Of Frogs & Herds: The Economics (and Behavioral Economics) of Privacy

Alessandro Acquisti
Heinz College & CyLab
Carnegie Mellon University

CyLab Research Briefing
March 8, 2010

Agenda

1. From the Economics of Privacy…

2. ... to the Behavioral Economics of Privacy

3. … to Soft Paternalism

… and how this relates to Information Security

1. The Economics of Privacy

The Economics of Privacy

- *Protection & revelation of personal data flows involve tangible and intangible trade-offs for the data subject as well as the potential data holder*
- Early 1980s
  - The Chicago school approach (Posner 1978, Stigler 1980, …)
- Mid 1990s
  - IT explosion (Varian 1996, Noam 1996, Laudon 1996, …)
- After 2000
  - Formal microeconomic models (Acquisti & Varian 2001, Taylor 2001, Calzolari & Pavan 2001, Katz & Hermalin 2003,…)

The economics of privacy

- Some of our studies in this area
  - Modeling
    - Conditioning prices on purchase histories (*Marketing Science 2005*)…
  - Empirical
    - Impact of breaches on stock market valuation (*ICIS 2006*)…
    - Impact of data breach notification laws on identity theft (*WEIS 2008*)…
    - Impact of gun owners DB publication on crime (*work in progress)…*
    - Impact of Facebook profiles on firms' hiring behavior (*work in progress)…*

However: Privacy Attitudes vs. Behavior

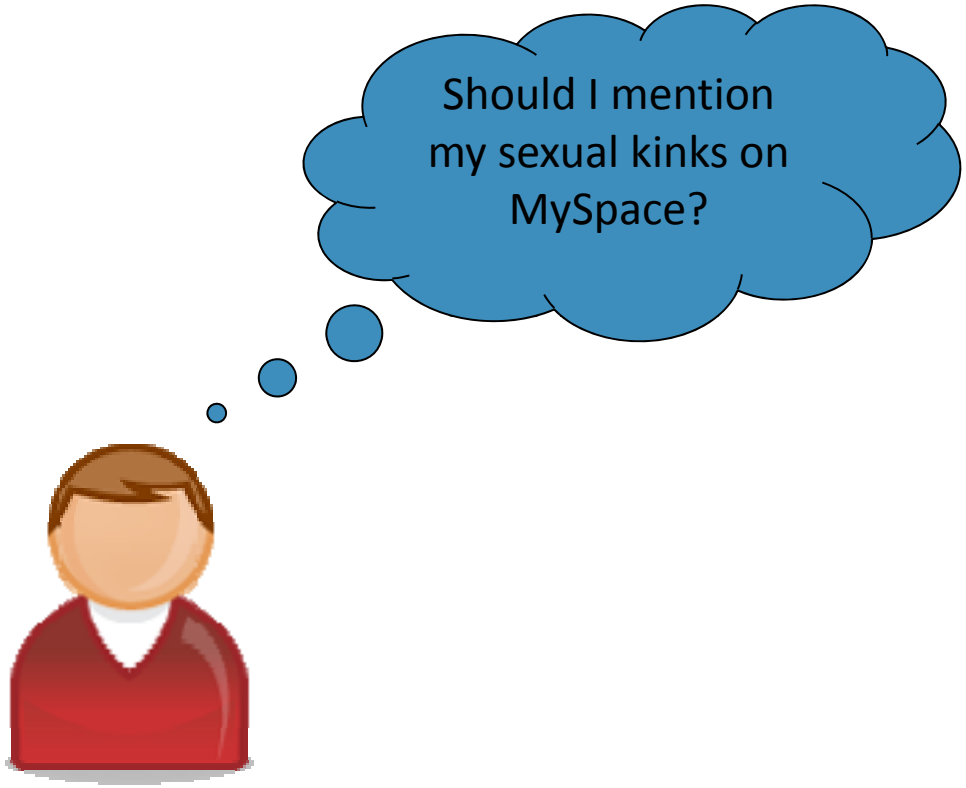
- Attitudes about privacy
  - (Ostensibly,) top reason for not going online… (Harris Interactive)
  - Billions in lost e-tail sales… (Jupiter Research)
  - Significant reason for Internet users to avoid Ecommerce… (P&AB)
- Actual behavior
  - Dichotomy between privacy attitudes and privacy behavior
    - Spiekermann et al. 2001, Acquisti & Gross 2006's Facebook study

Do people really care for privacy?

