Is Privacy Policy Language Irrelevant to Consumers?

Lior Jacob Strahilevitz[1] & Matthew B. Kugler[2]

Abstract

Consumers almost never read privacy policies. But what if consumers *did* read such policy language closely? Would the language of the privacy policy matter to them? This paper reports the result of an experiment in which a census-weighted sample of more than a thousand American email and social network users were asked to read short excerpts from Facebook and Google's privacy policies concerning the use of facial recognition software and automated content analysis on emails, respectively. Both policies are currently at issue in pending high-stakes class action lawsuits, where the defendants have argued that by agreeing to privacy policy language consumers have consented to the conduct at issue. Experimental subjects were randomly assigned to read either language from current policies, which lawyers would regard as explicitly describing the controversial Facebook and Google practices, or language from earlier policies, which lawyers and judges have deemed insufficient to notify consumers about the companies' practices. Strikingly, despite evidence that many experimental subjects read the privacy policies closely, subjects who saw the explicit policy language and those who saw the ambiguous / vague policy language did not differ in their assessment of whether their assent to that language would allow Facebook and Google to engage in the practices at issue. More surprisingly still, even though consumers rated both Facebook's use of facial recognition software and Google's use of automated content analysis as highly intrusive, they regarded their assent to even vague privacy policy language as allowing the companies to engage in those practices. Finally, respondents were asked whether they would be willing to pay some amount of money to avoid automated content analysis of their emails. A little more than a third of the experimental subjects expressed willingness to pay any money to avoid such privacy-invasive practices. Even among this subpopulation, the median willingness to pay for a more privacy-protective email product was only $15 per year. These results provide important evidence for the propositions that (1) social norms and user experiences with technological applications, not privacy policy language, will drive users' understanding of the nature of their bargain with firms, that (2) most users of email and social networking sites believe that Facebook and Google are authorized to engage in controversial and invasive practices implicating user privacy, and that (3) there is presently little reason to expect the development of a robust market for premium privacy-protective email and social networking applications in the United States.

---

[1] Lior Jacob Strahilevitz is the Sidley Austin Professor of Law at the University of Chicago.

[2] Matthew Kugler is a law clerk to the Honorable Richard Posner, U.S. Court of Appeals, Seventh Circuit. The authors owe thanks to [commentators], the editors, and participants in the Chicago conference on Contracting over Privacy, for helpful comments. Adam Woffinden provided terrific research assistance. The Russell J. Parsons Faculty Research Fund, Bernard Sang Faculty Research Fund and Coase-Sandor Institute for Law & Economics provided generous research support.

1. Introduction

There are few threats that can bring down companies as large and powerful as Google and Facebook. Privacy class actions are one such threat. In 2011, Google was sued in California for violating the federal Wiretap Act, which prohibits the intentional interception of emails. (In re Google, Inc. Gmail Litigation, 2013). The plaintiffs alleged that by scanning the contents of Gmail users emails to serve them with personalized advertisements, Google had been violating the law, and their claim was a plausible one. Google answered the complaint by arguing that Gmail users had consented to automated analysis of their emails, a defense that the district court ultimately rejected after parsing the language of Google's Terms of Service and Privacy Policies.

As a result of the district court's decision, Google faced extraordinary legal exposure. Under the Wiretap Act, plaintiffs who can show a violation are entitled to a minimum of $100 per day of the violation. (18 U.S.C. § 2520(2)(B)). Nearly all Gmail users send or receive some email every day, and Gmail has approximately 500 million users worldwide. If Gmail had an average of 50 million American users during the five-year period of alleged violations, then its exposure under the lawsuit would be nearly $9 trillion. The company wound up defeating the plaintiff's class certification motion in 2015 (In re Google II), which limited that exposure, so that the company's survival is no longer at stake. But the stakes in the privacy litigation remain significant.

Nearly contemporaneously with the Google class certification decision, Facebook also found itself in hot water as a result of another privacy class action. This time plaintiffs sued Facebook in Illinois for violating a state law that required any entity using biometric identification techniques to obtain the explicit consent of those subject to biometric identification before proceeding. (Illinois Biometric Information Privacy Act, 740 ILCS 14 et seq.) The law also requires such entities to disclose the purpose of their use of biometric identification and the amount of time for which biometric data would be retained. (Licata case, now pending in N.D. Cal after removal and transfer). The plaintiffs pointed out, again quite plausibly, that Facebook did not obtain the explicit consent of Facebook's Illinois customers before using facial recognition software to identify them in photos and suggest to their Facebook friends that they be "tagged" (i.e., identified in a caption) in photos their friends had posted. Here too, Facebook would try to defend itself by arguing that its users in Illinois had consented to the use of facial recognition software, requiring the courts to turn to Facebook's privacy policies. The litigation is pending, but the stakes are again quite high. The statute contemplates minimum statutory damages of up to $5000 per violation, and a very similar statute exists in Texas, which has an even larger population than Illinois. (ILCS 14/20(2); Wellinder article from Harv. L & Tech). A back-of-the-envelope calculation reveals that even if no Illinois Facebook user could sue for multiple violations of the law, Facebook's potential exposure is still approximately $37.5 billion.[3]

In ruling on Google's motion to dismiss the Wiretap Act suit, the district court assumed that Gmail users read the privacy policies in question and then found that agreeing explicitly with the terms of those policies would not have amounted to consent to the automated email content analysis as a matter of law. As the district court knew, and as scholars have long argued, consumers do not typically read privacy policies and other online disclosures, even for products like Gmail that they use every day.

