An informational theory of privacy

Ole Jann and Christoph Schottmüller

Outline

What is privacy (in this paper)?

- ability to take actions without being observed, and having interactions with others confined to the intended recipients
- (many other definitions)
- privacy $\subset$ asymmetric information

Privacy is hot topic

Privacy in news and life:

- government mass surveillance
- business mass surveillance
  - e-business: data as side product of any transaction
  - loyalty cards (AH Bonuskaart)
- voluntary provision of private data (facebook, twitter, mobile phone)
- micro targeting in election campaigns

What is at stake? trade-offs? mechanisms?

Economics of privacy: Chicago school (Stigler 80, Posner 81)

- privacy = asymmetric information
- asymmetric information = inefficiency (Akerlof, Mirrlees etc.)

$\Rightarrow$ privacy = inefficiency

Popular debate: ”Nothing to hide”

- reiterated by Google, facebook, NSA etc.
  - If you have nothing to hide, you do not need privacy.
  - If you have something to hide, you should not do it.
- (similar to Chicago school)

This paper

- our model:
  - information is not perfect (correlation: statistical discrimination literature, Phelps 72, Arrow 73)
  - privacy affects behavior
- main result: privacy can be efficient even when considering *informational effects only* (and we show when)
- other results:
  - lack of privacy changes behavior in one direction ("chilling effects")
  - effectiveness of privacy intrusion is easily overstated
  - privacy is redistributive (affects different people differently)
  - mandating privacy might be necessary to avoid unraveling

Model: An example

- Alice thinks drugs should be legalized and wants to write on her facebook profile about that.
- She is also looking for a job.
- employers do not want to hire drug users
- positive correlation between stand on legalization (θ_i) and drug use (τ_i)
- Two cases: The employer can see what Alice does online, or not.

Model I

- n individuals
- individual i has type (θ_i, τ_i)
  - $\theta_i \sim \Phi(0,1)$, τ_i positively correlated with θ_i (formally: distribution of τ_i at θ_i' fosd distribution at θ_i"$<\theta$')

Model II

1. information aggregation stage:
   - i observes (θ_i, τ_i)
   - i chooses $p_i \in \{0, 1\}$
   - policy $p \in \{0, 1\}$ is implemented with probability $q(m/n)$ where m/n is fraction of individuals choosing $p_i = 1$
   - $q' > 0$ and q point symmetric around 0.5
   - payoff for i: $p\theta_i$
2. interaction stage
   - opposing player (OP) chooses action A (aggressive) or M (mild) against i
   - payoff OP: A gives τ_i, M gives 0
   - payoff i: A gives $-\delta(\tau_i)$, M gives 0
   - δ weakly increasing