If they do, can they act on their concerns?

If they don't (or can't), should policy-makers do so on their behalf?

A Rational Model of Privacy Decision Making

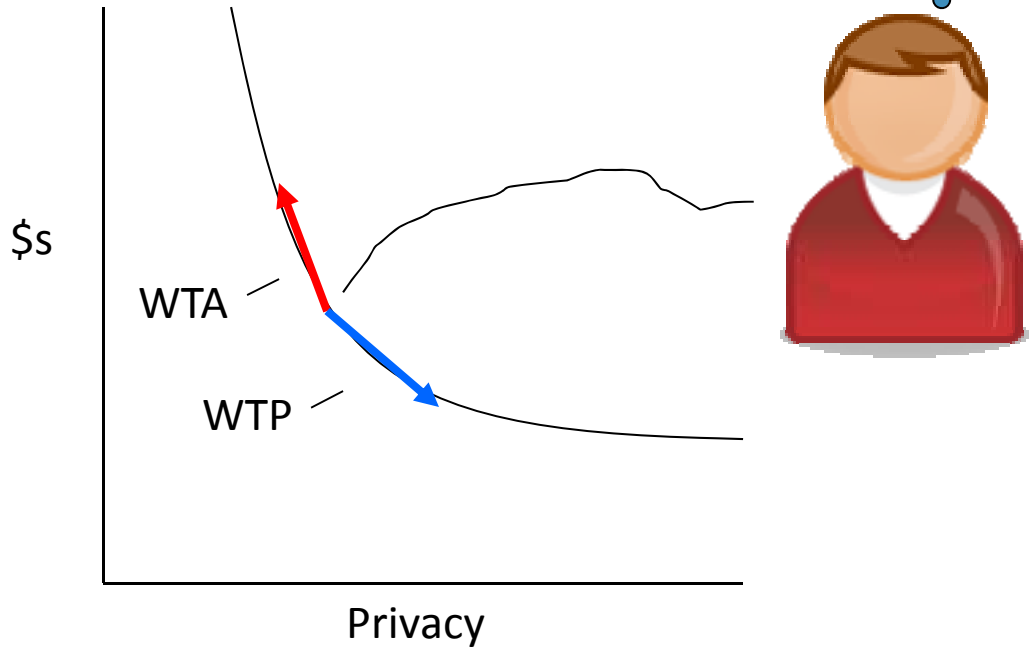


A Rational Model of Privacy Decision Making

Maybe I'll find a lover... But what about my future job prospects? And what if my parents happen to log on...

A Rational Model of Privacy Decision Making

$$\sum p_i \sum \frac{1}{(1+d)^t} u(benefits_{it}) - \sum q_i \sum \frac{1}{(1+d)^t} u(costs_{it})$$

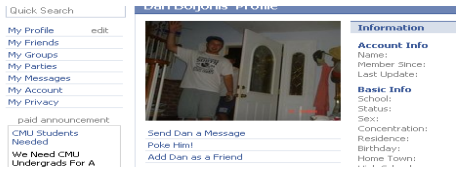


Hurdles Which Hamper (Privacy) Decision Making

1. Incomplete information
   - E.g.: using DOB/hometown to predict individual SSNs (PNAS 2009)

Predicting SSNs from Online Social Networks

Name	Birth	Death	Last Residence	SSN	Issued
JOHN SMITH	1 July 1987	Oct 2005	33540	022-10-4592	NJ



Name	Birth	SSN	Hometown
JOHNN FACEB			

44% of first 5 digits of SSNs issued from 1989 to 2003 predictable at first attempt

8.5% of complete 9-digit SSNs issued from 1989 to 2003 predictable with < 1,000 attempts

Name	Birth	Death	Last Residence	SSN	Issued
JOHN DOE	28 July 1987	Nov 2001	94720	022-12-6744	NJ

Hurdles Which Hamper (Privacy) Decision Making

1. Incomplete information
   - E.g.: using DOB/hometown to predict individual SSNs (PNAS 2009)
2. Bounded rationality
   - I.e., bounded cognitive power
3. Cognitive/behavioral biases, investigated by behavioral economics & decision research
   - E.g., optimism bias, hyperbolic discounting, ambiguity aversion, and so forth

2. The Behavioral Economics of Privacy

From Behavioral Economics, to Privacy

- Behavioral experimental economics has uncovered evidence for several systematic "deviations" from the theoretical rational behavior of the economic agent

- Many of those deviations have applications to the privacy arena (as well as information security)

