

# The privacy paradox II: Measuring the privacy benefits of privacy threats

Benjamin Wittes and Emma Kohse

## INTRODUCTION



**Benjamin Wittes** is a senior fellow in Governance Studies at The Brookings Institution. He co-founded and is the editor-in-chief of the Lawfare blog.



**Emma Kohse** is a J.D. Candidate at Harvard Law School, where she serves as editor-in-chief of the Harvard International Law Journal, and a 2012 graduate of Georgetown University's School for Foreign Service.

In 2015, one of the present authors, writing with Jodie Liu, published a Brookings paper entitled, “The Privacy Paradox: The Privacy Benefits of Privacy Threats,” advancing a simple thesis that cuts dramatically against the grain of contemporary privacy thinking: the very technologies that most commentators see as posing grave threats to privacy, the paper argued, in fact offer significant privacy benefits to consumers.<sup>1</sup>

While other scholarship has highlighted empirical data indicating that concerns over privacy do not seem to dampen enthusiasm for these new technologies, scholars have generally attributed this phenomenon to user value judgments prioritizing convenience or efficiency or service delivery over privacy.<sup>2</sup> The “Privacy Paradox” paper proposed an alternative explanation: countervailing privacy concerns may be a significant part of the consumer value judgment. In other words, it hypothesized that people who buy condoms online do so not just because they may be cheaper or more convenient to buy that way, but also, perhaps even most importantly, because of the privacy benefits of the online transaction.

These benefits are often invisible to privacy advocates and scholars who tend to focus their concerns on large remote corporate or government entities that collect data on consumers. As such, an online transaction that involves the creation of a digital footprint for a given person purchasing condoms and having those condoms shipped to a particular address may seem like pure privacy harm. So too might seem a Google search about a sensitive medical condition, a search which shows up in a database of an individual user’s search history that can then be tapped by investigators or litigants with appropriate legal process. So too might a decision to read a book on an Amazon Kindle, which creates records of which pages a user has read and which passages she found particularly noteworthy. But individuals, the 2015 paper argued, may be more concerned with keeping sensitive information from specific people: from neighbors, friends, parents, teachers, or community members. They may

even, somewhat irrationally, be more concerned about the woman behind the counter at CVS, with whom the purchaser actually has to interact in order to purchase condoms in those cash transactions that privacy advocates regard as more protective of privacy than online purchasing. Looking that person in the eye is hard for lots of people, after all. So such individuals, the paper hypothesized, might be willing to trade a certain degree of privacy from remote entities like corporations or governments in exchange for greater privacy from the people immediately around them from whom they have secrets to keep.

[Results] indicate that there is a quantifiable difference in consumer preferences that cannot be explained by factors like convenience and that likely reflect a privacy preference—at least some of the time—for doing business with remote entities that collect data, rather than immediately-present people who might judge us.

The initial paper laid out this idea on an intuitive level and supported it with anecdotes, some initial data, some circumstantial evidence, and a certain degree of personal experience. In this paper, however, we seek to sharpen the picture by testing the theory empirically using Google Surveys, an online tool that allows users to create and administer online surveys to a representative or targeted group of respondents.<sup>3</sup>

For a variety of different behaviors with respect to online shopping, self-checkout purchases, and e-reader use, we asked pairs of questions about different items, one of which was likely to trigger privacy concerns (condoms, for example), and the other significantly less so (dental floss, for example). Any efficiency or convenience benefits would presumably occur equally with both types of products, but the privacy effects are likely to be greater with products of a sensitive nature.

In this paper, we report the results for five such pairs:

- First, we tested the idea that more readers of *Fifty Shades of Grey* would prefer to do so on an e-reader than would readers of a substantially less titillating novel, *The Hunger Games*. Despite the privacy risks associated with online purchases generally and e-readers specifically, significantly more people who read *Fifty Shades* reported doing so on an e-reader than among those who read *The Hunger Games*.
- Second, we asked respondents about their preferred shopping habits. Very few survey respondents preferred to buy general household items online, but almost double that number preferred to buy products of a sensitive personal nature online.
- Third, many more women who bought or considered buying a “personal massager” preferred to do so online than did women who were considering purchasing an electric fan.
- Finally, we tested the hypotheses that young men and women would rather use self-checkout to buy products like condoms and tampons for reasons beyond convenience and efficiency. The effect here was notably weaker. Many consumers reported having no preference, though more respondents in both groups reported a preference for a human cashier when buying dental floss than when buying products they might consider embarrassing.

These results, which we lay out in detail in this paper, strongly support the thesis that consumers may have active privacy interests in dealing with the very remote, data-collecting companies toward which our privacy fears are

generally directed. They indicate that there is a quantifiable difference in consumer preferences that cannot be explained by factors like convenience and that likely reflect a privacy preference—at least some of the time—for doing business with remote entities that collect data, rather than immediately-present people who might judge us. Without taking this difference into account, the picture of the lived experience of privacy for hundreds of millions of real people, as opposed to the hypothesized privacy priorities of self-appointed privacy watchdogs and scholars, will necessarily remain incomplete. So too will our understanding of the privacy risks and benefits of doing business with large entities like Facebook, Google, and Amazon, which we tend to discuss in the language of privacy threats but which also are providing privacy benefits to consumers every day.

This paper proceeds in five parts. In the first section, we give a brief overview of the 2015 paper, its thesis, the hypotheses proposed that at the time lacked more than anecdotal support, and the academic reaction to it. In the next part, we offer an overview of Google Surveys, the tool that we used to gather quantitative data to test some of these hypotheses. We then describe the design of the study and our methods for developing and refining survey questions. As we shall explain, there are some significant problems with polling people about their privacy preferences directly, and Google Surveys has some significant limitations. As a consequence, we self-consciously did not ask about privacy preferences or attitudes but only about behaviors, and we did so without telling the user that we were trying to gauge attitudes about privacy. We also had to word certain questions in a suboptimal fashion to get around the somewhat prudish rules associated with Google Surveys. In the next section, we detail the results of the Google Surveys. In the paper's final part, we offer some conclusions and recommendations for further area of research.

## THE “PRIVACY PARADOX” THESIS AND THE RESPONSE TO IT

The core proposition advanced in “The Privacy Paradox” was that “the American and international debates over privacy keep score very badly and in a fashion gravely biased towards overstating the negative privacy impacts of new technologies relative to their privacy benefits.”<sup>4</sup> That is, many of the very technologies that privacy advocates find so concerning have both privacy-threatening *and* *privacy-enhancing* effects. Our debate, however, intellectually stacks the deck by largely ignoring or dismissing the privacy benefits of new technologies while agonizing endlessly over their privacy costs. While not denying or minimizing any of the hypothesized costs, “The Privacy Paradox” attempted to fill a gap in the research by focusing on the privacy boons that new technologies create for consumers.

It further argued that while the privacy community tends to ignore these benefits, consumer behavior indicates that individuals take them very seriously when making decisions about how to read, where to shop, and where to look for information.<sup>5</sup> Weighing privacy against privacy in this way, many individuals will choose to give sensitive data to remote entities in exchange for greater privacy from those around them.

Some preliminary data and a bunch of anecdotes supported this thesis in various settings. The paper looked at the way Google automatically completes user searches for sensitive subjects, for example, by way of showing that people routinely use Google searches to garner information on sensitive medical and personal conditions.<sup>6</sup> It looked at online

While not denying or minimizing any of the hypothesized costs, “The Privacy Paradox” attempted to fill a gap in the research by focusing on the privacy boons that new technologies create for consumers.

shopping, noting the prevalence of online retailers devoted to sensitive personal products like condoms.<sup>7</sup> It noted preliminary data suggesting that people were more comfortable checking out library books of a sensitive nature using self-checkout machines<sup>8</sup> and some anecdotal suggestions that many people prefer buying condoms using self-checkout machines as well.<sup>9</sup> It noted press reports about the relative popularity of the book *Fifty Shades of Grey* on e-readers, as opposed to in physical copies.<sup>10</sup> And it looked at data regarding pornography consumption to show that people are willing to give all sorts of data about themselves to remote, online pornography merchants in exchange for the ability to consume pornography privately with respect to the people around them;<sup>11</sup> the highest rates of consumption of gay pornography online in the United States turn out to be in the Deep South.<sup>12</sup>

But the original “Privacy Paradox” paper had gaps. While it offered the thought experiment of whether condom sales would rise in CVS stores that install automatic self-checkout machines, major retailers were unwilling to release data on the effects of self-checkout on purchasing behavior. So the hypothesis, while intuitively sensible, was hard to measure.