[3] Approximately 58% of Americans had Facebook accounts as of 2015, and Illinois had about 12.9 million residents at that time. Assuming Illinois residents use Facebook at national average rates, that means there are about 7.5 million Facebook users in the state. Multiplying that figure by $5000 yields $37.5 billion. But if each instance of unauthorized tagging is a separate violation, then Facebook's potential liability could quickly escalate from there. The statutory text seems ambiguous on the question. (740 ILCS 14/20 et seq.)

(McDonald & Cranor 2008, Marotta-Wurgler 2011; Schneider & Ben-Shahar, Ayres & Schwartz, Porat & Strahilevitz.) But the "duty to read" is nevertheless well established in contract doctrine. (citations). There may be something beneficial about incentivizing consumers to read at least high-stakes consumer contracts, and holding consumers to the terms of those contracts may help encourage consumers to educate themselves and enter into contracts on the basis of fuller information. There may also be formalist explanations for a duty to read as well as justifications grounded in suppositions about what boilerplate terms will be produced in a competitive marketplace. (Wilkinson-Ryan, Iowa; other scholarship.) And this duty to read can have external implications as well. For example, the government sometimes argues that consumers have no Fourth Amendment expectation of privacy in information consumers shared with online service providers under the terms of the providers' privacy policies. (United States v. Graham, 2015).

The law therefore relies on a legal fiction. Courts know that consumers do not read privacy policies but pretend otherwise for the purposes of contract law. Suppose that consumers actually read consumer contracts and privacy policies. What would they understand from them? That is the basic question this article poses and addresses through an experimental approach. The results of the research are surprising. After reading actual policy language from Gmail and Facebook, American users of email and social networking websites largely believe that by using those products they have consented to automated content analysis of their emails and the use of facial recognition biometrics to suggest photograph tags. That is true regardless of whether consumers read Google and Facebook's current privacy policies, which are explicit efforts to obtain the consent necessary to satisfy the law's requirements, or whether consumers read those companies' older privacy policies, which (at least in the Gmail litigation) the court deemed inadequate to obtain users' consent. In short, American consumers who read privacy policies believe they have consented to practices to which courts believe they have not consented.

Interestingly, it does not appear that Americans' views that they have consented to automated content analysis of their emails or the use of facial recognition to identify them in photos posted by others stems from normative approval of Google and Facebook's respective practices. When asked about the intrusiveness of Google and Facebook's practices, respondents rate these practices as highly intrusive. In light of these reactions, the most plausible interpretation of the data presented here is that email and social networking users understand these practices are part of the bundle associated with Gmail and Facebook, believe themselves to have accepted that bundle, all the while preferring that the bundle included greater privacy protections. (Wilkinson-Ryan, 2014). In short, even when consumers read privacy policies, their beliefs about the nature of their bargains with technology companies seems to depend more on their pre-existing expectations than on either the terms of the policies or their normative preferences.

2. Literature

There is a slowly growing experimental literature on consumer contracts. Some of it employs random assignment techniques to determine what effects changes in contract language or structure have on consumer behavior. For example, Zev Eigen randomly assigned online survey participants to conditions that mimic standard contract boilerplate, a compelled choice between two terms, and notice plus choice. (Eigen 2012) He found that respondents assigned to the boilerplate condition were less likely to read contractual terms and also devoted less energy to performing the task the experiment asked them to do. Joshua Mitts randomly assigned a mix of real and fictitious contract terms to respondents and identified surprising / unexpected terms. Such terms were then highlighted with

warnings for consumers. (Mitts 2014). Mitts found that the more times warnings about unexpected terms were given to consumers, the less effective each warning was in helping consumers understand the terms of the agreement.

[To be completed: Discuss the literature on psychology of contracting; more experiments on contracting. Also explore the privacy paradox literature and the question about why consumers will not have options to pay for less privacy invasive versions of popular products.]

3. Data and Empirical Approach

3.1 The Sample

Toluna, a professional survey research firm with an established panel, administered a survey to a weighted sample of 1,441 adult US citizens between May 26, 2015, and June 2, 2015. Data from some of these was discarded because of abnormally fast survey completion times and failed attention checks, leaving a final sample of 1382. The median age of respondents was 47 (range 18-89, mean: 46.62, SD = 16.37). Females comprised 49.8% of the sample. Compared with the population in the US census, a slightly higher percentage of the panel had completed high school or at least some college coursework, but the educational attainment of the sample was otherwise similar to that of the adult census population. 79.9% of the sample self-identified as White, 13.0% as Black, and 4.1% as South or East Asian. On a separate question, 16.2% of the sample reported that they are Latino or Hispanic. Respondents were asked their political orientation on a scale of 1 (very liberal) to 7 (very conservative), with a mean response of 4.16 (SD = 1.78), indicating an ideologically moderate sample. The Gmail and Facebook questions were administered at the end of a 10-15 minute survey that included questions for other papers on topics such as Fourth Amendment privacy expectations and trademark questions designed to assess attributions of product sponsorship.[4]

Participants were screened on the basis of whether they have email accounts for the Gmail questions and whether they have Facebook accounts for the Facebook questions. That left 1377 potential respondents to the email questions and 1,052 potential respondents for the Facebook questions.[5] Approximately 76.1% of the sample were therefore Facebook users. This utilization rate is close to the one produced by a Pew Research study conducted a few months earlier, which found that 72% of American adult Internet users used Facebook.[6] In each instances, eligible respondents were randomly assigned one of three "privacy policies" for both the Gmail and Facebook questions. In each

[4] These survey results are discussed in Strahilevitz & Kugler 2016, and Kugler 2016, respectively.

