# Consumer Heterogeneity, Privacy, and Personalization: Evidence from the Do-Not-Call Registry

Khim-Yong Goh*, Kai-Lung Hui**, I.P.L. Png†

This version: October 2009

Abstract

We develop a new theory which suggests that the demand for privacy could increase or fall with consumer heterogeneity. The key premise of our theory is that even when two populations have the same mean-level characteristics, the distribution of consumer utility *per se* may contribute to different consumer demands. We tested our theory in the context of U.S. Do-Not-Call (DNC) registration. Using detailed DNC registration and consumer demographics, we show that heterogeneity in ethnicity, age, education and religion systematically affected DNC registrations. Our results were robust to measurement and inclusion of key variables, sample specifications, and controlling for possible confounding factors. Our theory can be generalized to scenarios in which goods and services are offered at a particular price to heterogeneous consumers even when there is no issue of privacy. Based on our findings, we draw managerial implications on segmentation and public policy toward the protection of privacy, and the pricing of information products and personalization.

Key words: consumer heterogeneity, privacy, Do-Not-Call registry, personalization

* Department of Information Systems, National University of Singapore
** Department of Information Systems, Business Statistics, and Operations Management, Hong Kong University of Science and Technology
† NUS Business School and Department of Information Systems, National University of Singapore

---

Corresponding Author: Kai-Lung Hui, Dept. of Information Systems, Business Statistics, and Operations Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. We are grateful to Hal Varian for pointing us to the Do-Not-Call registry, and for comments at the City University of Hong Kong, Fudan University, ESMT, IOMS 2007, Singapore Management University, Telecom Paris, and University of Paris-Sud, and to Qiu-Hong Wang for assistance with the data. This version supersedes an earlier draft with the title, “Consumer Privacy and Marketing Avoidance: Evidence from the Do-Not-Call Registry”. This research was supported in part by the Singapore Ministry of Education, grants R253-000-030-112 and R253-000-034-112, and Hong Kong SAR Central Policy Unit and Research Grants Council, Project HKUST 1001-PPR-3.

## 1. Introduction

With limited exceptions, U.S. federal law prohibits unsolicited telemarketing calls to telephone numbers on the Do-Not-Call (DNC) registry. The U.S. Federal Trade Commission (FTC) opened the DNC registry on June 27, 2003. Within 24 hours, over 10 million telephone numbers were registered. Registrations reached almost 50 million within 9 weeks, and 62 million within a year.[1]

The rapid rate of DNC registration was phenomenal, and indicated that U.S. consumers are very concerned about privacy.[2] In general, if telephone marketing provides sufficient benefits (such as convenience and attractive offers) to the extent of offsetting the costs (such as annoyance and time spent on the telephone) that it inflicts on consumers, then consumers would not register for DNC. The large number of registrations, however, suggests that, for many people, the costs imposed by telemarketing outweigh the benefits.

From the perspective of public policy, the strong demand for DNC suggests that industry self-regulation, the position adopted by FTC in safeguarding consumer privacy (see Laudon 1996, Culnan and Armstrong 1999, Milberg et al. 2000, and Culnan and Bies 2003), may not sufficiently align the interests of businesses with consumers. Perhaps not surprisingly, various U.S. states are now proposing a similar "do not mail" registry to curb direct mail, and some privacy advocates have proposed a "do not track" registry to stop businesses from tracking consumers' Web surfing.[3] Internationally, Australia and Canada established federal do-not-call registries in 2007 and 2008 respectively.

---

[1] If we make the simplistic assumption that every registration came from a different household, then 62 million roughly represents 60% of all U.S. households. Obviously, the real percentage should be lower because some households may register multiple telephone numbers, but this estimate suffices to highlight the rapid penetration of the DNC registry.

[2] In this paper, privacy refers to seclusion and autonomy, that is, individual's desire for peace and quiet, and to be free from observation and avoid being manipulated, dominated or exposed by others (Westin 1967; Hirshleifer 1980; Posner 1981; Hui and Png 2006).

[3] In 2004, the FTC has recommended against a "do not email" registry because such a registry would only reduce the volume of spam when an email authentication system is in place.

The key question, however, is whether and the extent to which such "no contact" registries benefit consumers. In particular, how effective are public policies like DNC when the consumer populations exhibit different mixes and characteristics? What can DNC registration tell us about the interaction between direct marketing and consumer privacy, and technology-enabled marketing practices such as personalization and pricing of digital products?

Using data from the DNC registry and U.S. Census, Varian et al. (2005) identified various demographic factors, including income and race, which may affect DNC registrations. Their research focused on *average* demographics, for example, the proportion of Hispanics in the population. Here, using data from similar sources, we propose theoretically and test empirically the proposition that consumer heterogeneity is an important factor that dictates the benefits brought by the DNC registry. Instead of average demographics, our emphasis is on *heterogeneity*, for example, ethnic heterogeneity among Hispanics.

The key feature of our theory is that it accounts for how the distribution of consumer utility would interact with privacy harms and other costs to determine the *number of people* and hence net benefits brought by privacy protection such as DNC. In particular, we illustrate that even when two populations have the same average demographics and so would appear identical at the aggregate level, the variances of the populations (i.e., heterogeneity) *per se* would affect the resultant demand for DNC. Hence, it is important to analyze the structure of population when devising public policies on privacy. Our empirical analysis provides an initial exploration on the relevant heterogeneity dimensions.

Our theory is sufficiently general and should apply to settings in which the quality of vendors' marketing offers depends on how much they know about consumers, and consumers make cost-benefit tradeoffs when evaluating vendors' offers that give them some benefits but at the same time may impose costs or infringe their privacy. Many real-life applications, including those supported by latest information technologies such as online personalization, email marketing, and reward/loyalty programs, exhibit these characteristics (see, e.g., Moe and Fader 2001; Loder et al. 2006; Cavusoglu et al. 2007; Chellappa and Shivendu 2008; Laudon and Laudon 2008).

Our study makes two important contributions. First, we show, theoretically, that the demand for privacy could either increase or fall with consumer heterogeneity. We generalize our theory to a more general setting to investigate how consumer heterogeneity would affect the demand for a good or service when consumers differ in their valuations for the item, and when they must pay a price for the item. We explore how alternative pricing strategies – skim vis-à-vis penetration pricing – would interact with the distribution of consumer utility to affect demand.

Our second contribution is to empirically test the importance of consumer heterogeneity with a very detailed data-set, in which real consumer choices are observable and can be matched to measures of consumer heterogeneity at a very high degree of granularity. The data-set comprises all registrations of telephone numbers at the county level with the U.S. federal Do-Not-Call (DNC) registry. By registering with the DNC registry, Americans could avoid telemarketing, and hence the DNC registry functioned as a protection against privacy harms from unsolicited telemarketing calls.

