

Shifts in Privacy Concerns

Avi Goldfarb and Catherine Tucker*

January 28, 2012

Abstract

This paper explores how digitization and the associated use of customer data have affected the evolution of consumer privacy concerns. We measure privacy concerns by reluctance to disclose income in an online marketing research survey. Using over three million responses over eight years, our data show: (1) Refusals to reveal information have risen over time, (2) Older people are less likely to reveal information, and (3) The difference between older and younger people has increased over time. Further analysis suggests that the trends over time are partly due to broadening perceptions of the contexts in which privacy is relevant.

*Avi Goldfarb is Associate Professor of Marketing, Rotman School of Management, University of Toronto, 105 St George St., Toronto, ON. Tel. 416-946-8604. Email: agoldfarb@rotman.utoronto.ca. Catherine Tucker is Associate Professor of Marketing, MIT Sloan School of Management, 100 Main St, Cambridge, MA. Tel. 617-252-1499. Email: cetucker@mit.edu. We thank seminar participants at Queen's University and the 2012 AEA meetings for helpful comments. This research was supported by an NBER Economics of Digitization and Copyright Initiative Research Grant and by SSHRC grant 486971.

The information and communications technology revolution has led to a seismic shift in the storage and availability of personal consumer data for commercial purposes. This has led privacy advocates to call increasingly for regulation to protect consumers' privacy. However, in order to shape how regulation should look, it is important to understand how consumer privacy concerns have evolved over time alongside this increase in the intensity of data usage.

To address this need, we measure how consumers' privacy concerns have changed using three million observations collected by a market research company from 2001-2008, covering whether consumers chose to protect their privacy by not revealing their income in an online survey. There are three key patterns in the data: (1) Refusals to reveal information have risen over time, (2) Older people are much less likely to reveal information than are younger people, and (3) Though younger respondents have become somewhat more private over time, the gap between younger and older people is widening.¹ We show this across many specifications and with a variety of controls.

There are several possible explanations for these trends. These include an increase in experience with information technology (and awareness of potential privacy concerns), a change in the composition of the online population, a change in the composition of survey respondents, a change in income, and a change in underlying preferences for privacy.

We document that there has been an increase in the number of contexts in which consumers perceive privacy concerns to be relevant. Consumers were always privacy-protective in contexts where they were answering personal questions about financial and health products, but they are increasingly becoming privacy-protective in contexts which have not been highlighted as standard areas of privacy concern (such as answering questions about consumer packaged goods or movies). This finding can also explain the observed differences in privacy-protective behavior across age groups: the difference in information revelation

¹This also fits in with popular discourse that suggests that younger people are relatively uninterested in privacy - 'Kids don't care,' [Walt Disney Co chief Robert] Iger said, adding that when he talked to his adult children about their online privacy concerns 'they can't figure out what I'm talking about.' (Keating, 2009)

between personal and non-personal contexts is largest for older people. Therefore, as more topics are seen as potentially personal, the difference between the total amount revealed by older and younger people grows. Though this is evidence that a change in underlying preferences for privacy explains a substantial part of the trend, this does not rule out the possibility that the other explanations can explain the remainder.

Consistently with recent discussions in law and philosophy, these results suggest a change in consumers' perceptions over what contexts warrant privacy. Zittrain (2009), Selove (2008), and Nissenbaum (2010) all discuss how changes in technology change our perception of the public and private sphere.² Nissenbaum (2010) argues that privacy protection should conform to context-relative informational norms (labeled 'contextual integrity'). As information is collected, aggregated, and disseminated in new ways, Nissenbaum argues that previously-perceived boundaries defining contextual integrity have changed. Taylor (2004) develops an economic model of information-sharing across firms that provides one explanation of why the context in which data is used might matter: if data is used to price-discriminate, and changes in technology mean that firms can sell the data to other firms even outside their sector, then many rational consumers will choose to keep information private in order to prevent future price discrimination.

Our results also are important for a new literature on privacy regulation.³ This literature has so far focused on documented the costs of privacy regulation and the inhibition of data flows (Miller and Tucker, 2009; Goldfarb and Tucker, 2011c; Miller and Tucker, 2011; Campbell et al., 2011). By contrast, in this paper we attempt to provide some of the first behavior-based evidence that traces the diffusion and the motivation for increasing privacy concerns among consumers. This adds to survey-based literature that has asked respondents

²Zittrain (2009) (p. 212) argues that changes in technology "threaten to push everyone toward treating each public encounter as if it were a press conference."

³There is an earlier and somewhat broader literature on economic models of privacy (including Posner (1981), Varian (1997), Acquisti and Varian (2005), and Hermalin and Katz (2006); see Hui and Png (2006) for a review.)

directly about perceptions of privacy (Sheehan and Hoy (2000), Marwick et al. (2010), etc.). For example, Hoofnagle et al. (2010) show that there is a perception of increased attention to privacy over time. In particular, people say they are more focused on privacy than they were five years ago. When asked for a reason, the most common response is that they know more about the risks. This is consistent with an interpretation of our results that the reason that privacy has increased is that more contexts are seen as private.

