significantly greater than zero by chance even if the true parameter is zero.

TABLE 5
PLACEBO CONTROLS FOR +/-1-MONTH WINDOW

	All States		Any Privacy		High Privacy	
	2011	2013	2011	2013	2011	2013
*Sensitive * GooglePolicy*	-2.96** (1.32)	1.13 (1.43)	-2.85** (1.39)	1.21 (1.43)	-3.36** (1.60)	.802 (1.57)
Fixed Effects	Y	Y	Y	Y	Y	Y
R^2	.706	.725	.706	.728	.737	.767
N	16,624	18,702	11,320	12,735	1,672	1,881

Dependent variable is Google Trends index score for term *i*, in state *j*, during week *t*. Estimates based on one month before and after Google policy change. Fixed Effects include search term, state, and week effects. Robust standard errors clustered on search term are reported in parentheses.

The 2013 regressions are positive and insignificant, as expected. The 2011 placebo results, however, are similar to those reported in Tables 1 and 3 for the true policy change window—negative, statistically significant, ranging from -2.85 to -3.36. A news search around the estimation window reveals some events involving Google and privacy, which could be responsible for the impact. For example, on March 1, 2011, it was reported that Google accidentally deleted several thousand emails.[50] On March 30, 2011, Google entered into a consent agreement with the FTC to settle privacy issues related to Google Buzz.[51] And on March 7, it was reported that Google was dealing with a virus affecting Android apps.[52] Further, in April 2013, Google was dealing with fallout from revelations involving Android geo-location tracking.[53] However, given that none of these events directly implicated consumer privacy using search, it seems surprising that they would produce a larger and more statistically significant impact on sensitive search than the 2012 policy change. Although the null result in 2013 militates against seasonal factors, without additional information, we cannot rule out the possibility that at least part of the measured impact in 2012 is merely seasonal.

5. Discussion and Conclusion

Privacy is complex and multidimensional. As such, invasions of privacy often involve subjective harm, which is inherently difficult to measure. One important dimension of privacy is autonomy: a space in which one can be free to make certain

[50] http://www.cnn.com/2011/TECH/web/03/01/gmail.lost.found/
[51] https://www.cnet.com/news/google-settles-ftc-charges-over-buzz/
[52] http://latimesblogs.latimes.com/technology/2011/03/google-removing-virus-infected-android-apps-from-phones-tablets-promises-better-secutiry.html
[53] http://usatoday30.usatoday.com/tech/news/2011-04-24-apple-iphone-google-android-tracking.htm

choices away from the gaze and judgment of society. Not only does autonomy give rise to personal growth, but it also can be instrumental to society. This paper attempts to measure autonomy losses by examining changes in consumer search behavior after Google's 2012 privacy policy change.

The identification strategy employed in this paper appears to have been successful at measuring a change in behavior resulting from an increase in surveillance, and a concomitant reduction in autonomy. Taken as a whole, the results suggest that there was a relatively small (4-5%) and brief (1 month) reduction in sensitive search relative to non-sensitive search, a magnitude similar to those from Marthews & Tucker's study of the impact of the Snowden revelations on searching terms that raise suspicions with the US government. Surprisingly, I find no difference in privacy response among states with high- and low-demands for privacy, as measured by privacy laws. The results are consistent with some consumers, at the margin, withholding sensitive searches from Google for fear of unwanted observation. That the impact appears to be short-lived is not surprising. As with other studies of the impacts on negative events on company performance, fundamentals tend to return in short order.[54]

It is important to note some of the potential limitations of this study. First, Google Trends is not an absolute measure of search volume, so the results only speak to directional changes, not magnitudes. Thus, there is no way to measure the economic significance of the measured short-term reduction in sensitive search. Second, although sampling techniques were designed to make the selection of terms as objective as possible, and checks suggest that the results are robust to the selection of sensitive terms, there is no perfect way to cover the universe of potentially sensitive search terms, or to capture the differences across the population regarding the sensitivity of these terms. Third, ideally, identification would have been based on random selection of a treatment group, which would have had their sensitive searches observed. Unfortunately, Google's new privacy policy went into effect nationwide (and worldwide) at the same time, leaving no variation in treatment. Thus, the identification necessarily relies on treatment and control behavior (non-sensitive and sensitive search) that was selected by the researcher. Finally, the results of the 2011 placebo test means that without

[54] *See*, *e.g.*, Myung Ko & Carlos Dorantes, *The Impact of Information Security Breaches on Financial Performance of the Breached Firm: An Empirical Investigation*,17 J. INFO. TECH. MGM'T 13 (2006); Acquisti, A., A. Friedman, and R. Telang, *Is There A Cost to Privacy Breaches? An Event Study*, *In* TWENTY SEVENTH INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS (2006); Lawrence A. Gordon et al., *The Impact of Information Security Breaches: Has There Been a Downward Shift in Cost*?, 19 J. COMP. SECURITY 33 (2011) (finding that stock market costs of security breaches has fallen due in part to a decreased tendency of consumers to refrain from doing business with companies that suffer security breaches); Katherine Campbell et al., The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market, 11 J. Comp. Security 431, 434 (2003) (explaining the short-lived impact on stock prices from data breaches); Guy Kaplanski & Haim Levy, *Sentiment and Stock Prices: The Case of Aviation Disasters*, 95 J. FIN. ECON. 174 (2010)(finding evidence that stock prices rebound quickly after large falls in stock prices immediately following aviation disasters).

additional information, the possibility that seasonality is responsible for at least a portion of the measured short-run results cannot be ruled out.