Using data from the DNC registry, the U.S. Census, and other sources, we found that the demand for DNC registration increased (the demand for telemarketing decreased) with heterogeneity in ethnicity, education and religion, while the demand for DNC registration decreased (the demand for telemarketing increased) with heterogeneity in age. Our empirical results were robust to variations in measurement and inclusion of key variables, sample specifications, and controlling for possible confounding factors such as social interaction. We further triangulated our findings by investigating how heterogeneity was related to spatial correlations across counties in DNC registrations. In summary, we found robust evidence that consumer heterogeneity (as contrasted with the main effects of average demographics such as ethnicity, education, religion, and age *per se*) had contributed to the observed variations in DNC registration rates and so the demand for telemarketing. The results of our empirical analysis are consistent with the theoretical proposition that consumer heterogeneity matters and that demand for privacy protection could either increase or fall with heterogeneity.

The remainder of this paper is organized as follows. Section 2 presents our theoretical arguments and analysis. Section 3 provides a brief synopsis of the DNC registry. Section 4 describes our empirical model and the data. Section 5 presents the estimates and robustness tests. Section 6 discusses the implications of this research. Finally, Section 7 concludes the paper.

## 2. Theory

Unsolicited solicitations through direct mail, telephone, and email target consumers with information about product, price, and distribution channels. To some consumers, such solicitations are valuable and welcomed. However, unsolicited solicitations affect consumer privacy. Vendors ignore such harms when sending solicitations. Consumers perform a "privacy calculus" to decide if the benefits of receiving solicitations outweigh their privacy costs (Laufer and Wolfe 1977; Dinev and Hart 2006). If the costs exceed the benefits, then consumers may engage in marketing avoidance (Hann et al. 2008; Johnson 2008).

How would new public policies that protect consumer privacy, such as the institution of "no contact" registries, affect consumer welfare when the population comprises different mixes of consumers? In general, vendors take consumer characteristics into account in planning marketing efforts. The ideal of direct marketing is an offer so well-targeted that the consumer accepts it willingly, and so would obtain consumption benefits from the offer.

However, if consumers are heterogeneous, vendors may not be able to identify and target their offers perfectly, and so, vendors may not be able to present every consumer with her ideal offer. Accordingly, any increase in heterogeneity might be associated with a reduction in the benefits that solicitations provide to consumers.[4] In other words, inherent heterogeneity in demographic characteristics could affect the extent to which consumers appreciate marketing solicitations. This could cause the demand for marketing avoidance to increase, because it is now more likely that privacy harms outweigh consumption benefits.

On the other hand, consumers may also differ in terms of sensitivity to privacy (Smith et al. 1996; Hann et al. 2007; Hui et al. 2007). Coupling with their (possibly dispersed) benefits from marketing solicitations, such variations may give rise to differing degrees of heterogeneity in the *net utility* from marketing.

Let $u = b - p - h$ represent a consumer's net utility from marketing solicitations, i.e., the benefits, $b$, less the price, $p$, of the good or service being offered and less the harm, $h$, to privacy, and let $c$ represent the cost of avoiding the solicitations. So, if the consumer does not engage in avoidance, she would receive net utility, $u$, while if she does engage in avoidance, her net utility would be simply $-c$. Then, any consumer for whom $u < -c < 0$, where $c$ is the cost, will engage in marketing avoidance.[5]

Referring to Figure 1, suppose that the solid red curve represents the distribution of net utility, $u$, and let the broken blue curve represent an increase in heterogeneity in the sense

[4] This can be illustrated with the classical Hotelling model. Suppose that, owing to scale economies, a vendor offers only a fixed number of products and it sets prices to maximize profit. If the consumer population is more heterogeneous in the sense of the "transport costs" being higher, then a larger number of marginal consumers would get negative net utility and so drop out of the market.

[5] If $-c < u < 0$, then the consumer receives net negative utility but would not register because the cost of registration is too high. Note that any consumer for whom $u \geq 0$ would not register because the benefits from telemarketing outweigh the costs, including any privacy harms.

of a symmetric mean-preserving spread of the distribution of the net utility. In the left-hand case, the increase in heterogeneity would increase the number of consumers with $u \geq -c$, i.e., reduce the number with $u < -c < 0$ and so reduce avoidance. By contrast, in the right-hand case, the increase in heterogeneity would increase the number of consumers with $u < -c < 0$, and so, increase avoidance.

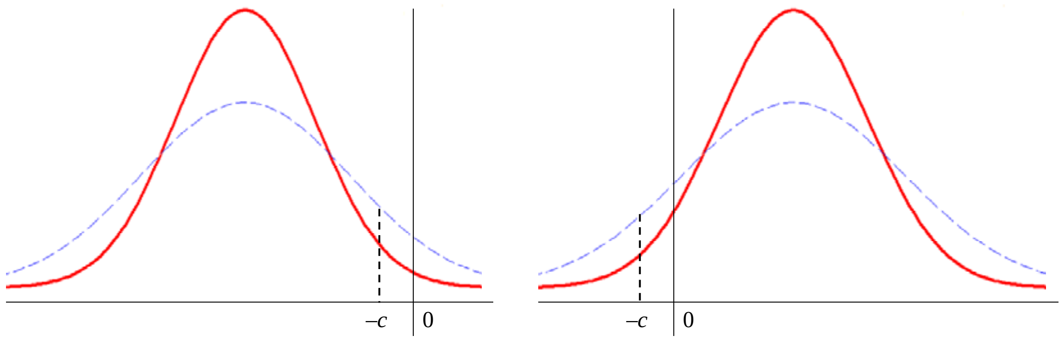


**Figure 1. Heterogeneity in net utility**

Consumer heterogeneity would also have a second-order effect on marketing avoidance due to strategic response of vendors. If low-utility consumers (those with $u < -c < 0$) engage in avoidance and so drop out of the market, then the remaining population would comprise consumers who receive relatively higher net utility from marketing solicitations. This would induce vendors to send out even *more* solicitations (Hann et al. 2008), which would increase the privacy harms on the *remaining* consumers, and cause additional marginal consumers to engage in avoidance. In equilibrium, such strategic interactions between consumers and vendors would lead to a higher overall marketing avoidance.

In particular, referring to Figure 2, the strategic response of vendors would increase solicitations to the remaining consumers, which would shift their net utility to the left. In the left-hand case, the extent of increase in avoidance is lower when consumers are more heterogeneous. By contrast, in the right-hand case, the extent of increase in avoidance is higher when consumers are more heterogeneous.