1 Data

1.1 Dependent Variable: Refusal to provide income

In economics, privacy has commonly been modeled as the transfer of information between consumer and firm and why consumers may find that undesirable (Posner, 1981; Acquisti and Varian, 2005; Hui and Png, 2006). A key challenge in identifying how privacy concerns have changed is that the ways in which people can reveal information have grown over time. Therefore, it is difficult to identify an underlying propensity to reveal information while controlling for the increased opportunities to reveal, changing potential benefits of revealing information, and changing risks and costs to revealing. In order to address this challenge, we look at a particular type of information revelation and examine it over time. Specifically, we look at the propensity to reveal income in a web-based survey by a marketing research company that received over three million individual responses from 2001 to 2008. The aim of this database is to provide comparative guidance to advertisers about which types of ad design are effective and to benchmark current campaigns against past campaigns. This database is among the main sources used by the online media industry to benchmark online display ad effectiveness. We have used the data to measure advertising performance directly in other studies, including Goldfarb and Tucker (2011b,a,c).

Despite its primary marketing focus, the database is particularly useful for understanding

changes in underlying privacy preferences over time because the benefits and costs for revealing this information relate to inherent preferences for privacy rather than to any explicit gain or risk associated with the information. Furthermore, the precise question asked has remained constant over time.

The survey began with various questions about the likelihood of purchase for multiple different products and brands. In this paper, we focus on what these different product categories meant for the context of the survey. Was the survey asking about purchase preferences for health insurance? Or was the survey asking about purchase preferences for detergent?

After answering questions on the product category, the respondent is asked to provide demographic information, specifically income, age, gender, internet use, zipcode, and income. When asked to give their income level, respondents were given the option of ‘Prefer not to say’. Refusals to provide income are several times higher than other refusals: 15% on average for income and less than 0.5% for all other questions. We use this decision to keep income information as our core dependent variable.

Specifically, our dependent variable is whether respondents provide income information, conditional on agreeing to participate in the survey overall. The survey changed little over time, with the main exception being that since September 2004 the survey has asked about hours online. This means that we can interpret the refusals to provide income information consistently across time. Table 1 shows summary statistics for this survey data and Table 2 shows the specific question asked on income.

It is well established that income nonresponse rates are substantially higher than nonresponse rates to other questions (Yan et al., 2010). There are many reasons why respondents might want to keep their income private, including concern about governments or firms exploiting that information or a general sense that income is something private that they do not want to release. Yan et al. (2010) emphasize that a likely reason income nonresponse

Table 1: Marketing Research Database: Summary Statistics

	Mean	Std Dev	Min	Max
Keeps Income Private	0.13	0.33	0	1
Years since 2001	5.95	1.72	1	9
Age	45.51	12.58	25	75
Female	0.50	0.50	0	1
Weekly Internet Hours	14.14	10.16	1	31
21+ hrs Internet	0.22	0.41	0	1
Observations	3319710			

Table 2: Income Question

Question	Possible Responses
What is your income?	Under 10,000
	20,000-29,999
	30,000-39,999
	40,000-49,999
	50,000-74,999
	75,000-99,999
	100,000-149,999
	150,000-199,999
	Over 200,000
	Rather not say

is high is that discussing income is often deemed inappropriate for everyday conversation. Therefore, they argue that “income nonresponse due to sensitivity is largely an issue of respondent motivation, and is driven by privacy concerns” (p. 146). They also mention that refusals to answer (in contrast to “don’t knows”) are generally due to privacy concerns rather than lack of knowledge of precise income and provide evidence that the trend toward increased nonresponse is more likely due to privacy-related reasons than information-related reasons. In our analysis we also attempt to rule out explanations based on lack of familiarity in income by exploiting the fact that around April 15 people are better informed about their income than at other times.

1.2 Other Explanatory Variables

In addition to this question about income which is our key explanatory variable, the survey also asked about age, gender, zipcode, and internet use.

The age question was open-ended so respondents simply type in their age. Our raw data includes respondents from age 10 to 90. We focus on respondents between ages 25 and 75 for two reasons. First, at the extremes (above 75 and below 18) the data become sparse and therefore cohort effects are less reliably identified. Second, between ages 18 and 25, many respondents appear to be students (judging from their locations in college towns) and therefore income refusals are as likely to be about a lack of income as about privacy. We show robustness of the core results to the full sample.

For hours online, the survey asks “How many hours per week do you use the internet” followed by six options: “31 hours or more”, “21 to 30 hours”, “11 to 20 hours”, “5 to 10 hours”, “1 to 4 hours”, and “Less than 1 hour”. However, the survey only asked this income question after September 2004. In our empirical analysis we show robustness to using focusing the period after September 2004 when survey format did not change.

The survey began with various questions about the likelihood of purchase for multiple

different products and brands. Underlying these questions was a prior treatment-control allocation of internet ads. Specifically, after the user arrived at the website they were randomly shown the ad to be tested (say, a detergent ad) or a placebo ad (say, an ad for a charitable organization). The survey was then focused on asking questions, for example, about the likelihood of purchase of that detergent brand. The difference between the treatment and control groups is considered the ad effectiveness. We explore questions related to advertising effectiveness in other work (Goldfarb and Tucker, 2011a,b,c). In this paper, though, we focus on what these different product categories meant for the context of the survey. Was the survey asking about purchase preferences for life insurance products? Or was the survey asking about purchase preferences for detergent?