Perhaps because it cuts so deeply against conventional thinking about privacy and Big Data technologies, and perhaps because it offered a more theoretical than empirical approach, the “Privacy Paradox” thesis entered the debate with more of a thud than a bang. New academic research on privacy since its publication has largely not

Consumer behavior indicates that individuals take [privacy benefits] very seriously when making decisions about how to read, where to shop, and where to look for information.

engaged the thesis. But there are a few exceptions. In an article published in the *University of Chicago Law Review*, Columbia law professor David Pozen discusses what he terms “privacy-privacy tradeoffs,” including “directional tradeoffs” in which a privacy threat is “redirected so that it comes from one source rather than another.”<sup>13</sup> As an example of such redirection, Pozen describes the privacy benefits of a Kindle e-reader in protecting a reader from his fellow subway passengers’ prying eyes, in exchange for the increased privacy threat of Amazon’s close tracking of reading habits<sup>14</sup>—an example that also features in the original “Privacy Paradox” paper. As does that paper, Pozen’s notes that

though “[p]rivacy is constantly being juxtaposed with competing goods and interests, balanced against disparate needs and demands,” the fact that “privacy also clashes with itself” warrants more attention than it gets.<sup>15</sup> Pozen, however, urges more focus on how “interventions that strengthen privacy on one margin can end up weakening it on another.”<sup>16</sup> This characterization is somewhat the inverse of the focus of the original paper, but it amounts to the same thing: technologies, like policies, often give privacy with one hand and take it away with the other.

In her paper on personal data collection by private companies, Rebecca Lipman also addresses the thesis, acknowledging that “it is definitely a privacy gain to keep your medical concerns or pornography preferences away from people you know, even if the tradeoff is sharing that information with Google.”<sup>17</sup> But Lipman uses a kind of nomenclature trick to minimize these privacy benefits. Americans are concerned not just about keeping their sensitive information away from others, she argues, but also about the tracking or recording of even non-sensitive behaviors or information.<sup>18</sup> So privacy, for Lipman, is having the capacity to protect *all* information, not just the sensitive material, from recording by companies and the government, while she characterizes protecting sensitive information from family members, neighbors, and storekeepers as something else: “secrecy.”

In her usage, “secrecy” is a relatively narrow and comparatively unimportant subset of “privacy,” and it seems to follow that gains in secrecy are goods of a smaller magnitude than are the goods lost when privacy as Lipman defines it erodes. The advantage of “Google keeping our embarrassing secrets” is thus balanced against the harm of Google’s access to the larger category that includes both secrets and everyday, non-sensitive information.<sup>19</sup> Furthermore, she argues that while “we are fairly adept at protecting our privacy in the physical world,” we are less able to do so online. And there, she argues, the consequences are “greater” because the information may be personally identified and stored for an indefinite period of time, probably longer than an eavesdropper’s memory.<sup>20</sup> But this is really a value judgment on Lippman’s part about what’s important, and it’s a value judgment that the “Privacy Paradox” hypothesized that millions of consumers do not share. Indeed, for an individual, the dignitary consequences of revealing sensitive information to a cashier may be more concrete, and may even be greater, than the abstract consequences of long-term personal data storage by Amazon.

The thesis has also sparked some discussion outside of academia. Journalist John Herrman, writing for *The Awl*, characterizes the “Privacy Paradox” as describing a “worry” that the ledger of “technology’s privacy impacts...doesn’t fairly represent the semi-related privacy boons we enjoy thanks to some of the most powerful companies in the world.”<sup>21</sup> In Herrman’s view, this faultiness is only a problem for these powerful companies, and not for consumers.<sup>22</sup> The implication, then, is that perhaps any inaccuracy in the weighing of privacy pros and cons that underestimates the pros provided by all-powerful companies is not such a bad thing. In fact, he worries that a version (or perversion) of the “Privacy Paradox” thesis might be used by companies as a means of “misdirection” from the privacy concerns of the new technologies they introduce to the market.<sup>23</sup>

Tim Cushing, a writer at the blog TechDirt, sees the threat as located not in the large companies, as does Herrman, but in governmental access to information collected by those companies.<sup>24</sup> While he acknowledges the “Privacy Paradox” thesis as “solid,” he argues that the paper “fails to closely examine government surveillance concerns.” While companies offer concrete benefits in exchange for the privacy losses caused by their technologies, the government only offers the intangible benefit of “security.”<sup>25</sup>

Whatever the merits of these criticisms, it is notable that neither argument disputes that technology’s privacy benefits are being undercounted in the discussion: Herrman is concerned about the potential consequences of more successful scorekeeping, and Cushing is arguing that the privacy harms side of the equation should be larger to take into account governmental surveillance.

On a discussion board on the website *Hacker News*, a social news website run by the startup incubator Y Combinator,<sup>26</sup> a different criticism arose. In response to a summary of the “Privacy Paradox” theory, a commenter made the familiar argument that consumers are not aware that they are being tracked by companies like Google and Amazon, or at least are not aware of the extent of the tracking.<sup>27</sup> If they were, the argument continues, then they would care a lot more. And consumer choices that elevate privacy from those around them over privacy from remote entities cannot be evidence of true privacy preferences if they are uninformed choices. But there is a fair amount of empirical evidence that most Americans are, in fact, aware that they are being tracked online,<sup>28</sup> and it is likely that awareness will increase with growing internet literacy and prominent news stories concerning online privacy and security.<sup>29</sup>

In this paper, we aim to sharpen the debate with a series of discrete empirical tests of the “Privacy Paradox” thesis. The goal is both to examine whether, in fact, a gap exists between the privacy expectations and preferences of



ordinary consumers and the theorized privacy expectations of scholars and activists and, to the extent that it does exist, to begin measuring the magnitude of that gap.

## MEASURING PRIVACY BEHAVIOR WITH GOOGLE SURVEYS

Testing the “Privacy Paradox” thesis, however, is difficult. After all, when a person buys *Fifty Shades of Gray* on an e-reader, rather than in a hardcover edition, Amazon does not know whether she did so for privacy reasons or for reasons of convenience, and it doesn’t know either which privacy reasons might have been salient: Is the reader embarrassed about the interaction with a cashier or is she shy about being seen on the subway with the book—or both? Or does she not want her housemates, spouse, or parents to know what she is reading?

Similarly, if a reader prefers a hardcover book, it’s not obvious whether that reader preference reflects an aesthetic or reading-experience preference for physical books when she’s not traveling or whether it reflects a strong aversion to Amazon’s knowing which passages she singles out as especially riveting and which pages she marks.

To glean how lots of people are thinking about these issues, one has to draw inferences from mass behavior, and

There is a fair amount of empirical evidence that most Americans are, in fact, aware that they are being tracked online, and it is likely that awareness will increase with growing internet literacy and prominent news stories concerning online privacy and security.

the companies in question generally do not release data on the mass behavior of their customers. So we don’t know in any kind of granular fashion how the raciness of a book correlates with mass preferences as to whether to read it in physical or electronic form. We don’t know whether people, on average, would rather brave the store clerk and pay cash for condoms or tampons and have no record linking them to the sale or whether they prefer to use a credit card and an address and have those products shipped to them without ever looking anyone in the face. Until companies start releasing data that bear on such questions, we can only look for proxies for these behaviors and preferences.

Surveys represent a complicated proxy for privacy-implicating behaviors. After all, if you ask someone whether he cares about privacy, he’s very likely to say that he does. That

same person, however, is also very likely to engage in all sorts of behaviors that may not comport with those the conventional privacy models would expect of someone who cares about his privacy.<sup>30</sup> The result is that the many studies of consumer attitudes toward privacy show a real gap between people’s stated attitudes and their behaviors. For example, researchers who compared participants’ self-reported opinions about privacy with their behavior during an online shopping experience found no correlation between greater concern for privacy and likelihood of taking privacy-protecting actions.<sup>31</sup> Participants who reported concerns about protecting their privacy online were no less willing to reveal “even highly personal information.”<sup>32</sup> In another study, researchers found that people tended to declare that they would refuse to provide certain personal information to marketers, but did, in fact, reveal that information when asked two weeks later.<sup>33</sup> What people say about privacy does not seem to match what they *do*—even with respect to disclosure of specific personal information.