[5] Facebook users were, on average, slightly younger than non-Facebook users (Users M = 45.06, SD = 16.17; Non M = 51.58, SD = 16.09). The Facebook user population was also more female (52.4% v 42.8%) than the general sample. The racial breakdown was roughly equivalent, however (79.0% White, 13.6% Black, 4.0% South or East Asian). Note that 28 respondents indicated that they had Facebook accounts but did not answer any of the other Facebook-related questions, so they were dropped from this experiment.

[6] Duggan, 2015. The Pew study reports that 62% of all US adults are Facebook users. Although our Toluna sample is census-weighted, Americans without Internet access were necessarily excluded from the online survey. This exclusion does not seem problematic given our interest in learning how consumers of privacy policies and online apps understand those policies. The exclusion of those without Internet access (16% of the adult population) largely explains the disparity in education levels between our sample and the adult population. Perrin & Duggan, 2015.

instance the privacy policy language subjects read was taken from actual language that Google or Facebook employed at some point in time.[7] The policy language varied in terms of how explicit it was about Google and Facebook's data practices. Not surprisingly, the current policy language (posted after the lawsuits at issue here were filed) is more explicit about company practices than the pre-lawsuit language.

3.2 The Survey Instrument

The randomization strategy in the experiment allows for a clean test about what effect differing policy language has on consumers' views of what they have agreed to. The difference in the new language and old language was (to these lawyers' eyes, at least) dramatic enough to warrant the following pre-experiment hypothesis: Lay understandings of privacy policies would depend significantly on the policy language chosen. Given the prominent display of just the relevant language to respondents, enough consumers would read the privacy policies closely to render the substantial differences between the old and new privacy policies significant.

Respondents were also asked other questions targeted at issues beyond the question of whether they had consented to the legally relevant conduct by Google and Facebook. Respondents to the Gmail survey were asked "On a scale of 1 to 10, how intrusive is the email provider's automated email scanning and ad personalization practice?" After answering this question, they were asked: "If there were an option to keep the same email account but pay some amount of money to avoid having the automated systems analyze email content for the purposes of showing you personalized advertisements, how would you respond? (1) I would keep the free email account with the automated email analysis and personalized advertisements. (2) I would be willing to pay some amount of money to avoid the automated analysis." Respondents who selected option 2 were asked how much money they would be willing to pay per year for a more privacy protective email product.

Respondents to the Facebook question, all of whom had Facebook accounts (N=1052) were shown various Facebook privacy policy language and then asked three questions designed to elicit responses that would shed light on whether Facebook had complied with its obligations under Illinois law. Namely, respondents who said they have Facebook accounts were asked: "Did Facebook's language (above) inform you that information about your facial features was being collected and stored?," "Did Facebook's language (above) inform you of the reason why information about your facial features was being collected, stored, and used?" "Did Facebook's language (above) inform you of the length of time for which information about your facial features would be stored?" (740 ILCS 14 et seq.) And then finally, they were asked the consent question: "Would your decision not to adjust your Timeline and Tagging settings allow Facebook to collect, store, and use information about your facial features?"

Respondents to the Facebook questions were then asked to rate on a scale of 1-10 the intrusiveness of Facebook's use of facial recognition software to suggest tags for people whose faces appear in uploaded photos.

[7] The Gmail questions used both Google's current language from 2015 and the circa 2011 language quoted in the Gmail litigation. The Facebook questions used Facebook's current language and earlier versions of related privacy policies obtained via the Internet Archive Wayback Machine. https://archive.org/web/

4. Results

Given the substantial differences between the language that email and social networking site users were shown, we predicted that our respondents who saw the highly explicit disclosures from Google and Facebook would be more likely to say that their decision to leave their privacy preferences unchanged after reading the relevant privacy policies allowed Google and Facebook to engage in the content analysis and facial recognition practices at issue. Surprisingly, the data did not bear out that prediction. Regardless of what language respondents were shown, they had statistically indistinguishable views about what practices their inertia would have authorized.

4.1 Experiment 1: Gmail Results

In the Gmail experiment, random assignment to one of three conditions – Google's very explicit current privacy policy, Google's moderately explicit historic section 17 language, or its least explicit historic section 8 language – had no significant effect on consumers' judgment about what they had authorized Google to do to their emails. Nor did the privacy policy language have any significant effect on the perceived intrusiveness of Google's automated content analysis of their customers' emails.

Table 1: Responses to Gmail Consent Question by Privacy Policy -- "Would your agreement to this provision allow the email provider to direct its automated systems to scan the contents of the emails you send and receive and show you personalized advertisements?"