Overall, from the marketing research data of over 3 million online surveys, we use whether they provided their income as the dependent variable that allows us to measure privacy concerns in a consistent way over time. We use information on the date of the survey and the age of the respondent to identify our main patterns of interest. In places, we also use information on gender, the broad category of the product advertised (e.g. automotive, consumer goods), the topic of the survey (e.g. Asian brand SUV between \$20k and \$40k, Urogenital preventative lifestyle symptomatic drugs), the category of the website where the advertisement appeared (e.g. portal, news), zipcode (to identify county), and time spent online.

1.3 Survey Respondent Selection

The setting we study is unusual in several ways. The responses are gathered when a website visitor is asked to fill out an online questionnaire that appears as a pop-up window as the website visitor tries to navigate away from the webpage where the treatment (e.g. detergent) or control (e.g. charity) ad is served. Therefore, we are looking at non-response to a question about income among a population who self-selected into responding to an entirely

optional online study. Encouragingly, the demographic variables reported in Table 1 appear representative of the general internet population at the time of the study as documented in the Computer and Internet Use Supplements to the 2003 and 2007 Current Population Surveys.

In order to assess further the degree to which the respondent age and income information is trustworthy, we use Direct Marketing Association data on zip code characteristics from 2000. We merge this data using the zipcodes provided by respondents. We find that zipcode-level income is strongly correlated with reported income by our survey respondents. Similarly, the fraction of the zipcode in each age category (ages 25 and 34, 35 and 44, between ages 45 and 54, between ages 55 and 64, and over age 65—as noted above, we exclude people under 25 from our main sample) is similar to the fraction of the zipcode in our data by age category, with strongly significant correlation coefficients between 0.18 and 0.36.

This clearly does not fully mitigate concerns that our sample is relatively willing to give information online. However, we believe this difference will reinforce, rather than contradict, many of our results: these are consumers who willingly gave other data to a firm but who, when given the option, choose to keep an answer private. In particular, as noted by Yan et al. (2010), given that overall response rates are falling, looking at response rates to income questions conditional on answering the survey understates declines in response rates over time. That said, general sample selection concerns suggest that our results on the overall trends in privacy might be driven by selection effects: if more privacy-conscious consumers are late adopters of the internet in general and online surveys in particular. In contrast, our results on the role of context are less likely to be changed as a result of such selection concerns.

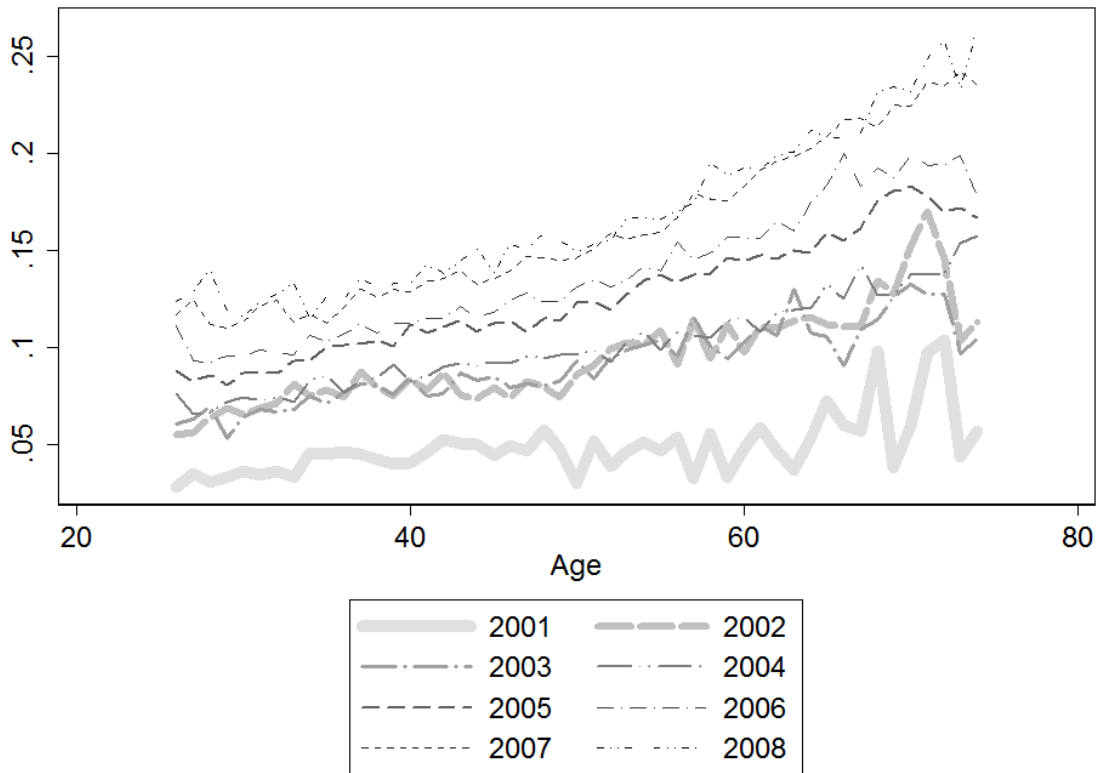


Figure 1: Cohort and Age effects over time

2 Data Analysis

Figure 1 displays an initial graphical representation of the correlation between age, privacy-seeking, and year. The y-axis is the percent of respondents who choose to keep their income private, and the x-axis is age. Each line represents the respondents in a particular year. The figure shows three patterns that will be confirmed in the econometric analysis. First, all lines are upward sloping: so older people are more privacy-protective than younger people. Second, the lines steadily rise over time, so privacy is increasing over time across all age groups. Third, though perhaps somewhat less clearly in the picture, the difference across age groups is increasing over time.