For this reason, in this study we specifically did not examine consumer *attitudes* towards privacy. That is, we did not ask about what people believe. We asked only about how they prefer to behave, and therefore—since we are working with a proxy—what those preferences say about what they *do*. The idea was to create as good a proxy as we could using survey instruments for the sort of data that companies refuse to release—data that sheds light on the question of whether consumer behavior is guided by, as in the traditional privacy model, a tension between convenience and privacy or, in the “Privacy Paradox” model, by competing and distinct privacy interests.

Like any survey instrument, Google Surveys (formerly called Google Consumer Surveys) is an imperfect and somewhat crude proxy for the sort of nuanced aggregations of privacy and service optimization balancing we are trying to measure. Sample sizes are relatively small, and the methodology is still controversial. On the other hand, Google Surveys allows for inexpensive polling on single questions by people and organizations who are not public opinion professionals.

We asked only about how they prefer to behave, and therefore—since we are working with a proxy—what those preferences say about what they do.

Google Surveys is a public opinion platform that piggybacks off of the world’s addiction to Google. It works by providing small incentives to users to answer individual questions on either their smartphones or in the course of their web browsing. Some web sites use Google Surveys as a gateway to premium content: answer a question in exchange for access. And Google also has a mobile app that allows users to answer questions in exchange for credits on Google Play, the company’s entertainment platform.<sup>34</sup> The idea is to leverage the gigantic sample of people using the internet into a public opinion research tool available to companies that pay Google for access to it. Surveys are easy to create on a simple and user-friendly interface. And results come in quickly.

Studies assessing the accuracy of Google Surveys data have concluded that it is relatively comparable to more traditional survey techniques. Nate Silver’s post-2012 election evaluation of polling accuracy ranked Google Surveys as the second-most-accurate of 23 polling firms, with less than half the average error of well-respected polls like Quinnipiac and Gallup.<sup>35</sup> More recently, his website has given Google Surveys a “B” in its “pollster rankings”—a ranking that is not top grade but certainly respectable.<sup>36</sup> It has been used in academic research in a variety of fields ranging from psychology to computer science.<sup>37</sup>

The Pew Research Center did a comprehensive study in 2012 comparing the results of Google Surveys with those from Pew dual frame (landline and cell phone) telephone surveys, and found that the median difference between the two groups of results was three percentage points, and the mean difference was six percentage points.<sup>38</sup> The researchers attributed some of the difference to differences in the structure and administration of the questions. Because Google Surveys does not use a true probability sampling method (i.e. random selection of respondents), Pew expressed concern about differences in the composition of the sample, but actually found that the sample “appears to conform closely to the demographic composition of the overall internet population.”<sup>39</sup>

Of course, only approximately 84 percent of American adults use the internet,<sup>40</sup> but it seems that the Google Surveys sample is a representative sample of at least this portion of the population. And though heavy internet users are slightly overrepresented in Google Surveys samples, this bias appears to be small.<sup>41</sup> Google’s inferred demographic information was not very accurate for individual respondents, especially with respect to age, Pew found, but the

overall pool is nonetheless representative.<sup>42</sup> Interestingly for our purposes, the largest difference between the Google Surveys sample and Pew sample of internet users was in the percentage who reported seeking medical information online: only 52 percent of the Google Surveys sample, as compared to 71 percent of the Pew respondents—suggesting, at least tentatively, that this is not a universe of people insensitive to privacy concerns relative to others.<sup>43</sup>

Google itself has done substantial research on the accuracy of its methods. In one study, it compared Google Surveys-obtained data and data from probability and non-probability based internet surveys to national media and health benchmarks.<sup>44</sup> Specifically, it compared Google Surveys results and the results of other, more traditional surveys, to very accurate data: viewership information from Video on Demand, Digital Video Record, and satellite dish information, and health information from the Center for Disease Control (CDC).<sup>45</sup> The study found that Google Surveys deviated least from the benchmarks in terms of average absolute error; it also had the smallest absolute error and the highest percentage of responses within 3.5 percentage points.<sup>46</sup>

Still, Google noted the limitations of the platform. The internet population tends to be younger, better-educated, and wealthier than the general population at large, for example.<sup>47</sup> What's more, because Google Surveys asks each respondent only one or two questions, it can be hard to assess the relationships between responses, which do not always involve the same survey samples. In addition, some questions might be regarded as suspicious when they appear to be blocking content on a website, leading to bias in the responses.<sup>48</sup>

To be sure, some reports are more critical, particularly about the unreliability of inferred demographic information, and they suggest that Google Surveys is more properly used as a supplement to probability based surveys.<sup>49</sup> Still, a recent report cautiously concluded that the inferred demographics “may be sufficiently sound for post-stratification weighting and explorations of heterogeneity.”<sup>50</sup> The researchers were able to reproduce the results of four canonical social science studies using Google Surveys, and determined that Google Surveys was likely to be a useful tool for survey experimenters.<sup>51</sup>

## STUDY DESIGN

The ambition of this study was to pose pairs of questions derived from the “Privacy Paradox” thesis, and to thus measure the response to privacy-sensitive matters against the response to non-sensitive but otherwise similar subjects. The idea was to isolate the privacy effect in user preferences and behaviors from those effects associated with efficiency, cost savings, or service delivery. In attempting this, we thought it important that respondents not know they were being asked about privacy but merely about how they behave or what they prefer to do—taking all variables and goods into account.

Working with Google Surveys involved significant constraints, some specifically debilitating to aspects of our project. For example, there are limits on whom surveys can target; age demographics start at 18, which makes it tricky to, say, poll LGBT teens about their use of Google searches to explore their sexual identities. Moreover, Google Surveys generally limits surveys to two questions per instrument, or one screening question to target a specific group, followed by one test question. Additionally, questions have a strict 125-character limit, and answer choices cannot exceed 44 characters.

Most troubling for our purposes, Google must approve the content of all questions, and the company is understandably a little prudish about the subjects concerning which it will and will not let you beam out questions to thousands of smartphone users around the world. The Google Surveys support team explains that they “don’t allow surveys



with content that could be considered adult in nature or contain adult material. This includes adult themes, adult activity, sexually suggestive material or other elements.”<sup>52</sup> Further, “[a]ny surveys regarding any types of contraceptives/birth control is [sic] not allowed.”<sup>53</sup> These restrictions made questions about pornography consumption impossible, and it required the use of euphemisms to refer to some of the products we would have preferred to ask about explicitly. Indeed, most of the “test” questions from our five pairs contained language that was blocked as we originally worded them.

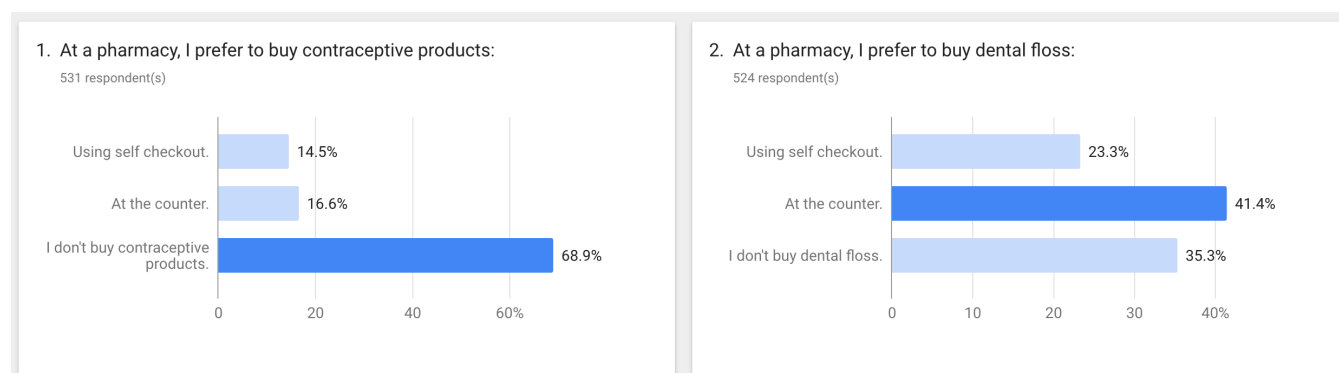
Fortunately, Google Surveys’ censorship is relatively half-hearted and inconsistent. So our challenge was to replace the offending language with phrasing sufficiently euphemistic to be approved by the Google Surveys team, but still specific enough that our data would accurately reflect people’s behavior relating to the kind of sensitive products that Google Surveys prohibited us from asking about by name. So “vibrator” became “personal massager,” which managed to slip past the censors. And “condoms” became “birth control products,” which was rejected in one instance and replaced with “contraceptive products.” This got through the filter twice. Interestingly, however, for an updated survey run two weeks later, this same language was rejected, but “birth control products” survived.