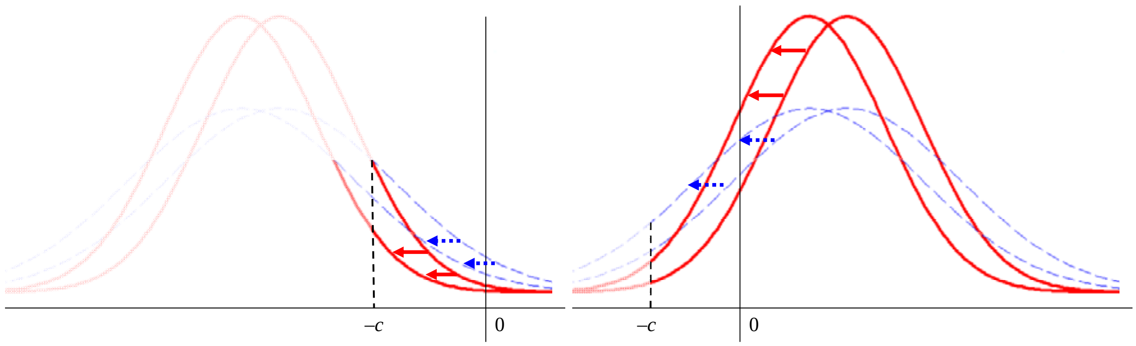


**Figure 2. Vendors' strategic responses**

The above analysis shows that, other than demographic characteristics (Milberg et al. 2000; Varian et al. 2005), the structure and composition of consumer population would also affect the demand for privacy and telemarketing. In summary, our theory provides the following insights: (i) Consumer heterogeneity may limit vendors' ability to target consumers with ideal offers and hence reduce the benefits provided by marketing solicitations. (ii) Depending on the cost of marketing avoidance relative to the distribution of consumer utility, consumer heterogeneity could either increase or decrease the demand for marketing avoidance. (iii) The strategic interaction between vendors and consumers would further intensify the effect of heterogeneity on consumer demand.

With this theoretical analysis, our next challenge is to identify relevant dimensions of consumer heterogeneity, and establish their impact on consumer choice. In the remainder of this paper we address this challenge in the context of the U.S. federal DNC registry.

## 3. Do-Not-Call Registry

The U.S. federal Do-Not-Call registry was established by authority of the *Telemarketing Sales Rule*. With limited exceptions, U.S. federal law prohibits unsolicited telemarketing calls to telephone numbers on the DNC registry. The registry applies to both interstate and intrastate telemarketing calls. It accepts registrations from both fixed-line and mobile but not business telephone numbers. Telemarketers are required to remove numbers on the DNC registry from their call lists no less frequently than every 31 days.[6] Calls for political campaigning and survey research, by nonprofit and charitable organizations, and by businesses with a recent commercial relationship with the consumer are exempted from DNC restrictions. Until February 2008, DNC registrations were effective for five years, but, following passage of the *Do-Not-Call Improvement Act of 2007*, registrations were indefinite.

Prior to the opening of the federal DNC registry in June 2003, 29 states in the U.S. had already implemented state-level "do not call" registries. The telephone numbers on these state registries were subsequently merged into the federal registry.

The DNC registry provides an excellent setting to study how consumer heterogeneity affects the demand for a personalized service, viz. telemarketing:

- It provides a direct, field observation (rather than stated choice, preferences or attitude) of consumer choice between telemarketing and privacy.[7]
- The service is free to consumers, is accessible through the Internet or toll-free telephone number, and there is no competition. These facilitate characterization of the impact of population characteristics on consumer demand without confounding effects due to pricing, adoption barriers, or competitive actions.
- The registry covers the entire United States, and so the distribution of consumer demographics across geographical regions allows us to assess how registration and so the sensitivity to telemarketing varied with the population mix.

[6] Violators are subject to fines of up to $11,000 per offence. As of September 2004 (one year after the registry became effective), the total number of complaints was 579,856. This was less than one percent of the total number of telephone numbers registered at that time (Federal Trade Commission 2005).

[7] In general, survey responses may not accurately reflect consumers' true attitude toward privacy (Berendt et al. 2005). Hui et al. (2007) conducted a field experiment to observe consumer responses to online information requests. Their study focuses more on the secrecy/anonymity but not the seclusion and autonomy aspects of privacy.

## 4. Empirical Model

The theory in Section 2 analyzes how consumer heterogeneity affects DNC registration in a steady state. Hence, we adopted a static approach to the empirical analysis (Hann et al. 2008). An immediate question was when DNC registrations reached equilibrium.

Referring to Figure 3, over time there were multiple waves of registrations, but the registrations tapered off and stabilized after several months. We used the cumulative registration rate in June 2004 (one year after the registry was opened) for the equilibrium analysis because, by that time, the number of further registrations was minimal, and Americans tend to move during the summer (Hansen 1998).[8]

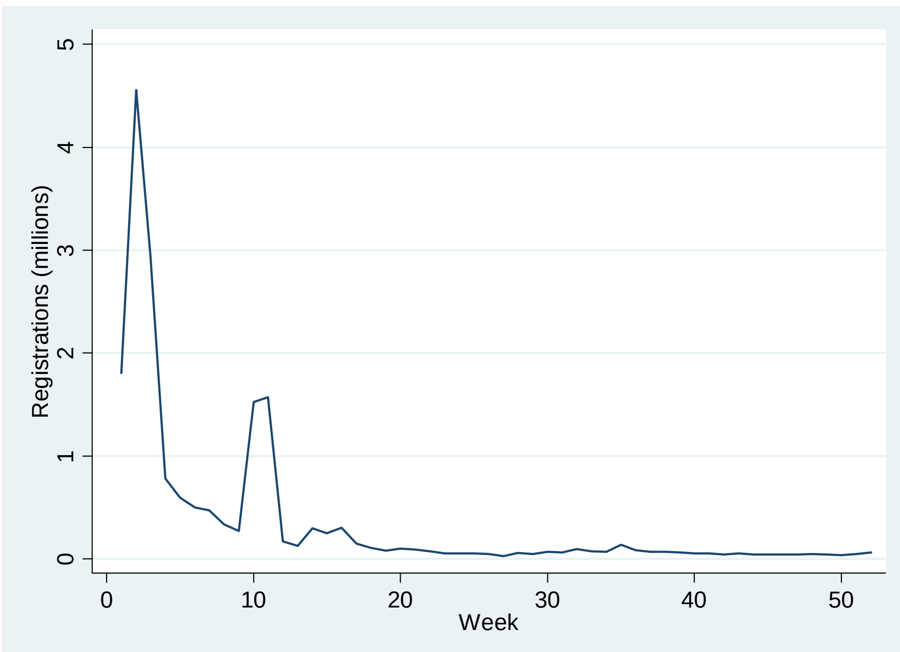


**Figure 3. Weekly registrations in states without state DNC registries**

### 4.1 Specification

We estimated the following equation using least squares regressions:

$$r_{ij} = x_{ij}'\beta + z_{ij}'\alpha + s_j'\gamma + \varepsilon_{ij} , \qquad (1)$$

[8] We shall assess the robustness of our results by using the cumulative registration rate on a different date as the hypothetical equilibrium state.

where $r_{ij}$ is the DNC registration rate in county $i$ in state $j$ at the end of June 2004, $x_{ij}$ is a vector of variables measuring consumer heterogeneity, $z_{ij}$ is a vector of control variables, $s_j$ is a vector of state-specific dummy variables, $\varepsilon_{ij}$ captures county-specific random errors, and $\beta$, $\alpha$, and $\gamma$ are vectors of least squares parameters that we estimated.