In order to better document and understand these patterns, we next turn to our re-

gression analysis. Our econometric estimation approach examines the relationship between individual-, survey-, and location-level factors and whether the respondent reveals their income. The dependent variable is refusal to reveal income, and therefore measures whether respondents choose to keep their income private. We use a simple linear regression model:

$$Private_{ist} = \alpha age_{ist} + \beta date_{ist} + \gamma age_{ist} \times date_{ist} + \theta_1 X_{it} + \theta_2 Z_{st} + \epsilon_{ist} \quad (1)$$

where i marks the individual respondent, s marks the survey, and t marks the date; X_{it} contains respondent-specific controls such as gender, internet use, changes in local payroll, and county fixed effects, Z_{st} contains survey-specific controls such as broad product category, website category, and whether the product mentioned in the survey is something that is generally considered a private topic.⁴ Because each respondent i appears only once, it is not a panel; we include the subscripts for s and t to clarify the key variation on the right hand side of the equation. One way to interpret this framework is the following simple cost-benefit model of keeping information private: people refuse to reveal their income if the perceived benefits of privacy exceed the perceived costs. Perceived costs and benefits might be affected by perceptions of how income data is used and whether filling out the survey is a worthwhile endeavor (Demaio, 1980; Yan et al., 2010).

We report our main results using a linear probability model, though we also show robustness to a probit formulation. This robustness is in line with the findings of Angrist and Pischke (2009) that there is typically little qualitative difference between the probit and linear probability specifications. We focus on the linear probability model because it enables us to estimate a model with thousands of county fixed effects using the full data set of over 3 million observations (the linear functional form means that we can partial out these fixed

⁴The data identify 22 consistently applied product categories (17 of which appear in at least seven years), 38 consistently applied website categories (26 of which appear in at least seven years), and 364 highly varying survey topics (only 33 of which appear in at least seven years). The survey topics are not consistently identified over time: they are identified and developed for each new campaign. Therefore we can include fixed effects for product categories and website categories but not specific survey topics.

effects through mean-centering), whereas computational limitations prevent us from estimating a probit model with this full set of fixed effects. Likely because the mass point of the dependent variable is far from 0 or 1, the predicted probabilities of our main specifications almost all lie between 0 and 1. This means that the potential bias of the linear probability model if predicted values lie outside of the range of 0 and 1 (Horrace and Oaxaca., 2006) is less of an issue in our estimation.

Table 3: Privacy concerns, demographics, and time trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years since 2001	0.017*** (0.00015)				0.013*** (0.00015)	0.012*** (0.00016)	0.013*** (0.00022)
Age		0.0024*** (0.000029)			0.0022*** (0.000030)	0.0023*** (0.000031)	0.0023*** (0.000029)
Female			0.045*** (0.00072)		0.049*** (0.00077)	0.045*** (0.00074)	0.046*** (0.00073)
Average Payroll (00,000)				0.085*** (0.017)	0.078*** (0.0096)	0.085*** (0.0094)	-0.011 (0.014)
Constant	0.028*** (0.0011)	0.019*** (0.0014)	0.10*** (0.00081)	0.094*** (0.0062)	-0.11*** (0.0037)	-0.11*** (0.0042)	-0.083*** (0.0052)
Month	No	No	No	No	No	Yes	Yes
Site Category	No	No	No	No	No	Yes	Yes
Product Category	No	No	No	No	No	Yes	Yes
County Fixed Effects	No	No	No	No	No	No	Yes
Observations	3319710	3319710	3319710	3319710	3319710	3176706	3176706
Log-Likelihood	-1047065.8	-1046176.7	-1051880.4	-1058133.3	-1027740.1	-989796.9	-985945.6
R-Squared	0.007	0.008	0.005	0.0008	0.02	0.02	0.02

Dependent Measure is whether respondent refused to provide income in survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.1 Privacy trends across age groups and over time

Table 3 documents the basic correlations between the respondent characteristics and refusals to provide income information. Column 1 shows the increase in the propensity to keep income private over time. Column 2 shows that privacy increases with age. These are two of the main effects of interest in the paper. Column 3 shows that women are more likely to keep income private and Column 4 shows that counties with higher payroll on average keep more information private. Column 5 shows that these results do not change if all four covariates are included, and Column 6 shows no change in the correlations if we include fixed effects for 22 categories of products (e.g. consumer package goods, entertainment, finance) and fixed effects for 38 categories of websites where the advertisement appeared (e.g. cooking, portal, sports). Column 7 adds county fixed effects and shows that the results on the respondent-specific covariates do not change: privacy increases over time, privacy increases with age, and women keep income more private than men. County fixed effects drop the significance and reverse the sign on average payroll. In most results that follow, we use the county fixed effects model in Column 7 as the base model.