Hence, the need arises for the application of behavioral, experimental economics to the understanding of privacy decision making (as well as security decision making)

The Behavioral Economics of Privacy

- Some of our previous and ongoing results (2004-2010) Some previous and ongoing results
  - Online social network mining studies
    - Over-confidence, optimism bias in online social networks (*WPES 2005, PET 2006*)…
  - Experiments in the lab or in the field
    - Hyperbolic discounting in privacy valuations (*ACM EC 2004*)…
    - Confidentiality assurances inhibit information disclosure (*SJDM 2007*)…
    - Individuals more likely to disclose sensitive information to unprofessional sites than professional sites (*SJDM 2007*)…
    - Privacy and the illusion of control (*iConference 2009*)…

Can Non-normative Factors Determine Inconsistencies in Privacy Concerns/Valuations?

- Privacy valuations may not just be context-dependent (on this, most researchers would agree) but also:
  - Malleable to non-normative factors
  - In fact, possibly internally inconsistent
- Disclosure likely to be influenced by subtle contextual factors, which can
  - Downplay privacy concerns
  - Act like 'alarm bells' – triggering concern for privacy that is often latent
- Possible explanation for inconsistencies in information revelation

Two Experimental Studies

1. Study 1: The effect of framing on privacy valuations
2. Study 2: The “herding” effect on information disclosure

- *All studies with Leslie John and George Loewenstein*

Study 1: How Framing Impacts Valuations of Personal Data

- Willingness to accept (WTA) money to give away information

  vs.

- Willingness to pay (WTP) money to protect information
- Hypothesis:
  - People assign different values to their personal information depending on whether they are focusing on **protecting it** or **revealing it**

Background: Various Disciplines Implicitly Assume Stable Privacy Preferences

- Alan Westin 1991's privacy "clusters"
  - Unconcerned, pragmatic, fundamentalists
- Sociological approaches: cost-benefit privacy calculus
  - E.g., Laufer & Wolfe 1977, Stone & Stone 1990
- Attempts to quantify "value" of privacy
  - E.g. Huberman et al. (2005), Hann et al. (2007), Danezis et al. (2005), Png (2007), …
- However: reasons to believe privacy valuations may not be consistent or stable (BE, BDR, attitude/behavior paradox, …)

Experimental Design

- Experimental subjects asked to choose between 2 gift cards
  - We manipulated trade-offs between privacy protection and value of cards
- Subjects endowed with either:
  - $10 Anonymous gift card. “Your name will not be linked to the transactions completed with the card, and its usage will not be tracked by the researchers.”
  - $12 Trackable gift card. “Your name will be linked to the transactions completed with the card, and its usage will be tracked by the researchers.”
- Subjects asked whether they’d like to switch cards
  - From $10 Anonymous to $12 Trackable (WTA)
  - From $12 Trackable to $10 Anonymous (WTP)

Two Versions

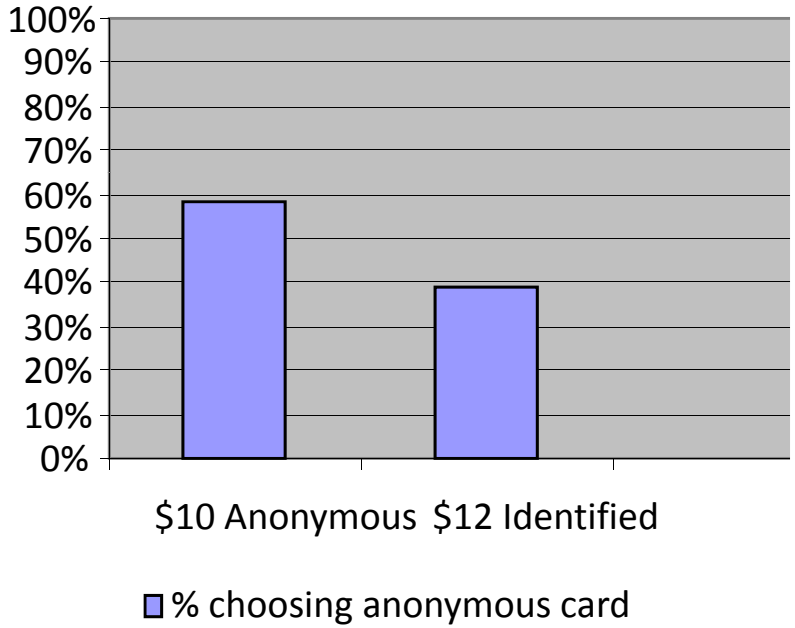
A. Experiment A: Hypothetical choice during a survey