	Current Language (most explicit)	Section 17 (moderately explicit)	Section 8 (least explicit)	Overall (ignoring condition)
Definitely Allowed	28.1%	28.1%	23.5%	26.6%
Probably Allowed	35.7%	40.2%	39.6%	38.5%
Probably Not Allowed	13.9%	10.6%	15.6%	13.4%
Definitely Not Allowed	22.2%	21.2%	21.3%	21.3%

The differences across condition are not significant. χ^2 (2, N = 1363) = 8.38, p =.21.

Even the coefficients are surprising. When confronted with the more explicit language that (to us lawyers) plainly authorizes email content analysis, respondents were actually *more likely* to say that Gmail's content was probably not allowed or definitely not allowed, though not by any significant margin. Differences in language that lawyers and judges would deem critical made no evident difference to a representative sample of adult American email users. (Compare Reidenberg et al. 2014).

Moreover, in every condition, most respondents say that if they read the short privacy language at issue and then did not change their privacy settings to prohibit content analysis, Google would be authorized to engage in the automated content analysis. Roughly two-thirds of the sample expressed this view in all three conditions, compared to roughly a third of the sample who stated that Google probably would not be allowed to engage in content analysis under all three conditions.

One possible interpretation of this result is that email users like receiving personalized advertisements and do not mind the automated content analysis of their email that facilitates this personalization. On this interpretation of the data in Table 1, consumers' normative views would be driving their answers to the question of what Google can do. Perhaps email users think that Gmail's practices are reasonable, and therefore authorized implicitly whenever a user signs up for a nominally "free" email account. This dynamic would explain why the privacy policy language at issue makes no difference. Table 2 shows why readers should be skeptical of that interpretation.

Table 2: Responses to "On a scale of 1 to 10, how intrusive is the email provider's automated email scanning and ad personalization practice?"

Email Condition	Mean	Std. Dev.	N
Most Explicit	7.60	2.47	445
Moderately Explicit	7.62	2.34	463
Least Explicit	7.65	2.43	455
Total	7.63	2.41	1363

Dep. Variable – intrusiveness: $F(2, 1360) = .06$, $p = .95$ $\eta^2 = .000$.

Mean Intrusiveness responses for Google's conduct is 7.63 on a 10-point scale. Consumers are saying that they regard the automated content analysis as quite creepy, but nevertheless authorized, even when presented with language that few lawyers would regard as consenting to the practice at issue. Intrusiveness ratings, predictably, were not significantly affected by whether respondents saw more explicit or less explicit privacy policies. Those who believe Google is not authorized to scan emails view the practice as slightly more intrusive ($r(1363) = .192$, $p < .001$), but the effect size is very small.

Prior experimental research on privacy has identified a "privacy paradox." Privacy paradoxes arise because Americans often say they care a great deal about privacy and yet they are willing to permit third parties to obtain sensitive information about them in exchange for relatively inexpensive goods and services, or in exchange for longshot odds to win a prize in a random drawing (Acquisti 2010, Holland, 2010, Swire, 1999). Our results are consistent with the privacy paradox. Although the mean respondent rated automated content analysis of emails as a 7.63 out of 10 on an intrusiveness scale, just 35.4% of the sample expressed a willingness to pay any amount of money to receive a version of their email service that did not use automated email content analysis to deliver personalized ads.[8] Among the roughly one-third of the sample that was willing to pay some amount of money, the median willingness to pay was $15 per year. Just 3% of the sample expressed a willingness to pay more than $120 per year for such an email service.

Perhaps this data indicates that the intrusiveness ratings offered by our respondents are not to be taken seriously. Maybe the 7.63 intrusiveness figure is just cheap talk. On this reading of the data automated email content analysis is not a serious concern for most Americans, which explains why they feel that Google is allowed to engage in the practice even without explicit ex ante warnings. Another

[8] Interestingly, the amount respondents were willing to pay correlated significantly with neither responses to the authorization question nor the intrusiveness question, though there is a relationship between being willing to pay any amount and the other measures. Those willing to pay some amount rated the intrusiveness at 8.65 (1.83) while those not willing to pay anything rated it at 7.06 (2.50).

possibility is that users of the Internet have grown accustomed to "free" email, news, weather, media content, etc., such that putting email behind a paywall prompts significant resistance even when doing so would create a substantially more privacy-protective product. (Citations here on the free internet). Alternatively, perhaps consumers say they are reluctant to pay any dollar amount for a privacy-protective email account precisely because they know that other email services (Yahoo!, Hotmail, etc.) exist and suppose they do not engage in automated content analysis. So Gmail users who object to Google's practices would prefer to stay with their present account (because of high switching costs) but say that were they to switch, they would switch to Yahoo!'s free service instead of a paid Gmail product. In any event, the shortage of consumers willing to pay meaningful sums for more privacy-protective email services suggests there may be a limited market for premium products that protect user privacy. Recent estimates suggest that a year's worth of data is worth $50 to $5000 per consumer to Google and $45 to $190 per consumer to Facebook. (Howe, 2015). The sorts of fees they'd be able to obtain from users for greater privacy-protection are relatively small potatoes, though it is conceivable that enhanced data security would prompt a more robust response from consumers. In any event, the shortage of consumers willing to pay meaningful amounts for more privacy protective email accounts suggests a limited market for services that protect data privacy.