The fit of the regressions in Table 3 (measured by R -squared), and all those shown below, is never more than 0.03. Given the numerous reasons why a particular individual may refuse to provide their income, we do not view this as surprising. It is consistent with many other studies that use linear probability models to understand patterns in large data sets including Athey and Stern (2002) and Goldfarb and Tucker (2011b). What is important is the estimated effects are economically large and statistically relevant. Column 7 of Table 3 suggests this is the case. Refusals increase by 1.3 percentage points per year relative to an average of 13 percent. The age effects are also large: refusals rise 4.6 percentage points when age increases from 25 to 45.

Therefore, Table 3 documents that people are becoming more likely to keep income private over time and that older people are more private than younger people. In Table 4, we examine

the interaction of these effects: are older people becoming relatively more private over time? Using a number of distinct specifications, we show that they are: the interaction between years and age is positive, significant, and large. Column 1 regresses refusal to provide income on age, data, and their interaction without further controls; Column 2 includes the same controls as Column 6 of Table 3; Column 3 uses a probit specification; Column 4 includes the full set of controls from the main specification in Column 7 of Table 3; Columns 5 and 6 show that the results are robust to including only March and April, and to excluding March and April (to address the concern that people are differently aware of income and privacy during tax season); and Column 7 shows that the results are robust to including all ages rather than just those between 25 and 75.⁵

⁵Though Column 7 suggests that the negative coefficient on age is not robust, the interaction means that the effect is positive for all ages

Table 4: Privacy concerns over time across age groups

	OLS						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Probit	Panel	Tax Months	Non-Tax Months	Missing Age
Years since 2001	0.0022*** (0.00049)	0.00051 (0.00049)	0.050*** (0.0028)	0.0017*** (0.00051)	0.0015 (0.0011)	0.0019*** (0.00055)	0.0048*** (0.00048)
Age	0.00037*** (0.000061)	0.00070*** (0.000061)	0.0088*** (0.00036)	0.00068*** (0.000061)	0.00085*** (0.00012)	0.00066*** (0.000066)	-0.0012*** (0.000065)
Years × Age	0.00027*** (0.000011)	0.00026*** (0.000011)	0.00035*** (0.000058)	0.00026*** (0.000011)	0.00021*** (0.000021)	0.00027*** (0.000012)	0.00025*** (0.000010)
Female		0.045*** (0.00074)	0.22*** (0.0037)	0.045*** (0.00073)	0.042*** (0.0011)	0.046*** (0.00080)	0.041*** (0.00063)
Average Payroll (00,000)		0.084*** (0.0094)	0.40*** (0.045)	0.0088 (0.013)	0.011 (0.030)	0.0086 (0.014)	0.022 (0.017)
Month	No	Yes	Yes	Yes	Yes	Yes	Yes
Site Category	No	Yes	Yes	Yes	Yes	Yes	Yes
Product Category	No	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	No	Yes	Yes	Yes	Yes	Yes
Observations	3319710	3176706	3176660	3176706	497325	2679381	3740177
Log-Likelihood	-1036993.0	-989378.7	-1180644.7	-985538.6	-142561.3	-840682.9	-1315238.3
R-Squared	0.01	0.02		0.02	0.03	0.02	0.02

Dependent Measure is whether respondent refused to provide income in survey. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.2 Understanding privacy trends across age groups and over time

Next we explore possible drivers of the increase in refusals over time, and why this increase is sharper for older people.

Table 5 explores one possible explanation: that people are using the internet more over time (and this increase is larger for older people over the 2001 to 2008 time period). If spending more time online makes people more aware of the potential for data misuse, it is possible that they become increasingly likely to refuse to provide private information such as income. However, Table 5 shows that the people who spend the most time online are the most likely to reveal their income.

Column 1 shows that as time online rises (the base is less than one hour online per week—by definition, an online survey means all respondents are online sometimes), people become less likely to keep their income private. Columns 2 and 3 split the data at whether the respondent spends 21 hours online per week. Column 2 shows only respondents who spend less than 21 hours per week and Column 3 shows respondents who spend more than 21 hours per week. The results show little substantive difference between the groups. If anything, we see increasing price sensitivity over time among those who spend few hours online per week. Because internet use is increasing over time, it is not clear whether this is a selection effect or a respondent-level trend. Column 4 simply verifies the main results hold when using the data since September 2004, when the question about hours online entered the survey. Overall, the results show no change in the trend toward privacy when we control for time online. If anything, it works against the trend in that internet use is increasing and we see more privacy among those who spend less time online.

Table 7 provides evidence that changing perceptions of what is private help to explain the trend. Applying Acquisti et al. (2011), who showed that questions that appear early in a survey frame privacy-protective responses, we use the topic of questions early in the survey to identify situations that do and do not obviously warrant privacy. As people become