B. Experiment B: Field experiment with actual gift cards

Experiment A: Hypothetical Survey

- "Imagine you have received a gift card…"
- "You have the option to exchange your card for…"
- 2x2 conditions between-subjects design
  - Initial endowment (anonymous vs. identified)
  - Value of tracked card ($12 vs. $10, and $14 vs. $10)
- Run at cafeterias in hospitals in Pittsburgh area
  - 190 participants

Results



Pearson chi2(1) = 4.3631; Pr = 0.037

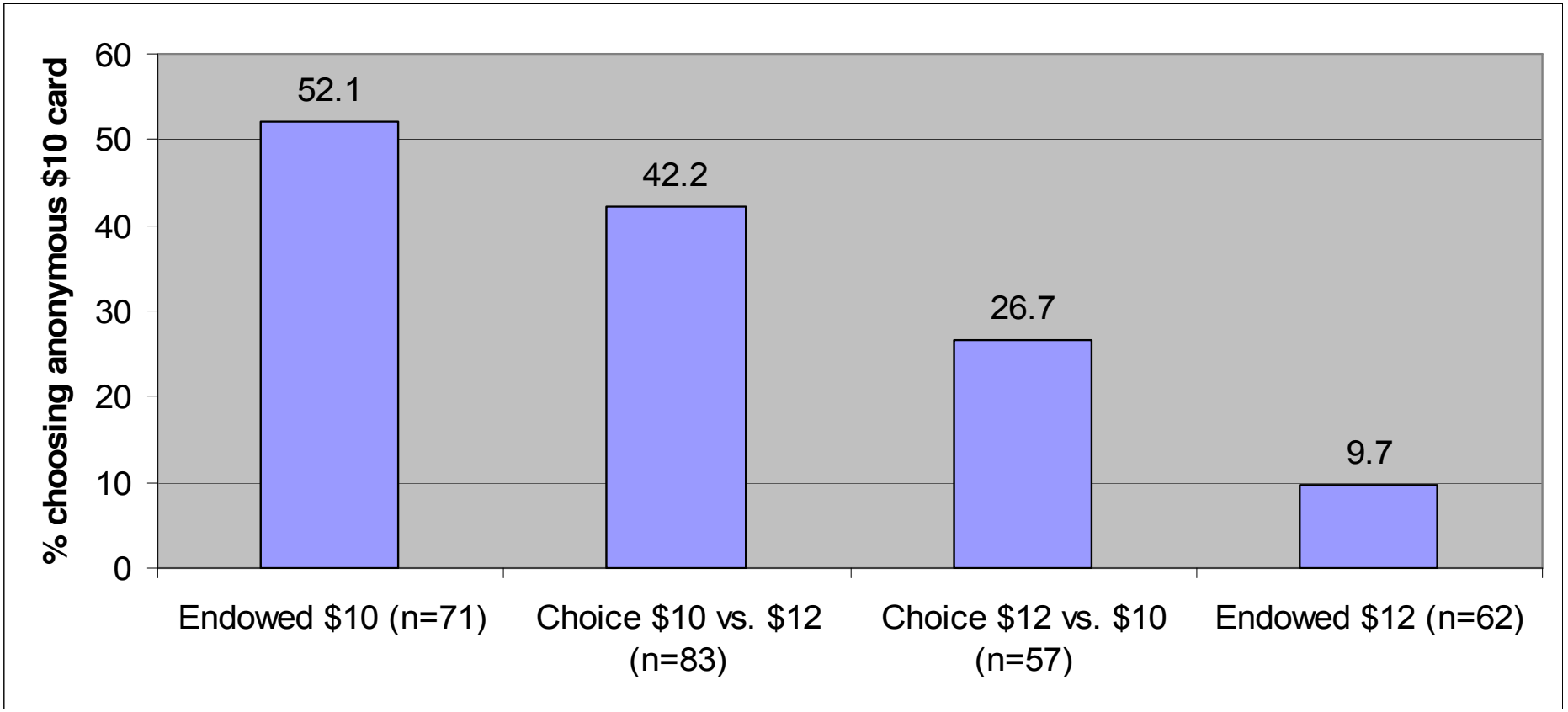
Experiment B: Field Experiment with Actual Gift Cards