Our data provide information that permits some inferences to be drawn about the dynamics at play. It does not appear that differential views about the intrusiveness of automated email content analysis are driving users to one email provider or another. Mean intrusiveness ratings were not significantly different among Gmail, Yahoo!, AOL, and Hotmail users. $F(3, 1136) = 1.44$, $p = .23$ $\eta^2 = .004$. Nor do consumers appear to be choosing their email providers based on their privacy preferences and company policies more broadly. When we analyzed responses to questions about the intrusiveness of Facebook's facial recognition software (discussed below) based on what email providers respondents use, there were no significant differences.

It is less clear if awareness of different company practices affect respondents' assessments of whether automated content analysis is permitted. Gmail users were significantly more likely than AOL and Yahoo! users to believe that email content analysis was permitted, but so were Hotmail users, and the effect sizes were small in any event.[9] (AAOL, Hotmail, and Yahoo! evidently do not use automated content analysis on their customers' emails.) Given that respondents were asked about whether their own email providers were allowed to engage in automated content analysis, it seems that at most a small sample of the population is attentive to the differences between Google's content-analysis and Yahoo!'s lack thereof.

Our study also generated mixed evidence on the question of willingness to pay. On the one hand, respondents willing to pay some amount of money to avoid content analysis rated the intrusiveness of the content analysis at 8.65 (1.83), whereas those unwilling to pay any amount rated it at 7.06. On the other hand, the amount people were willing to pay (above zero) bore no relationship to either the perceived intrusiveness or the authorization of automated content analysis.

4.2 Experiment 2: Facebook Results

Under the Illinois Biometric Information Privacy Act, two questions potentially involving consumer psychology are legally relevant. First, did Facebook comply with its obligations under the

[9] Gmail 5.23 (1.07); Yahoo! 5.39 (1.09); AOL 5.45 (1.06); Hotmail 5.19 (1.03); Total 7.61 (1.07); $F(3, 1136) = 2.93$, $p = .033$ $\eta^2 = .008$.

statute to inform its users about (a) the fact that information about their facial features was being collected and stored, (b) the reason why information about their facial features was being collected, and (c) the length of time for which information about their facial features would be stored. In addition, the law renders relevant the question of (d) whether Facebook had its users' permission to collect and store information about their facial features. Each of these four questions depends on Facebook users' understanding of Facebook's terms of service. This experiment was designed to test whether, if Facebook users had read the relevant information, they would feel that Facebook had adequately informed them of its practices and obtained their authorization to engage in them.

There were a clear consensus among respondents on all four questions, and the consensus is particularly interesting on the third of the four questions.

Table 3: Responses to Facebook Questions: Percentage of Respondents Answering "Yes"

	Policy – "We collect" (Least explicit on Facebook's actions and purposes)	Policy – "We use" (Least explicit on Facebook's actions, more disclosure on purposes)	Policy – "When Someone Uploads" (Most explicit on Facebook's actions, less on purposes)	Total
Did Facebook inform you about collection and storage?	67.8%	65.9%	70.7%	68.2%
Did Facebook inform you of reason for collection, use, and storage?	59.0%	61.8%	67.1%	62.5%
Did Facebook inform you about length of time information would be stored?	35.1%	34.0%	30.7%	33.3%
Does leaving settings unchanged allow Facebook to collect, use, and store information?	63.5%	62.3%	61.0%	62.3%

Despite a sample size of 1052 respondents, in none of the conditions presented in Table 3 is the language from Facebook's privacy policies having any significant effect on consumers' responses. More

than two-thirds of the sample regard themselves as having been informed of Facebook's collection and storage of their biometric information after having read any of Facebook's current or historic policy language. And an only slightly lower percentage of Facebook users view Facebook's language as informing them of the *purpose* of Facebook's use and collection. Again, the wording of the policy language at issue made no significant difference, even though some of the language was relatively explicit about the purpose's of Facebook's collection of information. Similarly high percentages of respondents said user inaction with respect to privacy settings authorized Facebook's facial recognition practices.

Viewed in context, the striking set of responses are those to question three, which asks about the length of time for which Facebook is retaining its information. In none of these conditions did the privacy policy language provided to respondents address the length of storage explicitly. About two-thirds of the respondents seem to have noticed this. This consistent reversal of the usual ratios across all three conditions suggests several possible implications. First, it seems that at the very least a third of the sample is reading the lengthy privacy policy language in the prompt carefully. These are the respondents who flip from a pro-Facebook stance on the other questions to an anti-Facebook stance on question 3.[10] Second, it is possible that whereas Facebook users have intuitions about the fact that a facial recognition algorithm is being used and the reason why it is being used (perhaps based on their use of the feature on Facebook), they lack a strong prior about the length of time for which facial recognition information should be retained, so the privacy policy language may loom larger. Third, unless there is other privacy policy language that Facebook can cite, it appears that Facebook's facial recognition feature is violating one provision – though on this analysis, only one provision – of the Illinois law.

Respondents were also asked about the intrusiveness of Facebook's practice of using facial recognition software to suggest tags for people whose faces appear in uploaded photos. Mean responses were a little lower than in the Gmail question (mean = 7.29 out of 10). Mean intrusiveness ratings did not differ by condition ($F(2, 1044) = 1.30$, $p = .27$ $\eta^2 = .002$). Thus, it does not appear that exposure to different policy language affected consumers' underlying beliefs about how problematic Facebook's practices are. Once again, majorities of consumers appear to regard Facebook's practice as troubling yet authorized. Comparing across conditions between authorization and perceived intrusiveness responses did not yield significant results. (p = .396).