The original versions of some of our questions also had problems with targeting, and with confusing or overbroad options. The result was that we fine-tuned certain questions that originally produced only noise. We report here all of our questions, including those that reflect nothing useful:

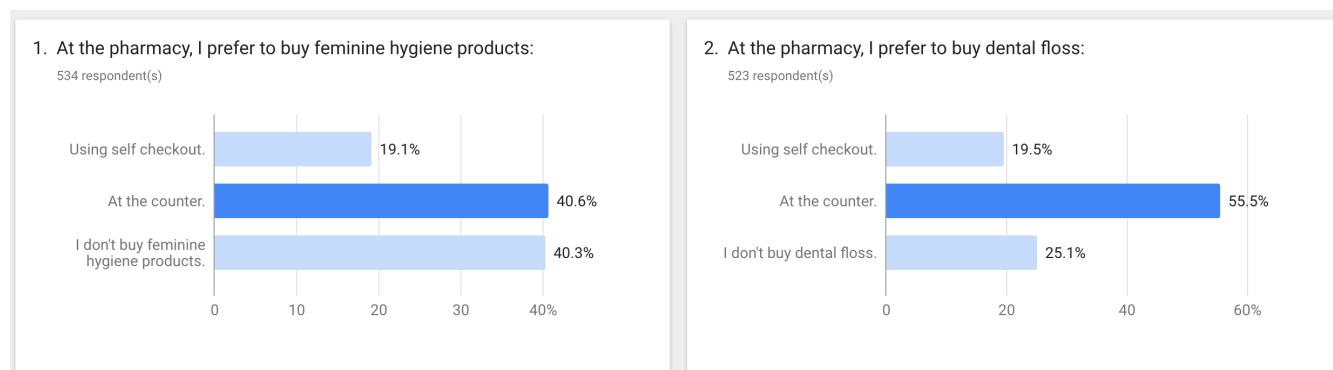
Question #1 screened respondents by asking “yes” or “no” to the following statement: “I have bought or considered buying a personal massager” (2,526 respondents). Among those who answered yes (8.3 percent), the question then asked for a “true” or “false” response to one of the following two statements: “I would rather buy a personal massager online than in a store” (204 respondents) or “I would rather buy an electric fan online than in a store” (204 respondents).

Questions #2 and #3 were failed efforts to measure the privacy value of self-checkout. Question #2 targeted men with what we thought would be a simple pairing of questions. It asked half (531 respondents) whether “At a pharmacy, I prefer to buy contraceptive products:” “At the counter,” “Using self-checkout,” or “I don’t buy contraceptive products.” It asked the other half (524 respondents) the same question about “dental floss.” Question #3 targeted women with a parallel question: “At the pharmacy, I prefer to buy feminine hygiene products:” “At the counter,” “Using self-checkout,” or “I don’t buy feminine hygiene products” (534 respondents). It used the same dental floss control question (523 respondents). Rather to our surprise, these questions produced something of a garble that certainly showed no preference for self-checkout in personally sensitive products:

**Figure 1: Results for Question #2 (Targeted at Men)**



**Figure 2: Results for Question #3 (Targeted at Women)**



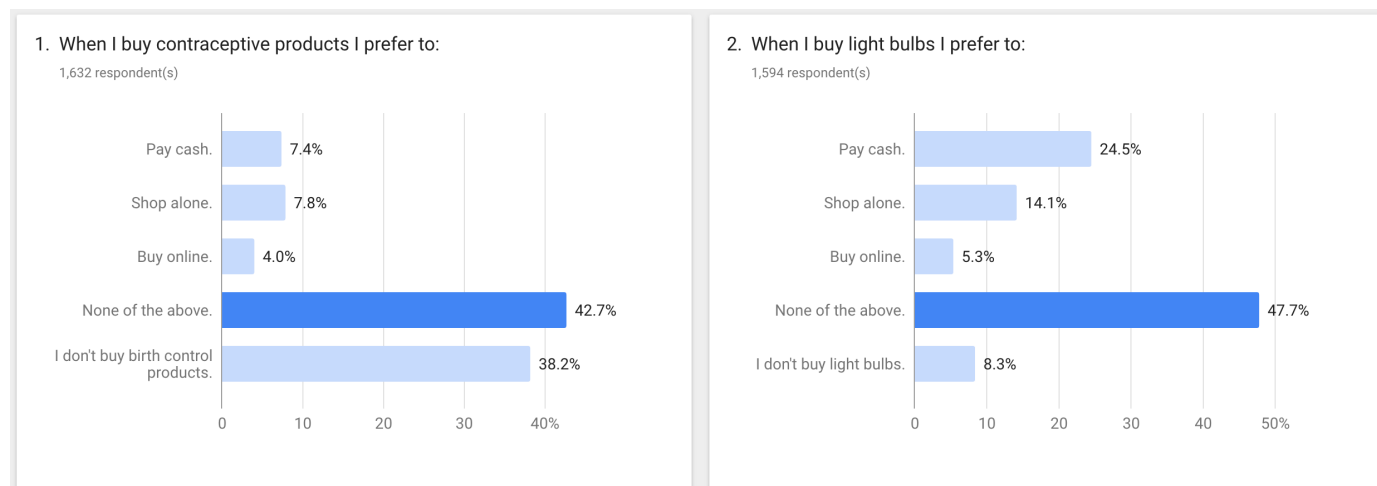
This outcome could be because there is, in fact, no effect present here, but we suspected a different reason: the astonishing percentage of men who claimed that they don't buy "contraceptive products" suggested to us that a great many users did not understand our euphemism for condoms and thought we were asking about prescription contraceptives. Moreover, by not restricting the demographic by age, we may have included enough post-menopausal women to skew our Question #3 sample, and in any event, the number of women who claimed not to buy "feminine hygiene products" was such as to render the sample too small to be useful.

In response, we produced four questions, each surveyed separately, designed to focus only on people who actually purchase the products in question and designed as well to focus on a young age demographic that has every reason to buy the products in question. Question #4 thus screened men between 18-24 years old by asking them the true-or-false question: "I regularly buy over-the-counter birth control products at a store that has an option to pay using an automated self-checkout machine with no human cashier" (1,965 respondents). Among those who affirmed that they made such purchases, the question then asked "When I buy over-the-counter birth control products at this store, I prefer to pay:" and gave three options: "At the counter with a human cashier," "Using the automated self-checkout machine," and "I have no preference" (200 respondents). (The fact that a mere 10.4 percent of the sample passed the screening question strongly suggests that our analysis of the flaws of Question #2 was at least partially correct.)

Question #5 posed the same screening question to the same demographic criteria with respect to dental floss (933 respondents) and asked the same survey question of the 21.7 percent of respondents who passed the screen (200 respondents). Questions #6 and #7 recast the "feminine hygiene product" questions in parallel fashion. Question #6 screened women between the ages of 18-24 with the question: "I regularly buy feminine hygiene products at a store that has an option to pay using an automated self-checkout machine with no human cashier" (527 respondents). It then asked those who affirmed that they did so how they preferred to pay (200 respondents). Question #7 posed the dental floss control question to 18-to-24-year-old women (783 respondents screened with 200 respondents polled).

Question #8, like Questions #2 and #3, also produced only noise and required subsequent refinement. We designed the question to probe whether people were more likely to shop online for sensitive items than for ordinary items. So the survey asked two questions: First, "When I buy contraceptive products I prefer to:" and offered five options: "Pay cash," "Shop alone," "Buy online," "None of the above," or "I don't buy birth control products" (1,632 respondents); and second, it asked the same question about "light bulbs" (1,594 respondents). The trouble, as reflected in Figure 3, is that an overwhelming majority of respondents said either that they did not buy contraceptive products or that they preferred "none of the above" and only a tiny percentage said they preferred buying the product online—making comparison impossible.

**Figure 3: Results for Question #8**



We sought to remediate this problem with Question #9, which approaches the problem from a slightly higher level of altitude in an attempt both to give respondents a wider array of shopping options and to control for people who don't buy birth control products. This question thus asked both "When I buy products of a sensitive personal nature, I prefer to shop:" (1,501 respondents) and "When I buy general household items, I prefer to shop:" (1,517 respondents), and offered the following six choices: "in the most customer-friendly environment"; "in a way that saves the most money"; "online"; "at a small local business"; "at a large national chain"; and "I have no preference."