Our unit of analysis was the county because households in a county are spatially proximate to each other and because counties lie within state boundaries. By contrast, telephone area codes may cross state boundaries and span households which are spatially disconnected. As already noted, individual states implemented state-level laws and regulations to govern telemarketing. Further, to estimate equation (1), we needed data on the demographics of the population to compute consumer heterogeneities. Such data are available at the county level from the U.S. Census and various other sources (see below).

We included state-specific dummy variables in equation (1) to capture fixed state effects that could affect consumers' registration decisions. These may include the broad legal framework, general intensity of marketing, consumption patterns, and culture.

Finally, other than influences from fixed state effects, consumer registrations may exhibit systematic disparity in variances because the consumers are located in different states with different characteristics. To account for this disparity and to improve the precision of the estimates, we treated the county-level observations as clustered samples (by states) and used robust, clustered standard errors in our estimation (Wooldridge 2002).

### 4.2 Variables

We constructed the dependent variable – DNC registration rate, $r_{ij}$ – by dividing the cumulative number of registrations in county $i$ at the end of June 2004 by the number of households in the same county.[9] Table 1 tabulates the number of registrations and registration rates in different states in the U.S.

<Insert Table 1 here>

We constructed profiles of consumer heterogeneity, $x_{ij}$, on seven dimensions – income, race, ethnicity, age, gender, education, and religion. Data on these characteristics are widely available, and vendors may use them to assist in designing direct marketing campaigns. Heterogeneity on these dimensions had been identified as important in social integration (Alesina and Ferrara 2000; McPherson et al. 2001; Marmaros and Sacerdote 2006) and new product adoption (Dekimpe et al. 2000; Talukdar et al. 2002; Van den Bulte and Stremersch 2004).

We computed heterogeneity of race, ethnicity, gender, education, and religion as the probability that any two individuals drawn at random from a county would not belong to the same demographic group.[10] We computed age and income heterogeneity using the Gini coefficient (Van den Bulte and Stremersch 2004).

Finally, we used a set of demographic variables, $z_{ij}$, to control for variations in DNC registration due to consumers' intrinsic preferences or needs for telemarketing. These include retail density,[11] median household income,[12] average household size,[13] commuting

---

[9] Note that we excluded all mobile phone registrations because there is no mandatory coupling of mobile phone numbers to the county of residence for a consumer, and so we were not able to determine the demographic characteristics of these registrants.

[10] Technically, the heterogeneity of consumer attribute $k$ in county $i$ is computed as

$$H_{ik} = 1 - \sum_{m \in M_k} s_{im}^2 ,$$

where $s_{im}$ is the percentage share of consumers in group $m$ of attribute $k$ (which comprises $M_k$ different groups; for example, male and female for gender).

[11] Consumers differ in the demand for shopping, and vendors fulfill these needs by setting up the corresponding retail channels. In general, we would expect to see more retail stores in regions where consumers make more purchases. Hence, the density of retail stores may serve as a good proxy for consumers' shopping needs. Retail density could have a negative influence on DNC registration, as people who have a higher shopping need, as exemplified by having more retail stores around them, may be more receptive of telemarketing calls (which is another channel to fulfill the shopping need),

time, and unemployment rate.[14] We selected this set of control variables largely based on cost-benefit analysis of telemarketing, that it provides convenience but may impose harms on consumer privacy (Milne and Gordon 1993; Hann et al. 2007).

### 4.3 Data

We assembled a rich data set that was culled from multiple sources. The FTC provided us with redacted DNC registrations by exchange on a daily basis from June 27, 2003 to January 6, 2006. These records showed registrations by redacted telephone number for each area code and exchange (called NPA-NXX), for example, (617) 363-xxxx, by date of registration. While the U.S. comprised 3,128 counties in our data set, it comprised 94,342 exchanges. On average, each county encompassed about 47 exchanges, but there was substantial variation in the number of exchanges per county.

To proceed, we needed to match the registrations with data on consumer and other characteristics that were available only at the county level. We procured the *North American Local Exchange NPA-NXX Database* (NALENND) from *Quentin Sager Consulting*. Using the NALENND database, we identified the counties served by each telephone exchange, and

---

and so may less likely register for DNC. On the other hand, a higher retail density could provide a better substitute for telemarketing, and hence increase DNC registration.

[12] Household income could influence DNC registration in two directions. On one hand, people whose income is higher may incur a higher opportunity cost of time (Pashigian and Bowen 1994; Aguiar and Hurst 2007), and so may appreciate the convenience provided by telemarketing. On the other hand, receiving telemarketing calls itself may be more harmful for these high income people, and so they may appreciate the DNC registry as well. On balance, the impact of household income on DNC registration is ambiguous.

[13] The likelihood that each person picks up a telemarketing call is lower in a larger household than a smaller one. Hence, the average harm per person due to telemarketing should decrease with household size, and so a larger household would be *less likely* to register for DNC.

[14] Similar to high income people, people who are employed or who need a longer commuting time may have a higher opportunity cost of time as well, and so may tend to appreciate the convenience provided by telemarketing. However, these people would spend relatively less time at home, and so may suffer less harm from telemarketing. Accordingly, their need for DNC may be lower. We expected unemployment rate to *increase* and commuting time to *decrease* DNC registrations.

so were able to map the DNC registrations to the respective counties.[15] We summed the individual DNC registrations up to the end of June 2004 to obtain the cumulative numbers of registrations in each county.[16]

We obtained the following county-level data from the U.S. Census 2000: median household income, education, gender, age, race, ethnicity, average number of people in a household (household size), average commuting time, unemployment rate, number of retail stores, and land area (the later two were used to compute retail density). We obtained religious affiliations of the population in each county from the *Association of Religion Data Archives*. Tables 2 and 3 report summary statistics and correlations of our variables.

<Insert Tables 2 and 3 here>

**5. Results**

We first estimated equation (1) by including only the control and state-specific dummy variables. The results are reported in Table 4, column 1.[17] Most of the control variables had the expected effects on DNC registration – the coefficients of retail density, household size, and commuting time were negative and significant. This is generally consistent with the view that consumer privacy and the demand for public policy (in this context, DNC registry) depend on the benefits that a person receives from telemarketing and her likelihood of receiving marketing solicitations.

[15] Some telephone exchanges spanned multiple counties. For these exchanges, we allocated the DNC registrations within the exchange to the respective counties according to the relative number of households in the counties as reported by the U.S. Census 2000.

[16] On reviewing the data, we identified three discrepancies. Registrations for two counties in Texas – Kenedy (FIPS 48261) and Loving (FIPS 48301) – were less than one each. Williamsburg, VA (FIPS 51830) had just 3,619 households but a massive 17,127 DNC registrations, or an average of 4.73 per household, which was 24.89 standard deviations higher than the mean registration rate of 0.39 per household. The county with the next highest registration was Pitkin, CO (FIPS 8097), with just 1.62 per household. Accordingly, we omitted Kenedy and Loving, TX, and Williamsburg, VA, from the subsequent analysis.