5. Discussion

The key lessons from both experiments is that users of email and social networking sites appear to regard even highly ambiguous privacy policy language as authorizing controversial company practices that implicate their personal privacy. Wilkinson-Ryan finds a similar result in the context of other boilerplate consumer contracts. (Wilkinson-Ryan 2014). Google and Facebook have both raised consumer consent as defenses in the class action privacy suits broad against them. To the judge in the

[10] It is almost certainly the case that some of the respondents who conclude that Facebook is failing to abide by its obligations to disclose the length of data retention but abiding by its other obligations have read the policy language carefully. If we just focus on the approximately 1/3 of respondents who "flipped" on question three, treating "no" on question 3 and "yes" on, say, question 1 as the "correct" response, we lose a lot of statistical power, but the headline results in this paper on the Gmail question do not change. Using the "flipped" responses as a super-strict attention check is post hoc and questionable, however, so we do not report those results in the paper.

Gmail litigation, and to at least some lawyers who read the clauses, Google's and Facebook's privacy policies seem inadequate. Consumers regard them quite differently. The presence of differential response among different audiences tracks previous findings by Reidenberg and co-authors. (Reidenberg et al., 2014).

What explains the divergence between lawyerly judgments and lay consumers' judgments about what constitutes consent? One possible explanation is that by the time they answered the questions in our survey, consumers had formed strong priors about the sort of privacy-related conduct that companies are permitted to engage in, and these priors inform their understanding about what they agree to when they use Gmail or Facebook without changing their privacy settings. (Martin 2014). Expectations are driven by neither formal law nor written policy language, even when consumers are familiar with the law and the language. Social norms seem to drive expectations and under these norms consumers view themselves as having consented implicitly to practices that many of them find somewhat troubling. (Kerr 2016). When consumers interpret contracts, they bring in these priors and integrate their beliefs with the policy language to produce an understanding of the bargain to which they are agreeing. (Hoffman & Wilkinson-Ryan, 2012) Consumers may not like the bargain in all material respects – and their intrusiveness scores suggest discomfort with automated email content analysis and the automated use of facial recognition software – but they seem to believe the privacy sacrifices inherent in their use of email and social networking sites outweighs those costs. Lawyers, by contrast, are more focused on privacy policy language itself and trained to identify ambiguity in it.

When faced with data like this and a consent defense by a defendant who invokes this sort of empirical evidence, what should a court do? In our view, data such as this, collected using rigorous techniques of survey design and analyzed by academics with no skin in the game, ought to play a large role in litigation over privacy policies in particular and consumer contracts in general. The goal of companies designing privacy policies and consumer contract language should be to inform consumers about what the companies are doing and why they are doing with it. They should field-test their policies on consumers and avoid presuming that the only information consumers have is what is disclosed in the policy language. It's precisely because lawyers are trying to cram so much information into policies that policies become unduly lengthy, and the result it they go unread entirely by rational consumers. (Ben-Shahar & Schneider). The meaning of a consumer contract is a product of consumers' expectations and the contract language, with the former seemingly looming larger than the latter in some contexts. The product is readily measurable, even if teasing out what work the expectations are doing and what work the language is doing is more complex. At least in the instance of Gmail, privacy policy language chosen by Google and the other information that consumers are receiving from various sources does adequately inform most consumers about the nature of the bargain.

Several important caveats remain. First, we know that both (a) consumers very rarely read privacy policies and (b) courts adjudicating class action cases nearly always impose a duty to read on consumers. There may be sensible reasons for the courts to proceed on that basis, particularly at the motion-to-dismiss stage or the summary judgment stage. But if they do assume that consumers read these contracts, it seems highly problematic to assume an interpretation of those contracts that relatively few lay readers of those contracts would share. The duty to read can't possibly mean a "duty to hire a lawyer to read in a lawyerly way." Can it?

Second, in assessing the generalizability of these results, it is important to recall that our respondents were only asked to read a short excerpt of an arguably relevant provision in a much lengthier privacy policy and asked to read that only. Respondents were not charged with scanning a much lengthier policy and finding the relevant provision. Had we asked respondents to do that, few

would have been incentivized to comply. On the other hand, the Gmail and Facebook questions were presented to our respondents toward the very end of a 10-15 minute online survey that also asked them a number of questions about Fourth Amendment privacy questions and trademark issues, so they may have been roughly as impatient and fatigued as a hypothetical consumer who wanted to read the entirety of a privacy policy for one reason or another. In any event, the results here should be conceived of as relevant to the question of "what would happen if consumers actually read the pertinent parts of privacy policies?" an inquiry that, though hypothetical, winds up being outcome-determinative in a great many litigated cases.