- Field experiment. Mall patrons stopped at mall, asked to participate in (unrelated) study, offered real gift card for participation in study
- Participants given choice between:
  - $10 anonymous gift card (card number not recorded)
  - vs. $12 identified card (card number and name recorded)
  - 349 participants

Design

- 4 condition between-subjects design
- Endowment conditions (2):
  - Endowed with $10 anonymous card
  - Endowed with $12 identified card
- Choice conditions (2):
  - $10 anonymous card listed first
  - $10 anonymous card listed second

Results



χ^2 (3) = 30.61, $p < 0.0005$

Implications

- Significant WTP vs. WTA discrepancy found in privacy valuations: Valuations 5 times as large under WTA!
  - Implication: What people say their data is worth depends on how problem is framed
  - People willing to forego cash for privacy if starting from a default position of “protection”
- Therefore, what “value” for privacy should be used in public policy?
  - Analogies to environmental policy

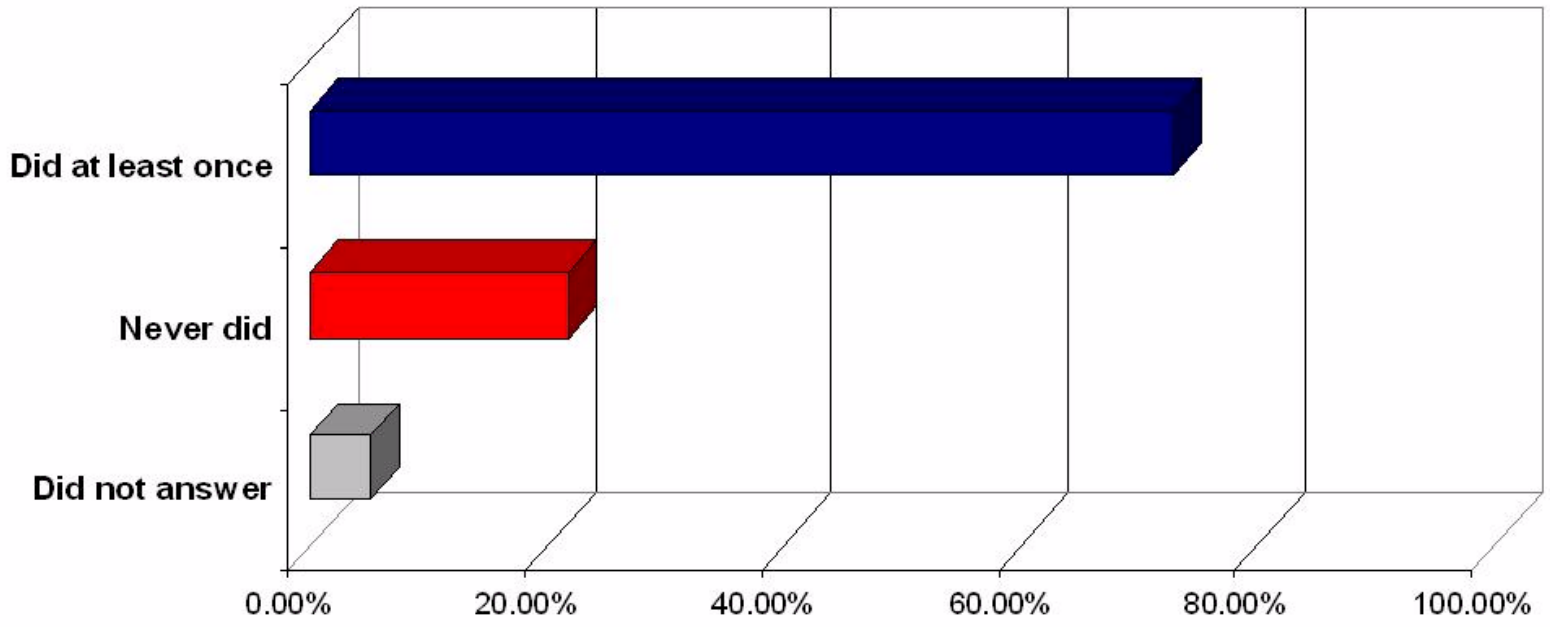
Study 2: Herding Behavior?