Finally, Questions #10 and #11 sought to measure user preference for e-readers versus physical books in explicit versus non-explicit literature. As noted in the "Privacy Paradox," the proportion of e-reader sales to total sales of *Fifty Shades of Grey* far outstripped that of other Random House publications.<sup>54</sup> The authors hypothesized that the reason for this phenomenon was the relative privacy afforded by e-readers: individuals embarrassed to be seen reading a novel that glorifies BDSM can read a Kindle version of *Fifty Shades* on the metro or in a café without their neighbors being any the wiser.<sup>55</sup> But high e-reader copy sales alone does not necessarily indicate a preference for the privacy afforded by e-readers: other factors, such as convenience or price, may have driven the sales. If these other factors were significant, we would expect to see a similar proportion of readers consuming other, less sensitive but comparably popular, titles on e-readers. Question #10 thus screened readers by asking them for a true-or-false answer to the following statement: "I have read *Fifty Shades of Grey*" (1,158 respondents). Among those respondents who affirmed that they had read the book, it then asked "I read *Fifty Shades of Grey*:" "In paper copy"; "On an e-reader"; or "As an audiobook" (201 respondents). Question #11 asked the same question about *The Hunger Games* (825 respondents screened with 212 respondents polled).

We thus ultimately created five workable comparison pairs to test various aspects of the "Privacy Paradox" thesis:

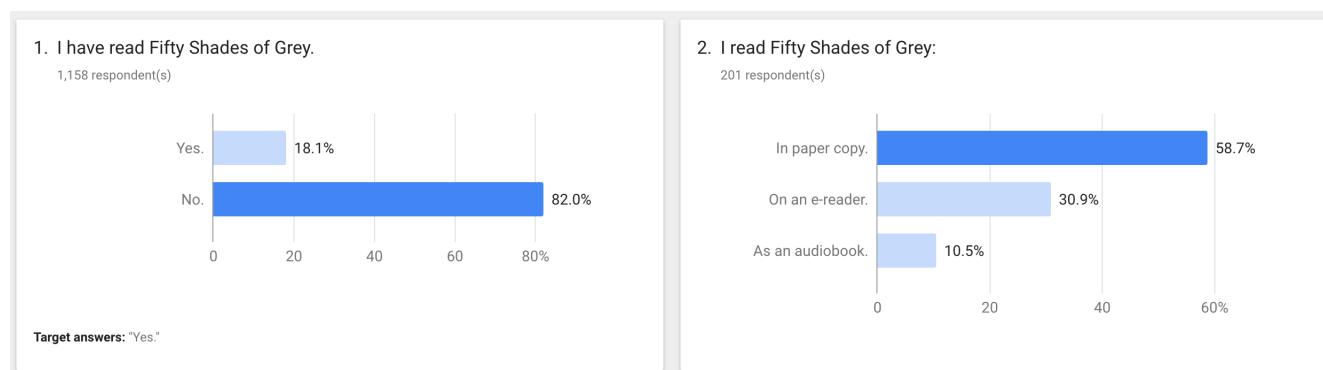
1. Are readers of *Fifty Shades of Grey* (Question #10) more or less likely to read the book on an e-reader than readers of *The Hunger Games* (Question #11)?
2. Do consumers prefer to shop online at a greater rate when shopping for "products of a sensitive personal nature" than when shopping for "general household items" (Question #9)?
3. Do consumers prefer to purchase a "personal massager" online at a greater rate than they do an "electric fan"? (Question #1);

4. Do young men (ages 18-24) who buy “over-the-counter birth control products” (Question #4) prefer to pay with self-checkout at greater rates than men who buy dental floss (Question #5)?
5. Do young women (ages 18-24) who buy “feminine hygiene products” (Question #6) prefer to pay with self-checkout at greater rates than women who buy dental floss (Question #7)?

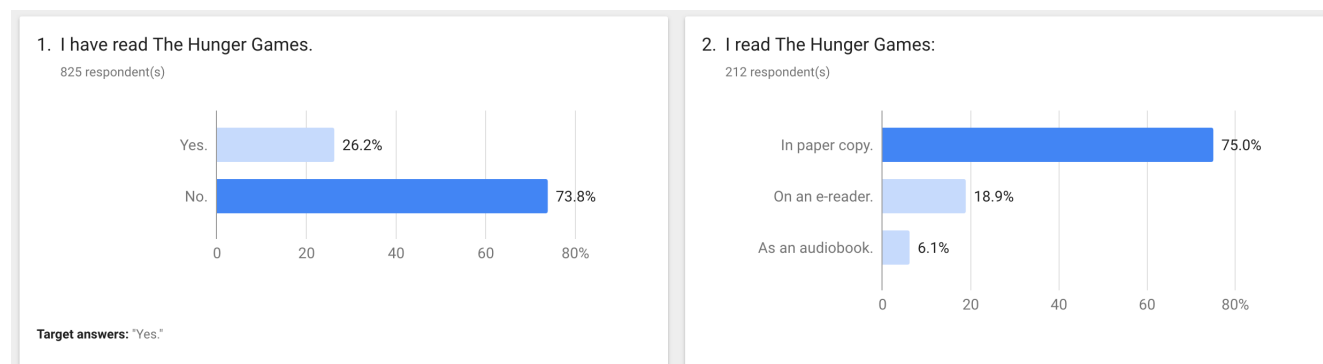
## RESULTS

Perhaps the most dramatic results came from the comparison between *Fifty Shades of Grey* and *The Hunger Games*: Survey respondents who had read *Hunger Games* were significantly less likely to do so on an e-reader than the respondents who had read *Fifty Shades*.

**Figure 4: Results for Question #10**



**Figure 5: Results for Question #11**



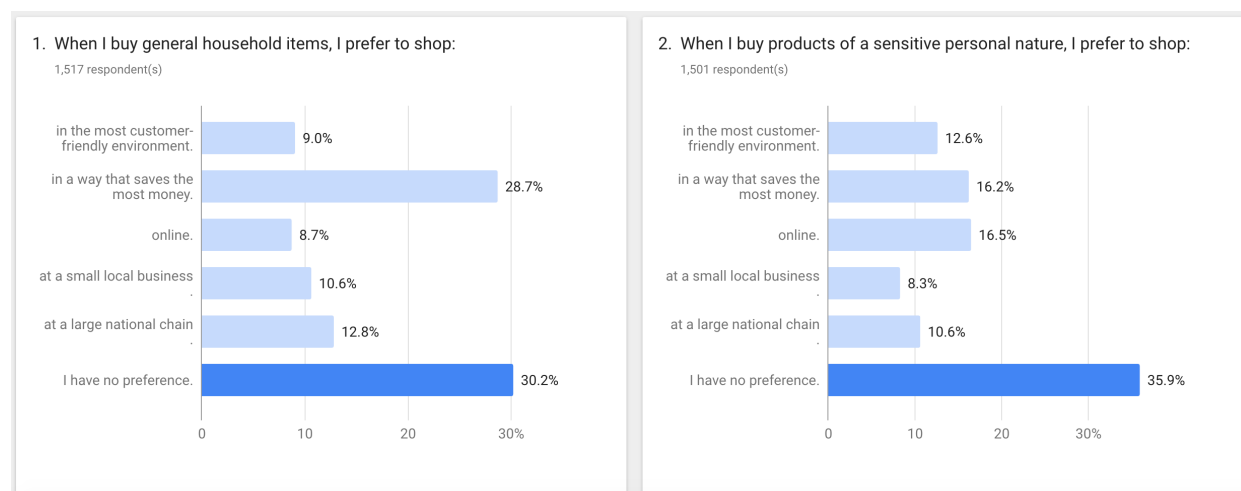
A full 30.9 percent of *Fifty Shades* readers read digital versions of the novel, and another 10.5 percent listened to the audiobook, which is comparably invisible to those around the reader and also involves an online transaction with a remote data-collecting entity. Only 58.7 percent of respondents who read *Fifty Shades* did so in paper copy. In contrast, only 18.9 percent of *Hunger Games* readers reported having read the book on an e-reader and only 6.1 percent listened to the audiobook, while three-quarters read a paper copy. That's more than 16 percentage points higher than *Fifty Shades*.