Table 5: Technology usage does not explain this

	(1)	Over 21 Hrs (2)	Below 21 Hrs (3)	All post Sept 2004 (4)
Internet 1-4 Hours	-0.012*** (0.0014)			
Internet 5-10 Hours	-0.012*** (0.0014)			
Internet 11-20 Hours	-0.014*** (0.0014)			
Internet 21-30 Hours	-0.016*** (0.0014)			
Internet 31+ Hours	-0.017*** (0.0015)			
Years since 2001	0.0033*** (0.0010)	0.0040*** (0.0012)	0.0022 (0.0017)	0.0033*** (0.0010)
Age	0.0014*** (0.00014)	0.0014*** (0.00016)	0.0013*** (0.00023)	0.0014*** (0.00014)
Age \times Years	0.00017*** (0.000021)	0.00016*** (0.000025)	0.00016*** (0.000035)	0.00017*** (0.000021)
Female	0.038*** (0.00078)	0.041*** (0.00090)	0.032*** (0.0012)	0.038*** (0.00078)
Average Payroll (00,000)	0.0081 (0.020)	0.046* (0.024)	-0.064* (0.037)	0.0073 (0.020)
Month	Yes	Yes	Yes	Yes
Site Category	Yes	Yes	Yes	Yes
Product Category	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Observations	2352197	1652905	699292	2352197
Log-Likelihood	-864444.5	-617650.4	-244645.5	-864536.9
R-Squared	0.02	0.02	0.02	0.02

$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

increasingly aware of the various uses of online data, it is possible that they become increasingly private in situations that do not obviously warrant privacy. Under this hypothesis, the increase in privacy would be sharpest in non-personal contexts such as surveys about entertainment, games, and consumer packaged goods. In contrast, people have always known that financial and health-related contexts require privacy and we expect less change over time.

The role of survey characteristics and using early questions to frame privacy concerns in privacy-protective responses is highlighted by .

Table 6 summarizes the survey topics that we defined as personal out of the 364 possible survey topics.

Column 1 of Table 7 confirms that people are indeed more likely to keep income information private when the survey is about a personal topic. Column 2 shows that this is decreasing over time: as we move from 2001 to 2008, the difference between personal and non-personal topics narrows. The coefficient values suggest that the difference in whether people reveal income in personal and non-personal topics disappears completely after eight years. Put another way, 24 percent ($0.0032/(0.0032+0.0100)$) of the increase in privacy can be explained by decreasing willingness to reveal income information in connection with ‘non-personal’ topics.

Column 3 shows the full set of interactions with age and date. The three-way interaction between age, date, and non-personal topic suggests that the increase in privacy for non-personal topics is sharpest for older people. Interestingly, the two-way interaction between age and non-personal topic is negative, suggesting that, early in the sample, the difference between older and younger people is larger for personal topics than non-personal topics.

Columns 4 and 5 show that the results of Columns 2 and 3 are robust to a broader definition of “personal” based on 5 of the 22 broad product categories rather than the detailed survey topics (specifically financial products, government services, health, pharmaceuticals,

Table 6: List of Personal Items

Urogenital preventative lifestyle symptomatic drugs	Auto insurance	tomatic drugs
Urogenital preventative lifestyle asymptomatic drugs	General insurance	Other erectile dysfunc- tion drugs
Other Urogenital drugs	Health insurance	Diabetes preventative life threatening symp- tomatic drugs
Intimates	Home insurance	Diabetes preventative lifestyle symptomatic drugs
Antibacterial/fungal/viral episodic lifestyle symp- tomatic drugs	Other insurance	Other diabetes drugs
Antibacterial/fungal/viral preventative life threat- ening symptomatic drugs	Investment brokerages	Credit cards-brand
Antibacterial/fungal/viral preventative lifestyle symptomatic drugs	General investment ser- vices	Credit cards-business
Other antibacte- rial/fungal/viral drugs	Market investment ser- vices	Credit cards-personal
Personal Care Feminine Hygiene products	Mutual funds	Credit cards-other
Home loans	Retirement investment services/IRA/401k	Cancer episodic life threatening symp- tomatic drugs
Mortgages	Banking	Cancer episodic life threatening asymp- tomatic drugs
Other loans	Other investment ser- vices	Other cancer drugs
	Other financial services	Birth control preven- tative lifestyle asymp- tomatic drugs
	Mental health lifestyle symptomatic drugs	Other birth control.
	Mental health lifestyle asymptomatic drugs	
	Other mental health drugs	
	Erectile dysfunction episodic lifestyle symp-	

and professional services). Here the main effect of the non-personal topic is collinear with the category fixed effects and is therefore dropped.

Table 8 shows the same set of results but using discrete covariates to facilitate interpretation of the three-way interaction. Again, Column 1 shows that people are less likely to keep their income private when the survey is not about a personal topic. Column 2 shows that this difference is much sharper from 2001 to 2004 than from 2005 to 2008. Column 3 suggests that about half of the trend toward older people being more private can be explained by differences in the trends of how older people respond to non-personal topics. In particular, older people became especially private about their income for non-personal topics. Columns 4 and 5 again verify robustness to the broader definition of “personal” for the survey topic.

This suggests a likely partial explanation for the observed rise in privacy over time: people are increasingly treating all online information sharing the same. The decision to keep income private depended much less on the topic of the survey in 2008 compared to 2001. To the extent that survey topics represent personal and non-personal settings, the level of privacy desired on average in seemingly non-personal settings has moved toward the level of personal settings.

This is consistent with Nissenbaum (2010): that technological advances in the ability to track, aggregate, and disseminate information have led to a change in the contexts that can be seen as private. Thus, a plausible interpretation of our result is that people are increasingly recognizing that even seemingly non-personal online contexts are not private.