- NYT online readers asked to judge the ethicality of, and then whether they had engaged in, a series of sensitive and/or illegal behaviors in a survey "about ethical behavior"
- Survey consisted of six intrusive questions (rated for intrusiveness during a pre-study survey)
  - E.g., `Have you ever had sex with the current husband, wife, or partner of a friend?`
- After answering each question, subjects were presented with the *ostensible* distribution of answers by other survey (fictional) participants
  - In fact, we manipulated those distributions

High Admission Condition

How did other survey participants answer this question so far?

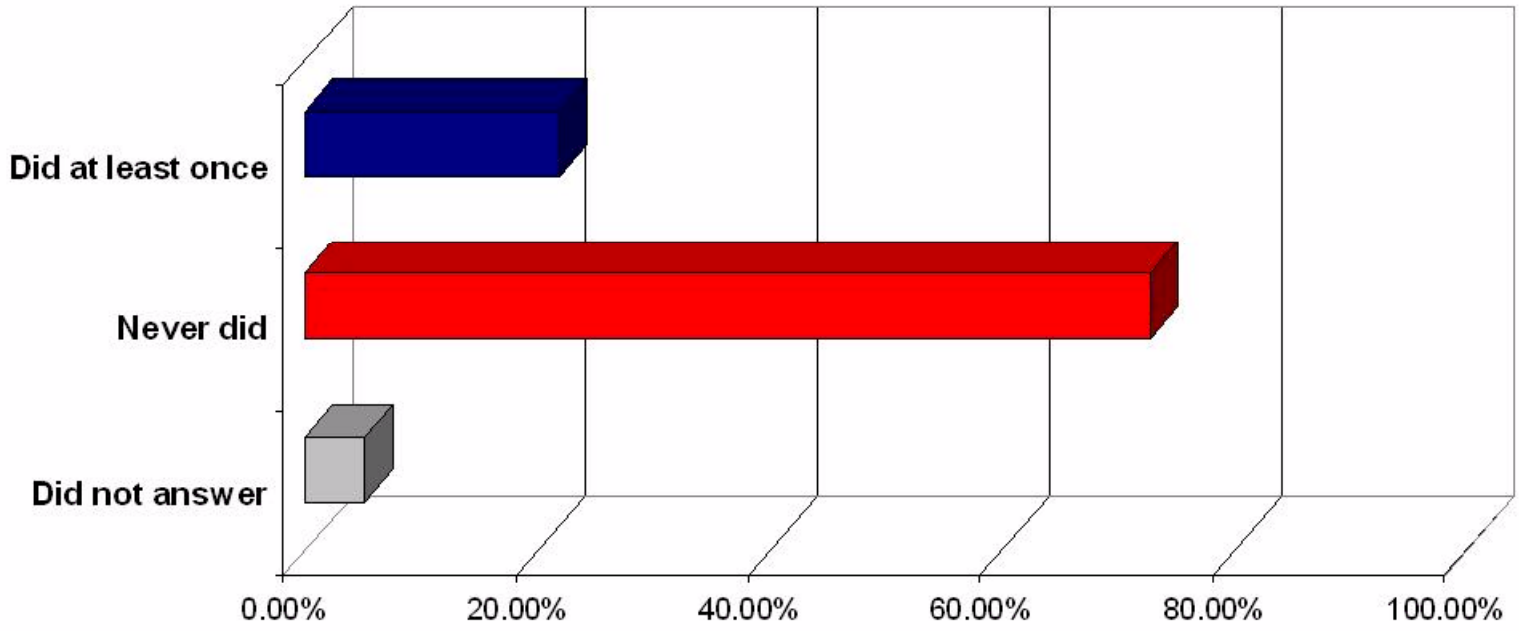
Have you made a false or even somewhat inflated insurance claim?



Low Admission Condition

How did other survey participants answer this question so far?