The result is particularly striking because the demographic differences between *Fifty Shades* readers and *Hunger Games* readers would suggest *less* enthusiasm for e-books—not more—among the *Fifty Shades* readers. We

would expect, after all, the group of *Hunger Games* readers to contain more young adults, who belong to the age group most likely to read e-books.<sup>56</sup> And indeed, the *Hunger Games* sample contained 35 respondents aged 24 and younger and 76 respondents aged 34 and below, while the *Fifty Shades* sample contains 33 respondents aged 24 and younger and 59 aged 34 and younger. In other words, with other factors held constant, one might expect *more* readers of *Hunger Games* to use e-readers and audiobooks, not fewer. It is thus reasonable to infer that the significantly higher consumption of electronic versions of *Fifty Shades* than *The Hunger Games* is, in fact, the result of its sensitive content triggering reader privacy concerns.

In our second set of paired questions, we specifically addressed the proposition that convenience, efficiency, and economic expediency fully explain the widespread use of technologies that allow entities to track consumer information.<sup>57</sup> If this were true, then consumers would be just as likely to buy products like toilet paper and dish soap online as they would condoms; presumably the savings in terms of money and effort occur equally with mundane products as with sensitive ones. However, survey respondents' self-reported preferences reveal substantial differences.

**Figure 6: Results for Question #9**



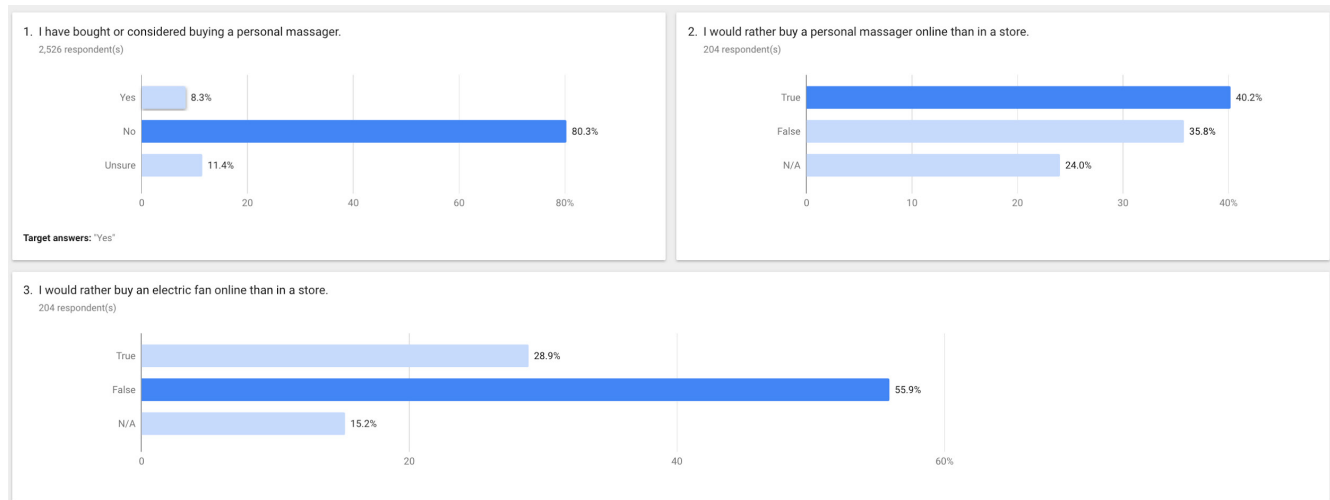
Price dictated the shopping preferences for general household items for 28.7 percent of respondents, whereas only 16.2 percent of respondents indicated that they preferred to shop for products of a sensitive personal nature in the way that saved the most money. A slightly higher proportion—16.5 percent—liked to shop online for such products, whereas only 8.7 percent of respondents preferred to shop online for general household items.

Though around a third of respondents indicated that they had no preference in shopping for either class of items, those that did express preferences did so predictably: People cared substantially less about price and substantially more about shopping online when buying sensitive products than they did when buying household products. Again, this seems to reflect an active preference for giving data to remote entities, rather than suffering the privacy indignities of immediate, face-to-face transactions. Indeed, this question suggests that some customers may even be prepared to *pay more* in order not to have to look someone in the face while buying products of a sensitive nature—even if that means that Amazon.com may be able to compose a more detailed profile of the consumer.

The third pair, the “personal massager” questions, tested a specific example of this more general hypothesis that consumers would prefer to buy sensitive products, as opposed to everyday items, online than in person. Again, the results are striking.



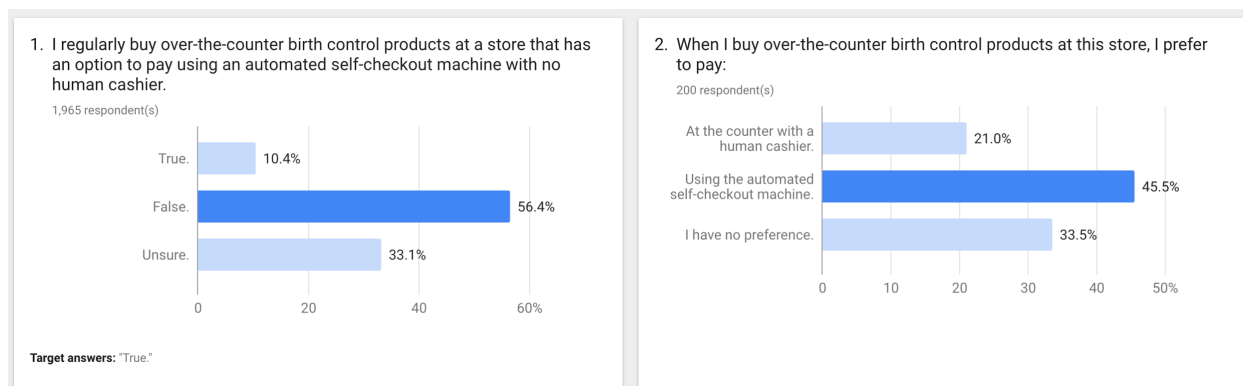
**Figure 7: Results for Question #1**

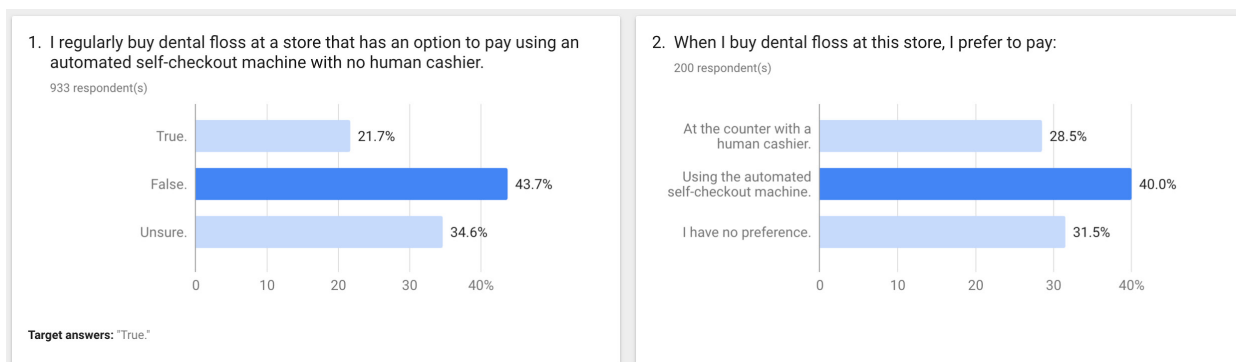


Fully 40.2 percent of respondents who answered “yes” to the screening question reported that it was true that they would rather buy a personal massager online than in a store; 35.8 percent of respondents reported that it was false. This means that about half of people who have bought or considered buying a vibrator and have a preference between in-person and online shopping would rather buy online than in a store. In contrast, a solid majority of these respondents (55.9 percent) responded that it was false that they would rather buy an electric fan online than in a store. And only 28.9 percent of the same group expressed a preference for buying an electric fan online. While the screening question targeted individuals based on whether they had bought or considered buying a personal massager, and did not ask the same question for electric fans, fewer respondents selected the “N/A” option for the question about fans than for the question about personal massagers. If a large portion of targeted respondents had never bought or considered buying an electric fan, thus skewing the results as based on hypothetical, rather than actual, preferences, we would expect a higher proportion to select “N/A.”