[17] For brevity, we do not report the fixed state effects in all tables. Most of the fixed state effects – more than 40 out of 50 in most of the estimations – were statistically significant. This indicates that some latent state-specific factors, such as regulation, culture, shopping habit, etc., might have systematically affected consumers' inclinations to register for DNC.

<Insert Table 4 here>

The coefficient of household income was positive and significant, which implies that for high-income people the costs of receiving telemarketing calls may outweigh the convenience provided by them. Contrary to our prediction, the coefficient of unemployment rate was negative, but it was insignificant.

The next seven regressions in Table 4 tested our main research question – whether consumer heterogeneity affected DNC registrations. Among the seven heterogeneity dimensions that we used, three were significant – heterogeneity in ethnicity (Hispanic vs. others) and religion had positive influences, and age heterogeneity had a negative influence on registration.

The result on age heterogeneity implies that in regions where people are more spread out in terms of age, their net utility toward direct marketing may skew more toward the positive side, and so they may have lower demand for privacy protection. By contrast, if we consider the ethnic or religious profiles of the population, then the more heterogeneous communities may tend to have net utility skewed toward the negative side, and so may demand for more privacy protection. Figure 4 depicts distributions that could possibly give rise to these results.

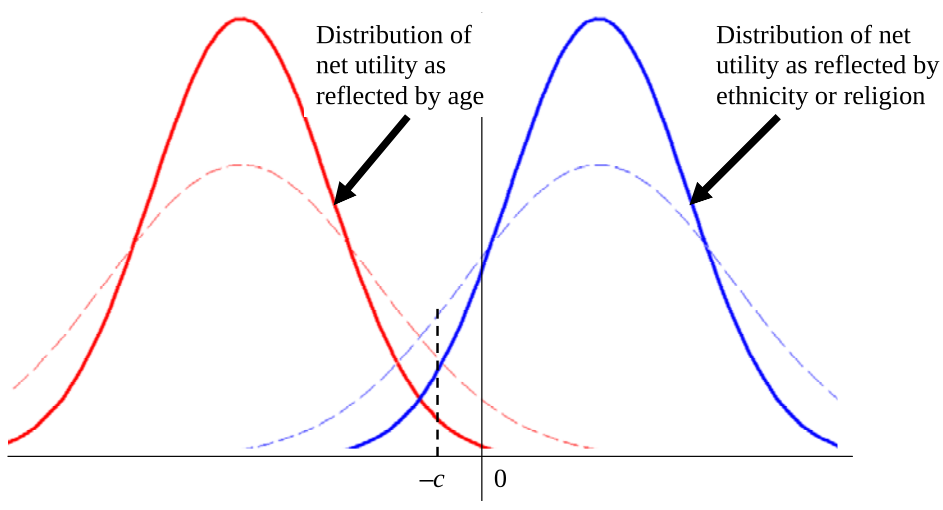


**Figure 4. Illustrative scenario**

To gauge the managerial importance of these findings, we calculated that a one standard deviation increase in age heterogeneity would *decrease* the registration rate by $0.959 \times 0.02 = 0.01918$, whereas a one standard deviation increase in religion heterogeneity would *increase* the registration rate by $0.083 \times 0.10 = 0.0083$. Relative to the overall mean registration rate of 0.39, these changes were about 4.92 and 2.13 percent of the mean, hence economically significant.

Finally, the last regression in Table 4 included all the heterogeneity and control variables. The results were largely consistent – almost all control variables had the same signs and significance.[18] Age and religion heterogeneity significantly affected DNC registrations. Ethnic heterogeneity was marginally significant ($p < 0.10$). Interestingly, the coefficient of education heterogeneity was positive and marginally significant ($p < 0.10$). Since this specification provided an integrated view of all dimensions of heterogeneity, we considered this to be our preferred specification.

**5.1 Robustness**

We conducted a number of robustness tests, all using the preferred specification. First, we used the registration rates as of the 28th week, which included the year end, December 31, 2003, as the equilibrium state. Referring to Figure 3, by the 28th week, the two major waves of DNC registration were concluded. As reported in Table 5, column 1, the estimates were very close to those that we obtained using data at the 52nd week. Hence, our results were not sensitive to choices of equilibrium date.

<Insert Table 5 here>

Second, we excluded the 29 states with state-level DNC registries that were established before the federal registry. States which established state-level registries were

[18] The coefficient of unemployment rate became positive, which was consistent with our expectation. However, it continued to be insignificant.

likely to be responding to relatively greater concern about privacy among their voters. Table 5, column 2 reports the estimation results. In general, the results were similar, but with weaker statistical significance. This was consistent with the reasoning that voters in the states without state-level DNC registries were less concerned about privacy.

Third, instead of excluding the three discrepant observations *ex ante*, we applied the robust regression method, which iteratively weighted the observations and dropped observations from the analysis using the regression diagnostics (Hamilton 1991).[19] The results are reported in Table 5, column 3, and they were similar to those obtained by directly excluding the discrepant observations.

Fourth, we estimated another specification in which the registration rates were computed as total registration counts divided by the number of residential telephone lines rather than the number of households.[20] As shown in Table 5, column 4, the fit was substantially worse – $R^2$ of 0.288 as compared with 0.508 from the estimates with registration rates computed using number of households (Table 4, last column).

The signs of the coefficients were largely the same as in our previous estimates, but the significance of the variables differed considerably. The poor fit and results were perhaps not surprising, as the decision about DNC registration is more likely made at the household level rather than by telephone line. Further, we had to estimate the number of telephone lines in the counties, which could have introduced additional measurement errors.[21]

---

[19] This was implemented using STATA 10's robust regression. Note that the robust regression method does not adjust the error variance by clustering the observations. Hence, this specification also serves as a robustness test for using a non-clustered error structure.

[20] We obtained data on the percentages of telephone numbers in use for each area code from Stroup and Vu (2003, Table 6), and data on the percentages of phone numbers associated with residential subscribers for each state from the Federal Communications Commission (FCC). On the assumption that that these percentages were uniform across all exchanges within area codes and all counties within states, we computed the number of residential telephone lines in each county.

[21] In another robustness test, we checked the impact of an unintended mistake in the FTC's implementation of the DNC registry. For the first 10 days, residents of states west of the Mississippi, Minnesota and Louisiana could register by phone and through the Internet, while residents of other

### 5.2 Alternative Explanation

Previous research has suggested that social interaction would be greater in communities that are less heterogeneous (Alesina and Ferrara 2000; Marmaros and Sacerdote 2006) and that social interaction influences consumer behavior (Dekimpe et al. 1998; Talukdar et al. 2002; Van den Bulte and Stremersch 2004). Could the significance of our heterogeneity variables be spurious because they actually reflected the impact of social interaction?

To address this possibility, we acquired data on social interaction from the *Social Capital Community Benchmark Survey*, which was administered by the Saguaro Seminar at Harvard University.[22] The Saguaro Seminar compiled a measure of informal social interaction, SCHMOOZ, which was constructed by principal components analysis of variables measuring the frequencies of visiting friends and relatives, and some other social gatherings. The same data set and variable was used in research into the impact of social interaction on the digital divide (Agarwal et al. 2009).