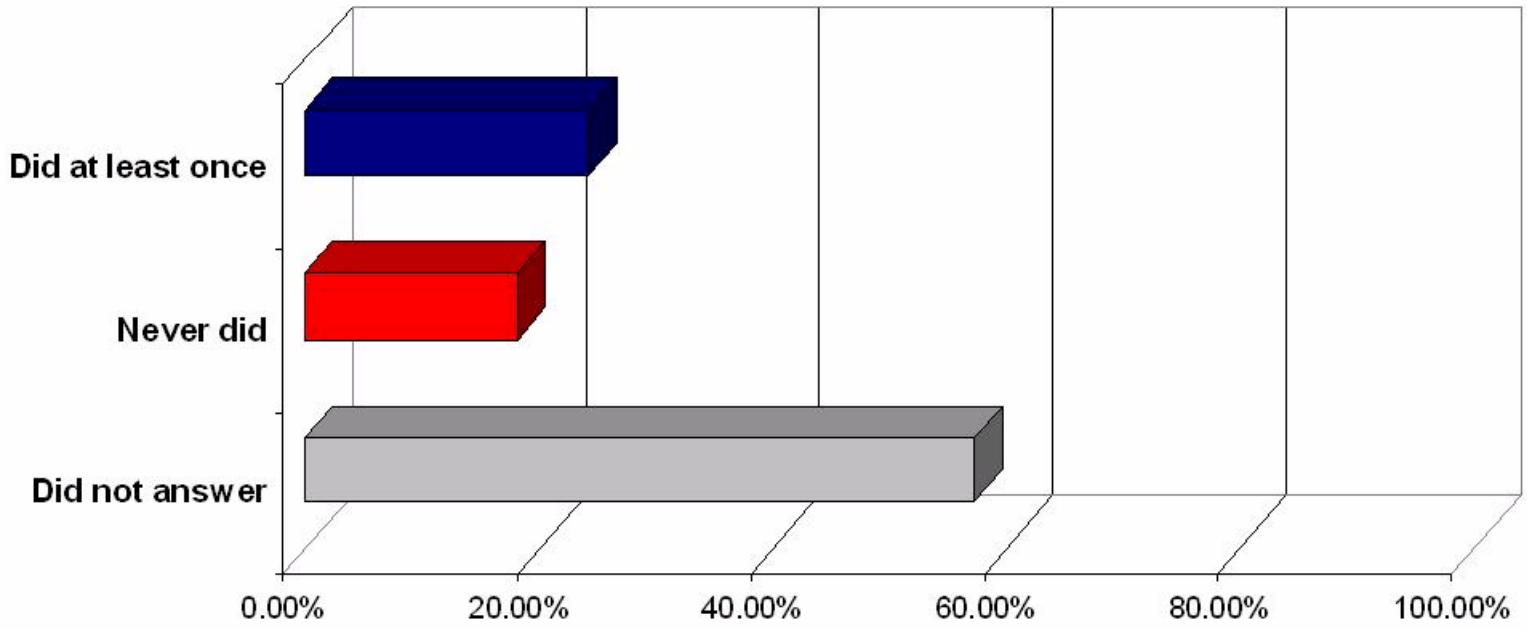
Have you made a false or even somewhat inflated insurance claim?



High “Decline to Answer” Condition

How did other survey participants answer this question so far?

Have you made a false or even somewhat inflated insurance claim?



Design & Hypothesis

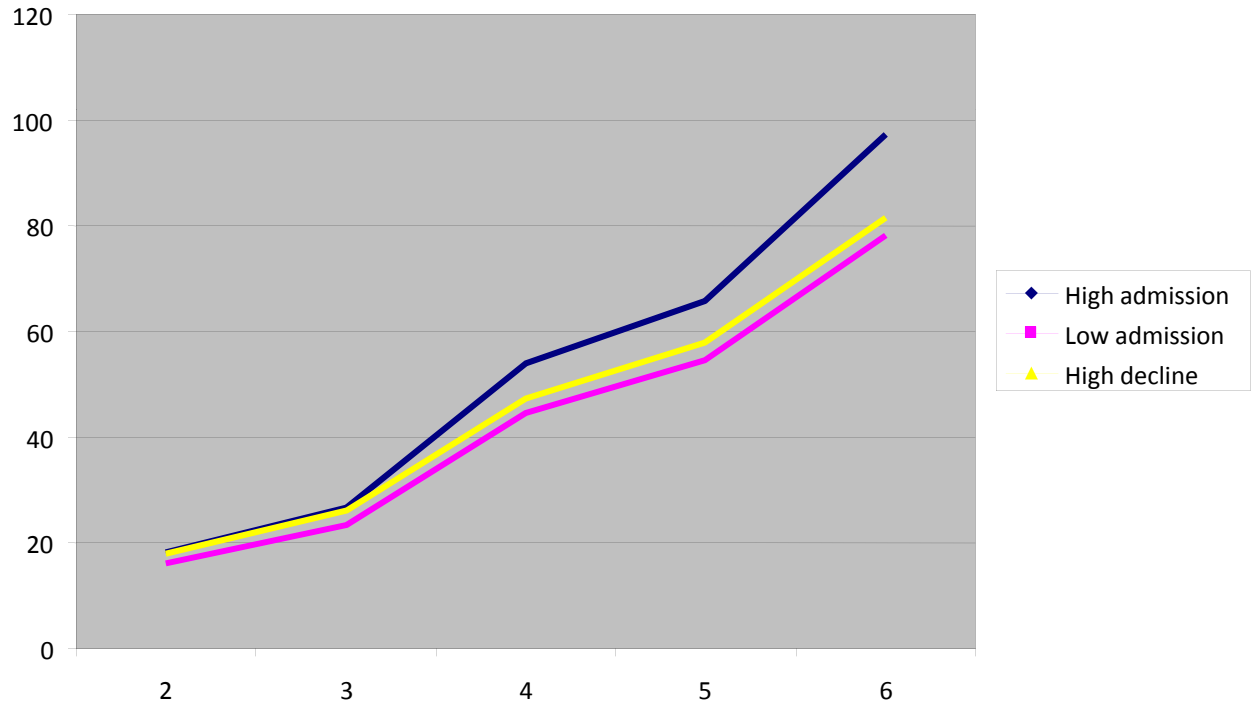
- Three conditions between-subjects design
  - High admission rates
  - Low admission rates
  - High “decline to answer” rates
- Hypothesis: Higher ostensible admission rates by others will trigger higher admission rates by subject. Lower admission rates will inhibit disclosure
  - An “herding” effect (see Asch 1955, “Opinions and Social Pressure”)