Finally, our questions designed to test the hypothesis outlined in the “Privacy Paradox” regarding self-checkout machines supported the thesis less strongly than we expected. Among men aged 18-24 who reported regularly buying over the counter birth control products at a store with a self-checkout option, 45.5 percent preferred to use self-checkout, while only 21 percent preferred paying a human at the counter.

**Figure 8: Results for Questions #4 & #5**

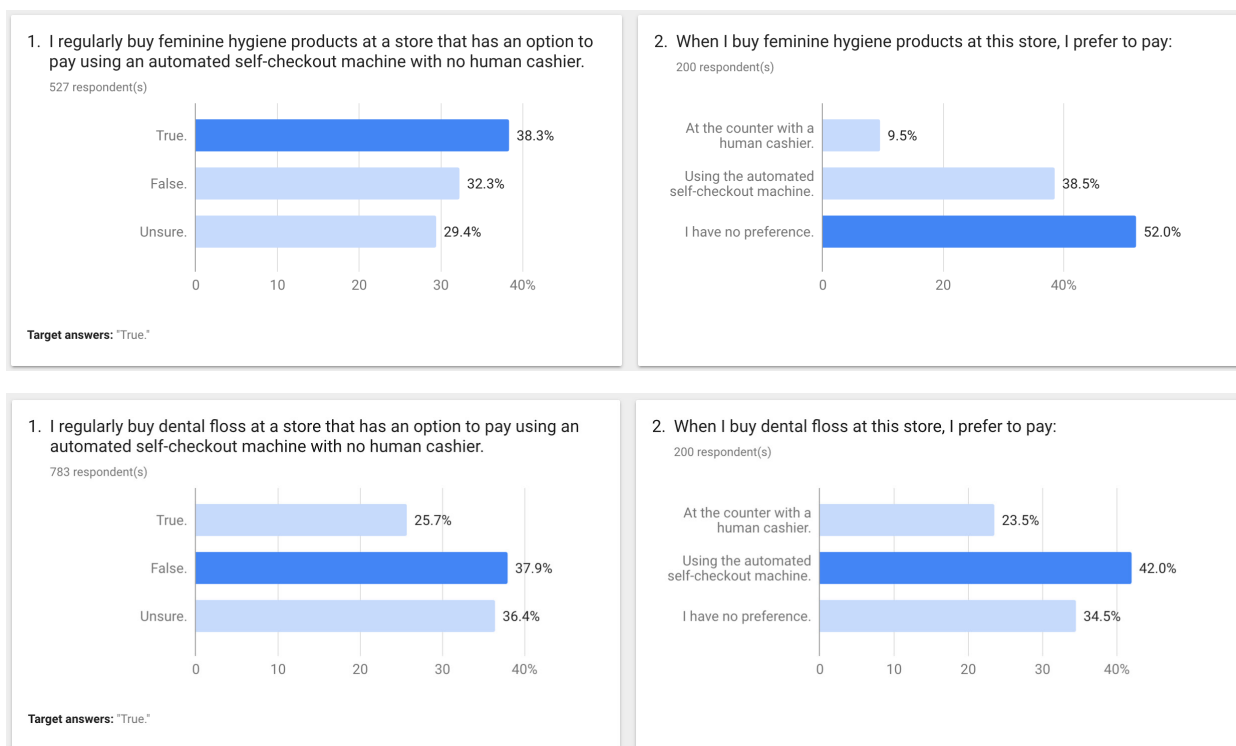




For dental floss purchases, on the other hand, 28.5 percent of respondents said that they would rather check out with a human, and only 40 percent preferred to use self-checkout. A substantial number reported having no preference for checkout method—33.5 percent when buying birth control products and 31.5 percent when buying dental floss.

Perhaps surprisingly, the difference was greater among young women than among young men, at least in terms of the aversion to human contact in buying sensitive products, though women did not prefer self-checkout machines as an alternative for these transactions any more than they did with dental floss.

**Figure 9: Results for Questions #6 & #7**



When buying dental floss, young women reported similar preferences to young men; 23.5 percent preferred a human cashier, 42 percent preferred to use self-checkout, and 34.5 percent had no preference. For tampon or sanitary napkin purchases, on the other hand, only 9.5 percent preferred paying a human cashier, while 38.5 percent preferred using self-checkout machine, and 52 percent had no preference.

In short, few young women want to buy feminine hygiene products from a human, and young men mind buying birth control products from a human somewhat more than they mind facing a person to buy floss for their teeth. The magnitude of these differences is relatively small, however, with a majority of young women reporting no preference when buying tampons or pads, and only a slightly higher percentage of young men preferring self-checkout for condoms as compared to dental floss.

A possible explanation for relatively small size of the effect is the age group surveyed. Young teens may be highly self-conscious when buying sensitive products, but a 24-year-old has potentially been buying tampons for at least a decade, and condoms for perhaps half that, and so the initial cheek-flushing embarrassment may have subsided.<sup>58</sup> Because Google Surveys prohibits targeting questions to individuals under 18, further research on this question will likely not be possible. Research using another platform, however, may be valuable to reveal whether any strong preference for self-checkout exists among minors buying condoms or tampons.

## CONCLUSION

While the strength of the effect varies by question and in one case is slight, all of our survey pairs showed at least some conformity with the “Privacy Paradox” thesis—although with respect to women and self-checkout, that conformity was limited to increased reticence about paying for feminine hygiene products with a human cashier. In some instances—with respect to preference for electronic books and the preference for online purchases, in particular—the effect is quite striking. Cumulatively, the results show that in their behavior, individuals, at least some of the time, prioritize privacy from those around them over privacy from the remote companies that collect data on us and, indeed, a preference for facilitating highly local privacy at the expense of privacy from remote data-collectors.

These results thus support a conclusion that privacy is, in fact, a highly contextual value—that it is not enough to talk about protecting individual privacy interests, but that the question of privacy *from whom* pervasively deserves more attention than it tends to garner in today’s conversations. Sometimes, the answer to that question will be privacy from Facebook, Google, and Amazon, but often, in the real lives of real people, the answer will not be; it will be something far more prosaic. And big, data-collecting companies can offer, at least sometimes, part of the answer to many such privacy problems.

These results also suggest promising areas of future research. For one thing, they suggest that Google Surveys is a valuable tool in making privacy research more empirical and less theoretical. Specifically, we have sought to demonstrate its value as a behavioral proxy to measure not what people think or feel but what they actually do—or, at least, what they report to do. The results suggest something of a mismatch between the privacy expectations revealed in consumer actions and those theorized by privacy scholars and activists. While the somewhat stuffy rules of Google Surveys with respect to sex and medical matters complicate its use as a tool to study privacy, we believe that we have only scratched the surface of the use of this tool as a means of exploring the empirical contours of that mismatch and, more generally, of the reality of privacy as a lived experience of billions of people. In particular, we think the following questions bear close examination:

- What are the circumstances in which people choose privacy from those around them over privacy from large remote entities? What can we learn about the specific subjects and behaviors that trigger the preference for giving information to the remote entity over the immediately present person? Conversely, what are the circumstances in which people prefer to take local privacy risks rather than take the risks associ-

ated with giving data to a large, remote entity?

- To what extent are people making this choice consciously and to what extent are they doing so naively? That is, are people actively choosing to engage these companies in order to protect their privacy as they perceive it or do they perceive their choices as matters of convenience and comfort? Or do they not feel like they have a choice?
- Are there demographic differences in the way people make these choices? Specifically, are younger people or people from marginalized communities more likely to make the judgment that Amazon and Google pose less of a threat to them than their friends, relatives, and neighbors? Or, conversely, are such subgroups more likely to fear the remote judgment of faceless entities? And are people in authoritarian countries more or less likely to entrust data to remote companies?

Finally, it's worth stressing that Google Surveys is not the ideal data source to address the questions at the heart of our inquiry. It is a proxy for data that actually exist: data on how and when people use self-checkout machines, on what books they do and do not read in electronic form, on what products they buy online and in person. One of the reasons that privacy remains such a theoretical field is that companies guard this data very jealously—often, ironically, in the name of privacy. Our understanding of privacy as a lived value would be immeasurably aided by both a scholarly focus on the lived experience of privacy by actual consumers and a willingness on the part of companies to facilitate serious research by making aggregate data available to researchers.

## ENDNOTES

1 Benjamin Wittes & Jodie Liu, *The privacy paradox: The privacy benefits of privacy threats*, Brookings Inst. (May 2015).