Furthermore, the growing difference between older and younger people can also be partially explained by these changing perceptions of context. Older people have always been more private in seemingly private contexts such as surveys on financial and health topics. As perceptions of contextual integrity change and all respondents increasingly see all contexts as having similar potential to raise privacy concerns, the relative impact on older people will be higher. In other words, surveys about non-personal topics were increasingly recognized as a

context in which privacy is relevant, particularly for older respondents. So, as the difference between personal and non-personal topics has narrowed for both older and younger people, the relative effect on older people has been larger.

Table 7: Broadening Privacy Concerns May Explain this

	Narrow Definition		Broad Definition		
	(1)	(2)	(3)	(4)	(5)
Non-personal Topic	-0.0052*** (0.0013)	-0.024*** (0.0023)	-0.012* (0.0062)		
Non-personal Topic × Years		0.0032*** (0.00033)	0.0013 (0.0011)	0.0053*** (0.00029)	0.0041*** (0.0010) -0.00019 (0.00013) 0.000034 (0.000022) 0.00024*** (0.000019)
Non-personal Topic × Age			-0.00031** (0.00014)		
Non-personal Topic × Age × Years			0.000051** (0.000024)		
Age × Years			0.00022*** (0.000021)		
Years since 2001	0.013*** (0.00022)	0.010*** (0.00036)	0.00045 (0.00094)	0.0090*** (0.00032)	-0.0020** (0.00083)
Age	0.0023*** (0.000029)	0.0023*** (0.000029)	0.00091*** (0.00012)	0.0023*** (0.000029)	0.00077*** (0.00011)
Female	0.045*** (0.00073)	0.045*** (0.00073)	0.045*** (0.00073)	0.045*** (0.00073)	0.045*** (0.00074)
Average Payroll (00,000)	-0.011 (0.014)	-0.0067 (0.013)	0.014 (0.012)	-0.0026 (0.013)	0.018 (0.012)
Month	Yes	Yes	Yes	Yes	Yes
Site Category	Yes	Yes	Yes	Yes	Yes
Product Category	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	3176706	3176706	3176706	3176706	3176706
Log-Likelihood	-985936.8	-985877.5	-985456.8	-985756.6	-985317.7
R-Squared	0.02	0.02	0.02	0.02	0.02

$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Broadening Privacy Concerns May Explain this: Discrete Specification

	Narrow Definition		Broad Definition		
	(1)	(2)	(3)	(4)	(5)
Non-personal Topic	-0.0055 (0.0013)	-0.021*** (0.0017)	-0.017*** (0.0020)		
Non-personal Topic \times >2004		0.020*** (0.0014)	0.015*** (0.0019)	0.019*** (0.0012)	0.018*** (0.0016)
Non-personal Topic \times Age >40			-0.0091*** (0.0017)		-0.0027* (0.0016)
Non-personal Topic \times Age >40 \times Year >2004			0.011*** (0.0021)		0.0033 (0.0020)
Age >40 \times >2004			0.012*** (0.0020)		0.018*** (0.0018)
Year >2004	0.028*** (0.0029)				
Age >40	0.041*** (0.00066)	0.041*** (0.00066)		0.041*** (0.00066)	
Female	0.043*** (0.00069)	0.043*** (0.00069)	0.043*** (0.00070)	0.043*** (0.00069)	0.043*** (0.00070)
Average Payroll (00,000)	0.38*** (0.084)	0.38*** (0.083)	0.39*** (0.081)	0.38*** (0.083)	0.39*** (0.082)
Year >2004		0.012*** (0.0024)	0.0048** (0.0024)	0.013*** (0.0025)	0.0022 (0.0025)
Age >40			0.033*** (0.0016)		0.027*** (0.0014)
Month	Yes	Yes	Yes	Yes	Yes
Site Category	Yes	Yes	Yes	Yes	Yes
Product Category	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	3176706	3176706	3176706	3176706	3176706
Log-Likelihood	-993014.3	-992880.7	-992642.3	-992877.1	-992639.1
R-Squared	0.02	0.02	0.02	0.02	0.02

$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3 Conclusions

This paper examines how privacy concerns have changed over time and across age groups. To create a consistent measure of privacy, we focus on refusals to provide income information in a particular online survey. Our results suggest that people are becoming more privacy sensitive over time, and this is particularly true of older people. Our results also suggest that the trend toward increased privacy and the growing difference between older and younger people in privacy concerns partly reflects changing perceptions of the contexts in which privacy is warranted. When online, people used to be much more private in health and financial contexts than other contexts. This was especially true of older people. This has changed, and now people are equally, and increasingly, private in all contexts.

As with any study, our results have several limitations. While we document an increase in privacy over time and with age, a key question is the extent to which we can think that the decision to not answer a potentially intrusive question on a survey is a good proxy for privacy-protective behavior or a consumer having privacy concerns in general. While Cho and Larose (1999), Day (1975), and others argue that privacy violations can arise from sampling, contact, and response methods in online surveys, at best we study one particular measure of privacy. Furthermore, our explanation for a substantial fraction of the trend over time relies on the particular context of the online survey topic. Any attempt to extrapolate to other contexts is necessarily speculative. Finally, we leave three-quarters of the trend unexplained, and cannot reject the possibility that some of this is due to changes in who is online and who is responding to the survey.