Dependent Variables and Demographics

- Dependent variable: Admission to engaging in behavior
  - Yes
  - No
  - No answer (Variation: coding “No answer” as “No”)
- 1,722 online participants

Results

Cumulative admission rates through questions 2 to 6, across conditions



Analysis

- Participants in “high admission” condition were more likely to admit to engaging in sensitive behaviors than participants in other conditions
  - For instance: subjects in “high admission” condition are 33% more likely to admit to having fantasized about non-consensual, violent sex, than subjects in the “low admission” condition

- Strongly significant:
  - Results are in hypothesized direction and statistically significant both when aggregating across questions (repeated-measure ANOVA: $p<0.0001$; RE logistic models) and when testing answers to specific questions
  - *Except, in fact, the first one*

Implications

- People seem more comfortable admitting to sensitive behaviors when other people also admit to (**other**) sensitive behaviors
  - Possible case in point: online social networks
- Note:
  - Robust to consideration of missing/skipped answers (i.e., coding “No answer” as “No”)
  - Most participants provided email addresses; those who didn’t were also *less sensitive* to manipulation

Overall Implications of the Studies

- People's concerns for privacy (and security) depend on priming and framing
- Hence, reliance on "revealed preferences" argument for privacy may lead to sub-optimal outcomes if privacy valuations are inconsistent…
  - People may make disclosure decisions that they stand to later regret
  - Risks greatly magnified in online information revelation
- Therefore, implications for policy-making & the debate on privacy regulation
  - E.g., Rubin & Lenard [2001] vs. Gellman [2001], or Chicago School approach vs. privacy advocates

Implications for Security

- Similar results also apply to information security decision making
- Both at the individual and corporate level, security decisions are affected by systematic (and therefore predictable) biases

So, can we anticipate, counter, or even exploit those biases?

3. Soft Paternalism

Soft paternalism

- "Soft" or asymmetric paternalism: design systems so that they enhance (and sometimes influence) individual choice in order to increase individual and societal welfare
  - **Nudging privacy**: *using soft paternalism to address and improve security and privacy decisions through policy and technology design that anticipates and/or exploits behavioral/cognitive biases* (*IEEE S&P 2009*)

Soft vs. strong paternalism vs. usability

- Consider online social networks users who post dates of birth online
- Imagine that a study shows some risks associated with revealing DOBs (e.g., SSN predictions)
  - Strong paternalistic solution: ban public provision of dates of birth in online profiles
  - “Usability” solution : design a system to make it intuitive/ easy to change DOB visibility settings
  - Soft paternalistic solution?

Nudging privacy

- Saliency of information
  - Provide context to aid the user's decision - such as visually representing how many other users (or types of users) may be able to access that information
- Default settings
  - By default, DOBs not visible, unless settings are modified by user
- Hyperbolic discounting
  - Predict and show immediately SSN based on information provided
- … and so forth

For more info

- Google: economics privacy
- Visit: http://www.heinz.cmu.edu/~acquisti/economics-privacy.htm
- Email: acquisti@andrew.cmu.edu