2 See, e.g., Jonathon W. Penney, *Privacy and the New Virtualism*, 10 Yale J.L. & Tech. 194, 232 (2008) (noting that “we want some information about us to be offered to sites (i.e. bits of data in cookies) to make our online travels more convenient”).

3 The authors would like to thank Google for donating a significant amount of Google Surveys time for this project.

4 Wittes & Liu, *supra* note 1, at 2.

5 *Id.* at 10.

6 *Id.* at 11–15.

7 *Id.* at 15–18.

8 *Id.* at 16–17.

9 *Id.* at 16.

10 *Id.* at 17–18.

11 *Id.* at 18–20.

12 Pornhub, a large online video-hosting platform for pornography, tracks and analyzes aggregate data about porn viewership by location. *Gay USA Porn Equality*, Pornhub, <http://www.pornhub.com/insights/gay-usa-porn-equality> (last visited Dec. 31, 2016).

13 David Pozen, *Privacy-Privacy Tradeoffs*, 83 U. Chi. L. Rev. 221, 229 (2016).

14 *Id.*

15 *Id.* at 222.

16 *Id.* at 246.

17 Rebecca Lipman, *Online Privacy and the Invisible Market for Our Data*, 120 Penn St. L. Rev. 777, 785 (2016).

18 *Id.*

19 *Id.*

20 *Id.*

21 John Herrman, *Privacy: The Sequel*, The Awl (Sept. 23, 2015), <https://theawl.com/privacy-the-sequel-236b74d1d12b#.7rr0dqmp>.

22 See *id.*

23 *Id.*

24 Tim Cushing, *Paper Says Public Doesn’t Know How To Keep Score In Privacy Discussion While Glossing Over Government Surveillance*, TechDirt (May 22, 2015), <https://www.techdirt.com/articles/20150521/14224431079/paper-says-public-doesnt-know-how-to-keep-score-privacy-discussion-while-glossing-over-government-surveillance.shtml>.

25 *Id.*

26 Discussion Post, *Hacker News*, <https://news.ycombinator.com/item?id=9795784> (last visited Jan. 1, 2016).

27 *Id.*

28 See, e.g., Emily Grumbling, *Privacy Research and Best Practices: Summary of a Workshop for the Intelligence Community* 23 (2016) (“[P]eople generally know that their online activities are being tracked—this is apparent from a range of surveys, as well as from anecdotal reports.”)



29 See Mary Madden & Lee Rainie, *Americans Attitudes About Privacy, Security and Surveillance*, Pew Research Ctr. (May 20, 2015) (describing how the “public has been awash with news stories detailing security breaches at major retailers, health insurance companies and financial institutions,” which contributes to public awareness.)

30 See, e.g., Leslie K. John, *We Say We Want Privacy Online, But Our Actions Say Otherwise*, Harv. Bus. Rev. (Oct. 16, 2015), <https://hbr.org/2015/10/we-say-we-want-privacy-online-but-our-actions-say-otherwise> (“In polls and surveys, consumers indicate profound and increasing concern for their privacy. Yet from the posting of suggestive photographs on social networking sites to the impulsive broadcasting of illicit activities on Twitter, consumers’ behavior often suggests a remarkable lack of discretion.”)

31 Sarah Spiekermann et al., *E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior*, 48 Comm. Ass’n for Computer Machinery 38, 38 (2005).

32 *Id.* at 45.

33 Patricia Norberg et al., *The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors*, 41 J. of Consumer Aff. 100, 118 (2007).

34 *Surveys: How It Works*, Google Analytics Solutions, <https://www.google.com/analytics/surveys/how-it-works/> (last visited Jan. 2, 2017).

35 Nate Silver, *Which Polls Fared Best (and Worst) in the 2012 Presidential Race*, N.Y. Times FiveThirtyEight Blog (Nov. 10, 2012), [http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/?\\_r=1](http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/?_r=1).

36 Nate Silver, *FiveThirtyEight’s Pollster Ratings*, FiveThirtyEight, <https://projects.fivethirtyeight.com/pollster-ratings/> (last visited Jan. 2 2017).

37 Google Consumer Surveys has even found its way into legal scholarship. In a student note published in the *Yale Law Journal* (Conor Clarke & Edward G. Fox, *Perceptions of Taxing and Spending: A Survey Experiment*, 124 Yale L.J. 1252 (2015)), the authors used Google Surveys to test the hypothesis that policies described as tax breaks will be more popular than policies identical in substance and total cost that are described as direct outlays. *Id.* at 1252. Though the authors acknowledge that their panel of respondents to Google Surveys is unlikely to be perfectly representative of the U.S. population, they determine that it is likely to be close. *Id.* at 1268. They further note that in the three years between its launch and the publication of the Note, Google Surveys was used “to produce peer-reviewed papers in a variety of fields, including political science, psychology, and business.” *Id.* at 1267.

38 Pew Research Ctr., *A Comparison of Results from Surveys by the Pew Research Center and Google Consumer Surveys 2* (Nov. 7, 2012).

39 *Id.*

40 Andrew Perrin & Maeve Duggan, *Americans’ Internet Access: 2000-2015*, Pew Research Ctr. (June 26, 2015), <http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/>.

41 Pew Research Ctr., *A Comparison of Results from Surveys by the Pew Research Center and Google Consumer Surveys 2* (Nov. 7, 2012).

42 *Id.* at 6.

43 See *id.*

44 Paul McDonald, et al., *Comparing Google Consumer Surveys to Existing Probability and Non-Probability Based Internet Surveys*, Google (2012), [https://www.google.com/insights/consumersurveys/static/consumer\\_surveys\\_whitepaper.pdf](https://www.google.com/insights/consumersurveys/static/consumer_surveys_whitepaper.pdf).

45 *Id.* at 5.

46 *Id.* at 6–9.

47 *Id.* at 2.

48 *Id.* at 10.

49 Erin Tanenbaum et al., *How Representative are Google Consumer Surveys?* Am. Ass’n for Public Opinion Research (2013).

50 Lie Philip Santoso et al., *Survey Experiments with Google Consumer Surveys: Promise and Pitfalls for Academic Research in Social Science* 24 Pol. Analysis 356, 372 (2016).

51 *Id.* at 371–372.

52 Email to Emma Kohse from the Google Surveys Team (Sept. 9, 2016).

53 *Id.*

54 See Mark Sweney, *Fifty Shades of Grey Publisher Random House Posts Record Profits*, The Guardian (March 26, 2013), <http://www.theguardian.com/media/2013/mar/26/fifty-shades-random-house-record-profit>.

55 See Wittes & Liu, *supra* note 1, at 17–18.

56 Kathryn Zickuhr & Lee Rainie, *A Snapshot of Reading in America in 2013*, Pew Research Ctr. (Jan 16, 2014), <http://www.pewinternet.org/2014/01/16/a-snapshot-of-reading-in-america-in-2013/>.

57 For a discussion of the idea that convenience trumps privacy concerns when it comes to technology, see, e.g., Felix Salmon, *Privacy is an afterthought when convenience is king*, Fusion (Jan. 1, 2016).

58 This idea is supported by data, reported in the original “Privacy Paradox” paper, describing the surfeit of articles, blog posts, memes, and discussion posts describing consumers’ relief at having the ability to use self-checkout to buy condoms, or horror when the option is unavailable. Wittes & Liu, *supra* note 1, at 16. Google Autocomplete similarly links self checkout with condoms. *Id.* Evidently, some subset of consumers, perhaps young teens, find the benefits of self-checkout for condom purchases substantial.

## GOVERNANCE STUDIES

The Brookings Institution  
1775 Massachusetts Ave., NW  
Washington, DC 20036  
Tel: 202.797.6090  
Fax: 202.797.6144  
[brookings.edu/governance](https://brookings.edu/governance).

## EDITING

Liz Sablich

## PRODUCTION & LAYOUT

Cathy Howell

## EMAIL YOUR COMMENTS TO [GSCOMMENTS@BROOKINGS.EDU](mailto:GSCOMMENTS@BROOKINGS.EDU)

The Brookings Institution is a nonprofit organization devoted to independent research and policy solutions. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views of the Institution, its management, or its other scholars. Emma Kohse has no relevant or material financial interests that relate to the research described in this paper.

Google generously donated the Google Survey questions used in this piece. The findings, interpretations, and conclusions in this piece are solely those of the authors and not influenced by any donation.

Brookings recognizes that the value it provides is in its absolute commitment to quality, independence, and impact. Activities supported by its donors reflect this commitment.