We re-estimated the preferred specification by adding SCHMOOZ as one of the independent variables to control for the effect of social interaction. The results are reported in Table 5, column 5. Social interaction had a positive effect on DNC registration, but the effect was not statistically significant. Most of the other coefficients had largely the same signs as the previous estimations, but their significance became weaker. This could be due to

---

states could only register through the Internet (Varian et al. 2004). We added a variable, WEST, to indicate the states west of the Mississippi, Minnesota and Louisiana. With the addition of this indicator variable, we could not add the fixed state effects into the estimation. The regression exhibited poorer fit, which is not surprising since WEST is much more restrictive than the state dummy variables. The coefficients of the control variables had the same signs and significance. The heterogeneity variables had mostly the same signs, but with much weaker significance. For brevity, we do not report the results of this estimation.

[22] For details of the Saguaro Seminar and the *Social Capital Community Benchmark Survey*, see http://www.hks.harvard.edu/saguaro/. We used data from the 2000 survey, which contained roughly 3,000 responses from a national sample and an additional 26,200 responses from 40 communities across 29 states in the U.S. The data spanned across roughly 1,323 counties that we included in this research. The purpose of the *Social Capital Community Benchmark Survey* was to measure civic engagement and social capital within U.S. communities.

the use of smaller and incomplete samples (data on SCHMOOZ were available only in a subset of the studied counties).

In any case, the coefficient of age heterogeneity continued to be negative and significant. The coefficients of religion and education heterogeneity were still positive, and their *p*-values were small and close to being significant. Ethnic heterogeneity had a positive coefficient, but it was insignificant. Although the addition of the social interaction variable considerably weakened the influence of the heterogeneity variables, some of them continued to play a significant role in affecting DNC registrations.

### 5.3 Further Evidence

In Section 2, we suggested that vendors may strategically respond to consumers' DNC registrations – when low-utility consumers register and drop out of the market, vendors may respond by sending more solicitations to the remaining consumers, which may cause some marginal consumers to register as well (Hann et al. 2008). In equilibrium, such strategic interactions may cause DNC registrations to exhibit a correlated pattern in the local community. If consumer heterogeneity had affected DNC registration through this second-order interaction effect (see the discussion around Figure 2), then DNC registrations in adjacent areas should be correlated.

To assess this possibility, we calculated spatial autocorrelations in DNC registrations among telephone exchanges within counties. Specifically, for each county, we calculated the registration rates in each constituent telephone exchange, and then computed the Moran *I* statistic, which measured the spatial autocorrelation of DNC registration in the exchanges within the same county.[23] Of the 3,128 counties in our data set, 3,060 contained between 2 and 2,932 telephone exchanges, with an average of 46.49 (±145.9).

[23] The Moran *I* statistic is commonly used to measure spatial autocorrelation. It is defined as

For each county, we then regressed the Moran *I* statistic as a dependent variable on the demographic variables including the various measures of heterogeneity. The estimation results are reported in Table 5, column 6.[24] Most of the variables that affected DNC registration played a significant role in affecting spatial correlations of registrations. Most importantly, the four dimensions of heterogeneity – age, ethnicity, education and religion – had significant influences on the spatial autocorrelations.[25] The same four variables were variously significant in the previous estimations with DNC registration as the dependent variable. Hence, the spatial autocorrelation analysis provided further evidence to support the importance of accounting for consumer heterogeneity in assessing the demand for privacy and telemarketing.

## 6. Discussion and Implications

Our empirical results show that, in the context of privacy and telemarketing, it is important to account for consumer heterogeneity in estimating demand. In particular, our theory and the illustrations in Figures 1 and 2 supposed that the means of the consumer distributions were identical. The distributions differed only in heterogeneity. Hence, any demand model that uses aggregate, mean-level data on consumer characteristics would fail to identify the type of

---

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2},$$

where $n$ is the number of telephone exchanges in a county, $x_i$ is the number of DNC registration in exchange $i$, and $w_{ij}$ is the distance between exchanges $i$ and $j$, computed from the actual geographic distance derived from the exchanges' latitudes and longitudes (Banerjee et al. 2004).

[24] We had fewer than 3,078 observations in this estimation because Moran *I* can only be computed when there are multiple neighboring exchanges within the same county. Some counties had only one telephone exchange serving the entire county.

[25] Note that although the sign of age heterogeneity was positive instead of negative as in the previous estimations, it should not be considered as an inconsistent finding because, as we illustrated in Figure 2, consumer heterogeneity could have a different impact on the magnitude of changes in consumer registration due to the second-order strategic interaction effect, and so could shift spatial autocorrelation in different directions. In other words, the coefficients in the spatial autocorrelation regression and the DNC registration regressions need not have the same signs. Our focus in this spatial autocorrelation regression is the *significance* of the heterogeneity variables.

effects that we have identified.[26] It is important to use information on the heterogeneity of population *directly* in the estimations.[27]

Our empirical results in the context of DNC registration and telemarketing are consistent with our general theory that the demand for privacy could vary with consumer heterogeneity. While Figures 1, 2 and 4 merely depict scenarios to illustrate our theory and do not purport to represent any actual distributions of consumers or any specific directions of the impact of heterogeneity, our empirical findings clearly show that heterogeneity does matter.

Note that our theory was founded on the privacy calculus framework (Laufer and Wolfe 1977). To the extent that this calculus applies to all settings involving privacy tradeoffs, the following managerial and policy implications should have broader significance for marketing and consumer privacy.

First, while it is common for direct marketing vendors to segment consumers based on the benefits provided by their goods and services, direct marketing solicitations in all forms, not just telemarketing, may be intrusive and harm consumer privacy (Petty 2000). The harm may be especially pronounced with the technology-enabled solicitations (for example, advances in database marketing and email technologies have hugely facilitated the use of spam for solicitations). It is important for direct marketers to recognize the extent of privacy harms that they impose on different consumers, and account for the harms when devising

[26] Empirical demand models that use aggregate data, such as those in the "New Empirical Industrial Organization" (NEIO) framework (e.g., Berry et al. 1995; Besanko et al. 1998), have attempted to account for consumer heterogeneity by imposing assumptions and structures (e.g., multivariate normality) on the consumer tastes for unobservable product characteristics. As our data and estimates show, however, the structures of consumer population could vary across regions, and could be approximated by different demographic characteristics. Hence, it is unlikely for any specific assumptions on consumer tastes or error structures to provide a sufficient fit to data that span wide geographical areas.

[27] The use of aggregate, mean-level data from the U.S. Census and comparable international sources on population, income, piracy rates, etc., is common in empirical information systems research (see, e.g., Gopal and Sanders 2000; Varian et al. 2005; Levina and Xin 2007; Agarwal et al. 2009). We do not know of any study that explicitly accounts for heterogeneities in the population.

segmentation strategies. The control variables that we used in the regressions may potentially serve as such "privacy segmentation" attributes. Further, our empirical results imply that segmentation by sensitivity to privacy might be especially important among consumers who are heterogeneous in ethnicity, education, or religion, and homogeneous in age.