Despite these potential limitations, the findings in this paper nonetheless have some important policy implications. First, they indicate that consumer choice appears to works when it comes to privacy. The policy change was widely publicized and gave consumers over a month of notice; those who were uncomfortable with the change had time to alter their behavior at little cost, and it appears that a small proportion did. As is the case when the price of any good increases, marginal consumers will exit the market, and the remaining inframarginal consumers continue to buy, albeit with less surplus. Thus, although we cannot directly measure the loss in autonomy, they appear to be small or swamped by benefits from combining personal data across Google platforms for most consumers.

Second, and more generally, the results here are consonant with a host of research suggesting that consumers are not terribly concerned with the type of data sharing involved in the day-to-day functioning of the online ecosystem that relies on advertising. For example, several studies find that consumers are willing to share sensitive data for very little compensation.[55] Similarly, recent work suggests that even when reading a privacy policy is made easy—for example, through a warning label approach—few consumers bother.[56] And, those who take the time to understand a firm's data practices—even ones that may be highly invasive—still provide the same amount of information, and are generally unwilling to pay to avoid sharing information.[57] The empirical analysis in this study likewise finds that few consumers were concerned with the policy change as measured by the long-term impact on their willingness to share sensitive information through search queries. That there is no long-run reduction in sensitive search suggests that despite initial misgivings—perhaps driven in part by the privacy advocacy community's vocal outcry—consumers quickly became comfortable with privacy implications of the new policy. The lack of a long-term impact on sensitive searching also may reflect

[55] For example, one study finds that consumers are willing to pay only an additional $1-$4 for a hypothetical smartphone app that conceals location, contacts, text content, or browser history from third-party collectors. Scott Savage & Donald M. Waldman, *The Value of Online Privacy* (2013), *at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2341311. *See also* Dan Cvrecek, Marek Kumpost, Vashek Matyas & George Danezis, *A Study on the Value of Location Privacy*, Proceedings of the 5th ACM Workshop on Privacy in the Electronic Society (2006); Hal R.Varian, Glenn Woroch & Fredrik Wallenburg, *Who Signed Up for the Do Not Call List?* (2004), at http://eml.berkeley.edu/~woroch/do-not-call.pdf; Ivan P. L. Png, *On the Value of Privacy from Telemarketing: Evidence from the "Do Not Call' Registry* (2007), *at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1000533. For a full review of this literature *see* Alessandro Acquisti *et al.*, *The Economics of Privacy* , J. ECON. LIT. at 41 (*forthcoming*, 2017), *at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2580411.

[56] *See* Omri Ben-Shahar & Adam S. Chilton, *Simplification of Privacy Disclosures: An Experimental Test*, (Apr. 2016), *at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2711474

[57] *See id.*; Lior Strahilevitz & Matthew B. Kugler, *Is Privacy Policy Language Irrelevant to Consumers ?*, J. LEG. STUD. (*forthcoming* 2017), *at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2838449.

Google's privacy benefits.[58] Although for some, having sensitive search data combined with data from other Google services is not ideal from a privacy perspective, being observed by a faceless algorithm that will use the information to target ads and customize services still may pose fewer privacy concerns than accessing sensitive information in other, more public ways. In this way, Google search—even under the new privacy policy—perhaps allowed for more autonomy than the available alternatives.

[58] *See* Benjamin Wittes & Jodie Liu, *The Privacy Paradox: The Privacy Benefits of Privacy Threats, Brookings Institute* (May 2015), *at* https://www.brookings.edu/research/the-privacy-paradox-the-privacy-benefits-of-privacy-threats/.

Appendix

TABLE A1
SENSITIVE AND NON-SENSITIVE SEARCH TERMS

Sensitive Terms	Average Trends Score (Jan.1, 2011 – Dec. 31, 2013)	Non-Sensitive Terms	Average Trends Score (Jan.1, 2011 – Dec. 31, 2013)
Abortion	49.1	Amazon	53.0
Acne	69.5	Apple	44.6
Adultery	37.9	Calculator	77.7
AIDS	48.0	CNN	30.3
Bankruptcy	54.8	Craigslist	73.7
Coming out	54.7	Ebay	81.0
Depression	66.8	Espn	56.3
Divorce	29.0	Facebook	77.1
Erectile Dysfunction	49.4	Games	63.0
Escort	65.6	Google	69.0
Gay	60.3	Iphone	37.4
Herpes	64.0	Mail	83.5
HIV	42.6	Maps	71.6
KKK	36.7	Netflix	58.9
Liposuction	45.5	News	53.8
Porn	85.8	Obama	11.7
Sexual Addiction	36.6	Target	40.6
Strip Club	54.7	Walmart	37.1
Suicide	49.1	Weather	39.7
Therapist	66.3	Yahoo	84.3
White power	43.2	Youtube	75.9
Total	52.8		58.1