Third, there is an adaptive preferences problem built into our survey methodology that could affect the interpretation of the results. The Facebook experiment was limited to respondents from a nationally representative sample who said they have Facebook accounts. The respondents therefore had already been exposed to Facebook's tagging suggestions, and many may have already realized that Facebook employed facial recognition software to suggest tags. This previous exposure had benefits and drawbacks. One benefit is that many consumers already understood a technology that might have been difficult to explain otherwise. (For reasons related to the complexity of the technology we did not ask non-Facebook users to answer the questions.) But a drawback is that by the time Facebook was sued and we presented them with our survey, Facebook had been employing facial recognition technology for nearly five years. (Ducklin 2010). Facebook users' initial understanding of Facebook's practices is arguably as relevant as Facebook users' contemporary understanding of Facebook's practices. The problem is present too in the Gmail survey, where the practice was again longstanding by the time the survey launched. That said, the lack of large differences in the responses of Gmail users and demographically similar Yahoo! users alleviates some concerns about conditioned responses. Still, as a result of these issues, our study lacks a clear "before" to go with its "after" result. Because it takes time to get a survey developed, approved, funded and launched, it is unlikely that third party researchers will ever be able to test consumers' understandings of companies' new practices before those practices have been implemented. But firms themselves might hire reputable academic researchers to obtain data that predates consumer adaptation to a new feature. That said, in both the *Licata* and *Gmail* litigation, plaintiffs are seeking continuing damages over a period of several years. Even if we cannot identify precise consumer sentiment at the time a controversial practice began, understanding contemporary responses may help place an upward bound on the damages that are appropriate in any given case.

Fourth and finally, there is a hard question of what to do with respondent heterogeneity. When presented with language that (to our lawyer eyes anyway) very clearly informs Facebook users of the reasons why Facebook is collecting facial recognition data, 37% of our respondents said that Facebook did not inform them of the reasons for the data collection. And when presented with language that (again, in our judgment) unambiguously informs readers that facial recognition of software is being used to collect data used for suggesting photo tags, 29% of our respondents said the language failed to do so. To come at the same issue from the other direction, 34% of the overall Facebook sample said that Facebook had informed its users about the length of time during which facial recognition data would be retained, a conclusion that cannot be squared with the information they were given, at least in this survey.[11] With any survey instrument, there are going to be some people who do not read very carefully or very well, and there will be others who have sufficiently strong views about the facts or morality of an issue to not be swayed by any exculpatory contract language. It appears that in our experiment, those groups combined to form somewhere between 25 and 40 percent of the overall sample. In a world

[11] It is theoretically possible that respondents saw such information previously in other contexts.

where lawyers have determined that contract or policy language should have some efficacy in shaping consumer expectations, the fact that 30% or 35% of a sample believes that particular policy language with which they were presented is inadequate should not sway a court unduly.

As one surveys pleadings in cases like *Gmail* and *Licata v. Facebook*, the absence of empirics about how consumers respond to terms of service language is striking. This is information that litigants (or better yet, social scientists) ought to be producing and courts ought to be evaluating. (Martin, 2014). The survey results presented here were neither particularly difficult nor costly to gather. Compared to a few billable hours of a good lawyer's time, such experimental research is a bargain.

6. Conclusion

[to be written]

References [incomplete]

Ducklin 2010 https://nakedsecurity.sophos.com/2010/12/17/facebook-friendships-get-creepier/

Maeve Duggan, *The Demographics of Social Media Users*, Pew Research Center: Internet, Science & Tech., Aug. 19, 2015, available at http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/ (reporting data from March and April of 2015).

David A. Hoffman & Tess Wilkinson-Ryan, Legal Promise and Psychological Contract, 47 Wake Forest L. Rev. 843 (2012).

Jared Howe, How Much is Your Personal Data Worth?, Private WiFi, June 9, 2015, available at http://blog.privatewifi.com/how-much-is-your-personal-data-worth/

Orin Kerr, Norms of Computer Trespass, 116 Colum. L. Rev. __ (forthcoming 2016), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2601707 .

Florencia Marotta-Wurgler, Some Realities of Online Contracting, 19 Sup. Ct. Econ. Rev. 11 (2011).

Kirsten Martin, Privacy Notices as Tabula Rasa: An Empirical Investigation into how Complying with a Privacy Notice is Related to Meeting Privacy Expectations Online, J. of Pub. Pol'y & Mktg. (forthcoming 2015).

Aleecia M. McDonald & Lorrie Faith Cranor, The Cost of Reading Privacy Policies, 4 I/S J.L. & Pol'y Info. Soc'y 543 (2008).

Andrew Perrin & Maeve Duggan, Americans' Internet Access 2000-2015, Pew Research Center, June 26, 2015, available at http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/

Joel R. Reidenberg et al., Disagreeable Privacy Policies: Mismatches Between Meaning and Users' Understanding, 2014 draft, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2418297 .

United States v. Graham, ___ F.3d ___ (4th Cir. 2015), available at 2015 WL 4637931.

Appendix – Experimental Questions

Gmail questions:

Random assignment among these three questions:

Q1139 Suppose that when you signed up with your current email provider you agreed that they could show advertisements next to your inbox in exchange for a free account. Suppose further than when signing up for the account, you read and agreed to the following terms and conditions: "Advertisements may be targeted to the content of information stored on the [email provider's] services, queries made through the provider's affiliated search engine, or other information"

Q1144 Suppose that when you signed up with your current email provider you agreed that they could show advertisements next to your inbox in exchange for a free account. Suppose further than when signing up for the account, you read and agreed to the following terms and conditions: "[Email provider's] automated systems analyze your content (including emails) to provide you personally relevant product features, such as customized search results, tailored advertising, and spam and malware detection."