Second, with regard to public policy, our results imply that communities where people are similar in age but differ in ethnicity, education, or religion would be more sensitive to direct marketing. Hence, to the extent that vendors ignore this relative sensitivity, it is less likely that self-regulation or a market in information would sufficiently protect consumer privacy. In this case, government intervention may be necessary.[28]

It is worth emphasizing that our study relied mostly on U.S. Census data, which are also available to direct marketers. If the telemarketing industry had fine-tuned their offers to the extent possible with census information, then we should not have found any significant impacts among the control or heterogeneity variables. Evidently, the industry did not sufficiently account for the characteristics and structures of the population in devising their marketing campaigns, which thereby contributed to the massive DNC registration.

### 6.1 Pricing

With slight modifications, our theory applies generally to any scenario in which goods and services are offered at a particular price to heterogeneous consumers even when there is no issue of privacy. Suppose that a consumer derives benefits *b* from some item which is sold at price, *p*, and the consumer derives zero benefit from not buying. Then, the consumer will buy the item if her net utility $u = b - p \geq 0$, and not buy if $u < 0$.[29]

---

[28] For example, both U.K. and U.S. regulations disallow firms from marketing to consumers using pre-recorded telephone messages unless consumers opt in to the service.

[29] This is a simple modification of the theory in Section 2. Absent any issue of privacy, the privacy harm, $h = 0$ and the cost of marketing avoidance, $c = 0$, and hence the net utility simplifies to $u = b - p - h = b - p$.

To see how heterogeneity matters in such a setting, refer to Figure 5, where now the curves depict the distribution of consumer benefits, *b*, from the item, and the price, *p*. The intersections between the benefit curves and the price would determine the number of consumers who buy the item.

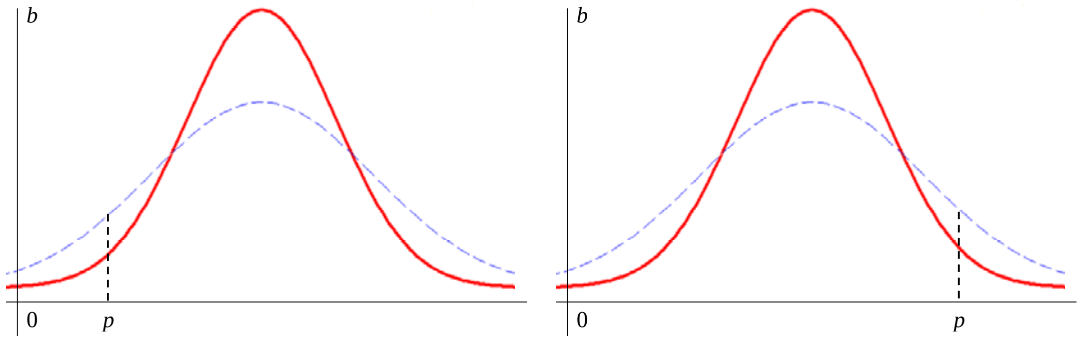


**Figure 5. Consumer heterogeneity and pricing**

In the left-hand case, when the price is low (a situation of penetration pricing), an increase in heterogeneity would increase the area under the benefit curve to the left of *p* and so reduce the demand for the item. By contrast, in the right-hand case, when the price is high (a situation of skim pricing), then an increase in heterogeneity would raise demand. Accordingly, consumer heterogeneity has a profound impact on the effectiveness of any particular pricing strategy.

As the marginal cost of producing information goods is very low, vendors of information goods have relatively wide discretion in pricing, and hence, pricing is a particularly important function (Whinston et al. 1997; Shapiro and Varian 1999). To our knowledge, we are the first to highlight the significance of consumer heterogeneity to the choice between penetration and skim pricing of information goods.

### 6.2 Personalization

The widespread use of the Internet and advances in information technology have enabled vendors to collect detailed information about consumers and analyze this information to increase personalization and better target personalized offers (Cavusoglu et al. 2007; Chellappa and Shivendu 2008). The challenge of personalization, however, has always been how to identify relevant consumer characteristics and data.

Our theory and empirical analysis further suggest that if aggregate data are used to select or fine-tune the potential bases of personalization, the vendor would need to account for one additional source of influence, viz. consumer heterogeneity. Also, heterogeneity can be defined along multiple dimensions, and our research has shown that different measures of heterogeneity (e.g., age inequality vis-à-vis religion heterogeneity) may affect demand in opposite directions. While we found robust evidence that heterogeneities in ethnicity, age, education, and religion mattered in the context of telemarketing, it is possible that heterogeneity on other dimensions may be relevant to the demand for other applications or products/services. Hence, it is important for any personalization efforts that use aggregate consumer data to scrutinize the structure of the consumer population.

## 7. Conclusions

This study proposes a new theory that consumer heterogeneity affects demand in a systematic way. We applied this theory to a context of consumer privacy, where consumers suffer harm from marketing solicitations and can choose to avoid marketing by incurring some avoidance cost. We then subjected the theory to empirical tests using county-level U.S. federal DNC registration data, and found that consumer heterogeneity in age, ethnicity, education, and religion consistently provided good explanatory power on observed DNC registration rates. Our empirical results were robust to measurement and inclusion of key variables, sample specifications, and controlling for possible confounding factors.

Our empirical results validated the theoretical proposition that consumer heterogeneity matters in consumer demand. In the context of consumer privacy, our theory and empirical findings provide specific managerial guidance on segmentation and implications for public policy toward protection of privacy. Applied to the pricing of information goods, our theory provides specific managerial guidance on the issue of skim vis-à-vis penetration pricing. Finally, our theory offers a new perspective on personalization – other than consumer characteristics *per se*, any personalization based on aggregate data should account for the relevant heterogeneity as well.

The immediate and important direction for future work is a deeper analysis of the economic, social, and cultural factors underlying consumer heterogeneity and the implications of the various dimensions of consumer heterogeneity for demand. With data on consumer choice and heterogeneity at suitable granularity, this would set the stage for further empirical testing of the impact of consumer heterogeneity on demand.

**References**


Agarwal, R., A. Animesh, and K. Prasad, "Social Interactions and the "Digital Divide": Explaining Variations in Internet Use," *Information Systems Research*, Vol. 20 No. 2, June 2009, 277-294.

Aguiar, Mark, and Erik Hurst, "Life-Cycle Prices and Production", *American Economic Review*, Vol. 97 No. 5, December 2007, 1533-1559.

Alesina, Alberto, and Eliana La Ferrara, "Participation in Heterogeneous Communities," *Quarterly Journal of Economics*, Vol. 115, No. 3, 2000, 847-904.

Banerjee, S., B.P. Carlin, and A.E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman and Hall/CRC Press, 2004.

Berendt, B., O. Günther and S. Spiekermann, "Privacy in E-Commerce: Stated Preferences vs. Actual Behavior," *Communications of the ACM*, Vol. 48, No. 4, 2005, 101-106.