Despite these limitations, we provide novel behavior-based evidence of the existence of, and the reasons for, increasing privacy concerns among consumers. Our results suggest that attention to a particular type of privacy is increasing. As more and more contexts appear to be private, people are revealing less information in the same situations. Furthermore,

because older people are particularly private in personal contexts, the difference between older and younger people has widened.

References

- Acquisti, A., L. K. John, and G. Loewenstein (2011). The impact of relative standards on the propensity to disclose. *Journal of Marketing Research Forthcoming*.
- Acquisti, A. and H. R. Varian (2005). Conditioning prices on purchase history. *Marketing Science* 24(3), 367–381.
- Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton Press.
- Athey, S. and S. Stern (2002, Autumn). The impact of information technology on emergency health care outcomes. *RAND Journal of Economics* 33(3), 399–432.
- Campbell, J. D., A. Goldfarb, and C. Tucker (2011). Privacy Regulation and Market Structure. *mimeo, University of Toronto*.
- Cho, H. and R. Larose (Winter 1999). Privacy issues in internet surveys. *Social Science Computer Review* 17(4), 421–434.
- Day, G. S. (1975). The threats to marketing research. *Journal of Marketing Research* 12(4), pp. 462–467.
- Demaio, T. J. (1980). Refusals: Who, where and why. *Public Opinion Quarterly* 44(2), 223–233.
- Goldfarb, A. and C. Tucker (2011a). Advertising Bans and the Substitutability of Online and Offline Advertising. *Journal of Marketing Research* 48(2), 207–227.
- Goldfarb, A. and C. Tucker (2011b). Online display advertising: Targeting and obtrusiveness. *Marketing Science* 30(3), 389–404.

- Goldfarb, A. and C. Tucker (2011c). Privacy regulation and online advertising. *Management Science* 57(1), 57–71.
- Hermalin, B. and M. Katz (2006, September). Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics* 4(3), 209–239.
- Hoofnagle, C. J., J. King, S. Li, and J. Turow (2010). How Different are Young Adults from Older Adults When it Comes to Information Privacy Attitudes and Policies? *Mimeo, Berkeley*.
- Horrace, W. C. and R. L. Oaxaca. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economic Letters* 90, 321–327.
- Hui, K. and I. Png (2006). *Economics and Information Systems, Handbooks in Information Systems, vol. 1*, Chapter 9: The Economics of Privacy. Elsevier.
- Keating, G. (2009, 23 July). Disney CEO bullish on direct web marketing to consumers. *Reuters*.
- Marwick, A. E., D. M. Diaz, and J. Palfrey (2010). Youth, Privacy, and Reputation. *Mimeo, Harvard Law School*.
- Miller, A. R. and C. Tucker (2009, July). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science* 55(7), 1077–1093.
- Miller, A. R. and C. Tucker (2011, 4). Can healthcare information technology save babies? *Journal of Political Economy* 119(2), 289–324.
- Nissenbaum, H. (2010). *Privacy in context: technology, policy, and the integrity of social life*. Stanford Law Books.
- Posner, R. A. (1981). The economics of privacy. *American Economic Review* 71(2).

- Selove, D. (2008). *Understanding Privacy*. Harvard University Press, Cambridge, MA.
- Sheehan, K. B. and M. G. Hoy (2000). Dimensions of privacy concern among online consumers. *Journal of Public Policy and Marketing* 19, 62–73.
- Taylor, C. R. (2004, Winter). Consumer privacy and the market for customer information. *RAND Journal of Economics* 35(4), 631–650.
- Varian, H. (1997). Economic aspects of personal privacy.
- Yan, T., R. Curtin, and M. Jans (2010). Trends in income nonresponse over two decades. *Journal of Official Statistics* 26(1), 145164.
- Zittrain, J. (2009). *The Future of the Internet—And How to Stop It*. Yale University Press.

Table A-1: Robustness Table—Discrete Specification

	Baseline (1)	Post Sept 2004 (2)	Probit (3)	Tax Months (4)	Non-Tax Months (5)
Age>40	0.026*** (0.00078)	0.043*** (0.00086)	0.18*** (0.0052)	0.023*** (0.0019)	0.026*** (0.00084)
Year>2004	0.017*** (0.0026)		0.21*** (0.0048)	0.016*** (0.0026)	0.017*** (0.0028)
Year>2004 × Age>40	0.020*** (0.0011)		0.043*** (0.0060)	0.019*** (0.0022)	0.020*** (0.0011)
Year>2006		0.020*** (0.0011)			
Year>2006 × Age>40		0.0053*** (0.0012)			
Female	0.043*** (0.00069)	0.035*** (0.00075)	0.21*** (0.0035)	0.040*** (0.0011)	0.044*** (0.00076)
Average Payroll (00,000)	0.39*** (0.083)	0.084** (0.034)	0.41*** (0.054)	0.29*** (0.060)	0.41*** (0.087)
Month	Yes	Yes	Yes	Yes	Yes
Site Category	Yes	Yes	Yes	Yes	Yes
Product Category	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	3176706	2352197	3176660	497325	2679381
Log-Likelihood	-992804.2	-869147.5	-1187983.4	-143499.5	-847029.4
R-Squared	0.02	0.01		0.02	0.02

$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.