Q564 Suppose that when you signed up with your current email provider you agreed that they could show advertisements next to your inbox in exchange for a free account. Suppose further than when signing up for the account, you read and agreed to the following terms and conditions: "[Email provider] reserves the right to pre-screen, review, flag, filter, modify, refuse or remove any or all content from any service. For some services, [email provider] may provide tools to filter out explicit sexual content."

All respondents then answer these questions:

Q1140 Would your agreement to this provision allow the email provider to direct its automated systems to scan the contents of the emails you send and receive and show you personalized advertisements? For example, if emails you send and receive regularly mention the words "tired" and "sleepy" the automated system might show you more ads from mattress sellers.

- ❍ They definitely would be allowed to do so. (4)
- ❍ They probably would be allowed to do so. (5)
- ❍ They probably would not be allowed to do so. (6)
- ❍ They definitely would not be allowed to do so. (7)

Q1141 On a scale of 1 to 10, how intrusive is the email provider's automated email scanning and ad personalization practice?

_____ - (1)

Q1142 If there were an option to keep the same email account but pay some amount of money to avoid having the automated systems analyze email content for the purposes of showing you personalized advertisements, how would you respond?

- ❍ I would keep the free email account with the automated email analysis and personalized advertisements. (3)
- ❍ I would be willing to pay some amount of money to avoid the automated analysis. (2)

Q1143 [For those selecting 2 above] How much would you be willing to pay per year?___ dollars __ cents

Dollars (1)
Cents (2)

Facebook Questions – Random Assignment Among these Three Questions:

Q1152 The following language appears in the Data Policy on Facebook's web site: "We collect the content and other information you provide when you use our Services, including when you sign up for an account, create or share, and message or communicate with others. This can include information in or about the content you provide, such as the location of a photo or the date a file was created. We also collect information about how you use our Services, such as the types of content you view or engage with or the frequency and duration of your activities. We also collect content and information that other people provide when they use our Services, including information about you, such as when they share a photo of you, send a message to you, or upload, sync or import your contact information. We are able to deliver our Services, personalize content, and make suggestions for you by using this information to understand how you use and interact with our Services and the people or things you're connected to and interested in on and off our Services. We also use information we have to provide shortcuts and suggestions to you. For example, we are able to suggest that your friend tag you in a picture by comparing your friend's pictures to information we've put together from your profile pictures and the other photos in which you've been tagged. If this feature is enabled for you, you can control whether we suggest that another user tag you in a photo using the 'Timeline and Tagging' settings."

Q1155 The following language appears in the Data Policy on Facebook's web site: "We use the information we receive about you in connection with the services and features we provide to you and other users like your friends, the advertisers that purchase ads on the site, and the developers that build the games, applications, and websites you use. For example, we may use the information we receive about you:• as part of our efforts to keep Facebook safe and secure;• to provide you with location features and services, like telling you and your friends when something is going on nearby;• to measure or understand the effectiveness of ads you and others see;• to make suggestions to you and other users on Facebook, such as: suggesting that your friend use our contact importer because you found friends using it, suggesting that another user add you as a friend because the user imported the same email address as you did, or suggesting that your friend tag you in a picture they have uploaded with you in it. Granting us this permission not only allows us to provide Facebook as it exists today, but it also allows us

to provide you with innovative features and services we develop in the future that use the information we receive about you in new ways .We are able to suggest that your friend tag you in a picture by comparing your friend's pictures to information we've put together from the photos you've been tagged in. You can control whether we suggest that another user tag you in a photo using the 'How Tags work' settings."

Q579 The following language appears in the Help Center on Facebook's web site: "When someone uploads a photo of you, we might suggest that they tag you in it. We're able to compare your friend's photos to information we've put together from your profile pictures and the other photos you're tagged in. If this feature is turned on for you, you can choose whether or not we suggest your name when people upload photos of you. Adjust this in your Timeline and Tagging settings. We currently use facial recognition software that uses an algorithm to calculate a unique number ('template') based on someone's facial features, like the distance between the eyes, nose and ears. This template is based on your profile pictures and photos you've been tagged in on Facebook. We use these templates to help you tag photos by suggesting tags of your friends. If you remove a tag from a photo, that photo is not used to create the template for the person whose tag was removed. We also couldn't use a template to recreate an image of you."

All respondents then answer these questions:

Q1153 Suppose you had previously read the Data Policy's / Help Center's language and had not adjusted your Timeline and Tagging settings. Suppose further that the following scenario occurs: A Facebook friend of yours uploads a photo of you and them to Facebook. Because Facebook already has analyzed other photos of you, its facial recognition software suggests to your friend that you be tagged (captioned) in the photo, and your friend agrees to tag you in the photo.

	Yes (1)	No (2)
Did Facebook's language (above) inform you that information about your facial features was being collected and stored? (7)	❍	❍
Did Facebook's language (above) inform you of the reason why information about your facial features was being collected, stored, and used? (8)	❍	❍
Did Facebook's language (above) inform you of the length of time for which information about your facial features would be stored? (9)	❍	❍
Would your decision not to adjust your Timeline and Tagging settings allow Facebook to collect, store, and use information about your facial features? (10)	❍	❍

Q1154 On a scale of 1 to 10, how intrusive is Facebook's use of facial recognition software to suggest tags for people whose faces appear in uploaded photos?