Berry, S.T., J. Levinsohn, and A. Pakes, "Automobile prices in market equilibrium," *Econometrica*, Vol. 63, No. 4, 1995, 841-890.

Besanko, D., S. Gupta, and D.C. Jain, "Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework," *Management Science*, Vol. 44, No. 11, 1998, 1533-1547.

Cavusoglu, H., H. Cavusoglu, and S. Raghunathan, "Selecting A Customization Strategy under Competition: Mass Customization, Targeted Mass Customization and Product Proliferation," *IEEE Transactions on Engineering Management*, Vol. 54, No. 1, February 2007, pp. 12-28

Chellappa, R.K. and S. Shivendu, "An Economic Model of Privacy: A Property Rights Approach to Regulatory Choices for Online Personalization," *Journal of Management Information Systems*, Vol. 24 No. 3, Winter 2008, 193-225.

Culnan, M. J. and K. P. Armstrong, "Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation," *Organization Science*, Vol. 10, No. 1, January-February 1999, 104-115.

Culnan, M. J. and J. R. Bies, "Consumer Privacy: Balancing Economic and Justice Considerations," *Journal of Social Issues*, Vol. 59, No. 2, 2003, 323-342.

Dekimpe, Marnik G., Philip M. Parker and Miklos Sarvary, "Globalization: Modeling Technology Adoption Timing Across Countries", *Technological Forecasting and Social Change*, Vol. 63, 2000, 25-42.

Dinev, Tamara, and Paul Hart, "Extended Privacy Calculus Model for E-Commerce Transactions", *Information Systems Research*, Vol. 17, No. 1, March 2006, 61-80.

Federal Trade Commission. *Annual Report to Congress for FY 2003 and 2004 Pursuant to the Do Not Call Implementation Act on Implementation of the National Do Not Call Registry*. 2005.

Gopal, R.D. and G.L. Sanders "Global Software Piracy: You Can't Get Blood out of a Turnip," *Communications of the ACM*, Vol. 43 No. 9, September 2000, 82-89.

Hamilton, L.C. *Regression with Graphics: A Second Course in Applied Statistics*. Duxbury Press, 1991.

Hann, I.H., K.L. Hui, T.S.Y. Lee, and I.P.L. Png, "Overcoming Online Information Privacy Concerns: An Information Processing Theory Approach," *Journal of Management Information Systems*, Vol. 42, No. 2, 2007, 13-42.

Hann, I.H., K.L. Hui, T.S.Y. Lee, and I.P.L. Png, "Consumer Privacy and Marketing Avoidance: A Static Model", *Management Science*, Vol. 54 No. 6, June 2008, 1094-1103.

Hirshleifer, J., "Privacy: Its Origin, Function, and Future," *Journal of Legal Studies*, Vol. 9, No. 4, 1980, 649-664.

Hui, K.L., H.H. Teo, and T.S.Y. Lee, "The Value of Privacy Assurance: An Exploratory Field Experiment," *MIS Quarterly*, Vol. 31, No. 1, 2007, 19-33.

Hui, K.L. and I.P.L. Png, "The Economics of Privacy," in *Economics and Information Systems, Handbooks in Information Systems*, Vol. 1, Chapter 9, ed. Terrence Hendershott, Elsevier, 2006.

Johnson, J.P., "Targeted Advertising and Advertising Avoidance," Johnson Graduate School of Management, Cornell University, November 7, 2008.

Laufer, R. S. and M. Wolfe, "Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory," *Journal of Social Issues*, Vol. 33, No. 3, 1977, pp. 22-42.

Laudon, K.C., "Markets and Privacy," *Communications of the ACM*, Vol. 39, No. 9, 1996, 92-104.

Laudon, K. and J. Laudon. *Management Information Systems*. Prentice Hall: Upper Saddle River, NJ, 2008.

Levina, N. and M. Xin "Comparing IT Workers' Compensation Across Country Contexts: Demographic, Human Capital, and Institutional Factors," *Information Systems Research*, Vol. 18 No. 2, June 2007, 193-210.

Loder, T., M. Van Alstyne and R. Wash "An Economic Response to Unsolicited Communication," *Advances in Economic Analysis & Policy*, Vol. 6 No. 1, 2006, Article 2.

Marmaros, David and Bruce Sacerdote, "How Do Friendships Form," *Quarterly Journal of Economics*, Vol. 121, No. 1, 2006, 79-119.

McPherson, Miller, Lynn Smith-Lovin and James M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, Vol. 27, 2001, 415-444.

Milberg, S. J.; Smith, H. J.; and Burke, S. J., "Information privacy: corporate management and national regulation," *Organization Science,* Vol. 11 No. 1, 2000, 35-57.

Milne, G. R. and Gordon, M. E., "Direct mail privacy-efficiency trade-offs within an implied social contract," *Journal of Public Policy and Marketing*, Vol. 12 No. 2, Fall 1993, 206-215.

Moe, Wendy W. and Peter S. Fader, "Uncovering Patterns in Cybershopping," *California Management Review*, Vol. 43 No. 4, 2001, 106-117.

Pashigian, B.P. and B. Bowen, "The Rising Cost of Time of Females, the Growth of National Brands and the Supply of Retail Services," *Economic Inquiry*, Vol. 32, No. 1, 1994, 33-65.

Petty, R.D. "Marketing Without Consent: Consumer Choice and Costs, Privacy, and Public Policy," *Journal of Public Policy and Marketing*, Vol. 19, 2000, 42-53.

Posner, R.A., “The Economics of Privacy,” *American Economic Review*, Vol. 71, No. 2, 1981, 405-409.

Talukdar, Debabrata, K. Sudhir, and Andrew Ainslie, “Investigating New Product Diffusion Across Products and Countries”, *Marketing Science*, Vol. 21, No. 1, Winter 2002, 97–114.

Smith, H. J., J. S. Milberg, and J. S. Burke, “Information Privacy: Measuring Individuals’ Concerns about Organizational Practices,” *MIS Quarterly*, Vol. 20, No. 2, June 1996, 167-196.

Stroup C. and J. Vu, *Numbering Resource Utilization in the United States as of June 30, 2003*. Federal Communications Commission Report, Industry Analysis and Technology Division, Wireline Competition Bureau, December 2003.

Van den Bulte, Christophe, and Stefan Stremersch, “Social Contagion and Income Heterogeneity in New Product Diffusion: A Meta-analytic Test”, *Marketing Science*, Vol. 23 No. 4, Fall 2004, 530–544.

Varian, Hal, Fredrik Wallenberg, and Glenn Woroch, “Who Signed Up for the Do-Not-Call List?” School of Information, University of California, Berkeley, June 15, 2004.

Varian, Hal, Fredrik Wallenberg, and Glenn Woroch, “The Demographics of the Do-Not-Call List”, *IEEE Security & Privacy*, Vol. 3, January/February 2005, 34-39.

Westin, A.F. *Privacy and Freedom*. New York, NY: Atheneum, 1967.

Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA, 2002.