no privacy: OP knows p_i
privacy: OP has only prior

Results: Chilling effect

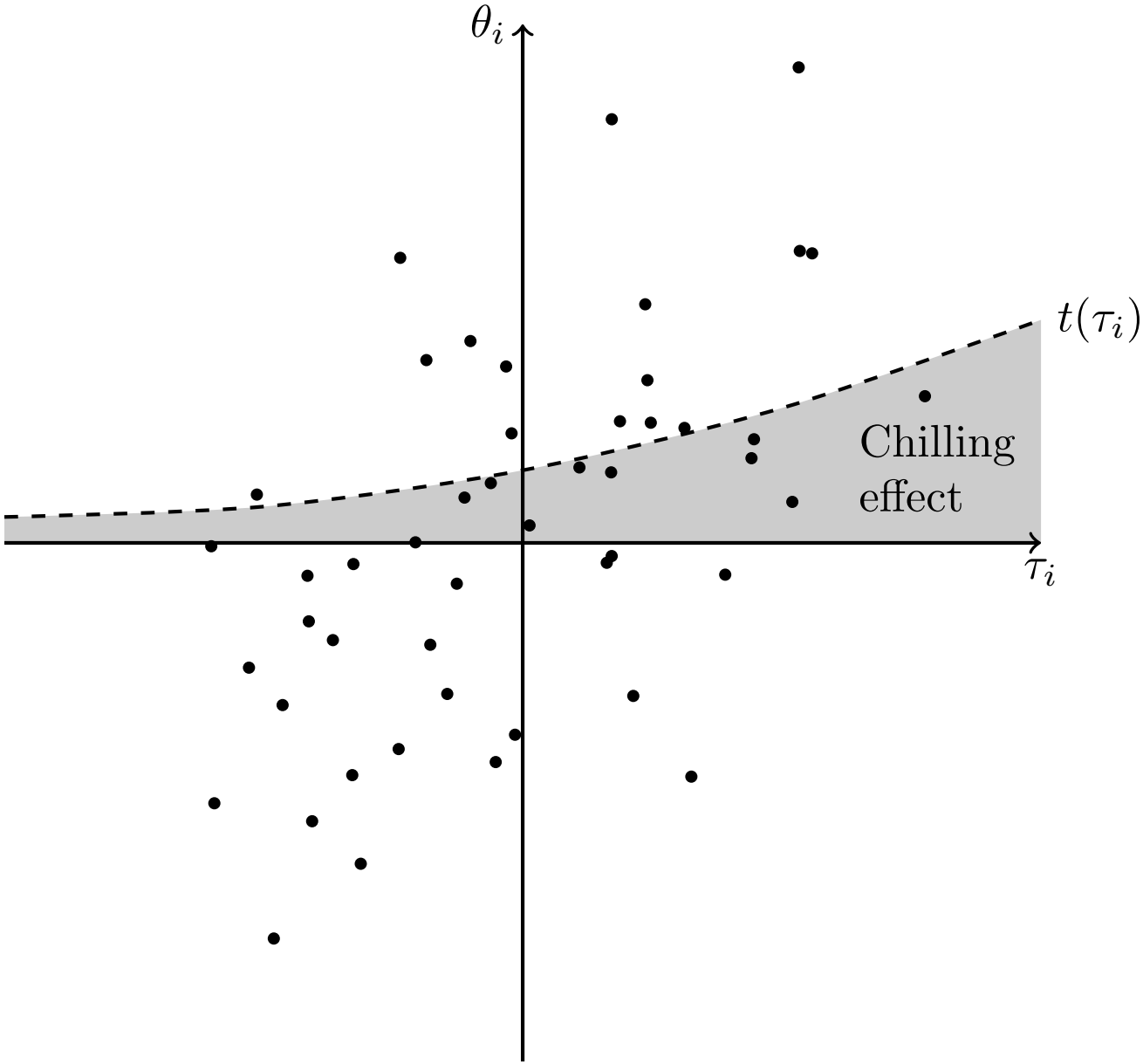
- individuals use cutoff strategies: $p_i = 1$ iff $\theta_i \geq t(\tau_i)$

Theorem (Chilling effect)

With privacy $t^p(\tau_i = 0)$. *Without privacy* $t^{np}(\tau_i) \geq 0$ *and* $t^{np}(\tau_i)$ *is increasing.*

- without privacy individuals with $\theta_i \in [0, t(\tau_i))$ change their behavior ("are chilled")
- without privacy OP plays M against $p_i = 0$ and A against $p_i = 1$
  (assuming privacy affects behavior and equilibrium in pure strategies)

Equilibrium without privacy



Welfare results

Lemma (Consumer surplus)

Expected consumer surplus in the information aggregation stage is maximal at the privacy cutoff $t^p(\tau_i) = 0$.

- welfare includes OP payoff (OP might loose from privacy)

Theorem (Welfare)

Let $\delta' > 0$.
For n sufficiently large, expected consumer surplus is higher under privacy and OP payoff is the same under privacy and no privacy.
Let the disutility of A be $-r\delta(\tau_i)$. For r sufficiently high, consumer surplus is higher under privacy and OP payoff is the same under privacy and no privacy.

Privacy is redistributive

- individuals with $\theta_i < 0$ loose from privacy
- individuals that are chilled $\theta_i \in [0, t^{np}(\tau_i)$ gain a bit from privacy
- individuals with $\theta_i > t^{np}(\tau_i)$ gain double from privacy (decision, OP interaction)

Surveillance performs worse than expected

- seems as eliminating privacy could give OP huge benefit
- but: behavior change (chilling) might reduce correlation between p_i and τ_i
- technical assumption (TA): $\mathbb{E}[\tau_i|\theta_i = 0] \geq 0$
  (if OP knew $\theta_i > 0$, A would be his best response)

Lemma

Assume TA. OP's payoff without privacy is lower if individuals use t^{np} than if they used $t^p = 0$.

Surveillance performs worse than expected

Hello [redacted],

We thought you might be interested in knowing that customers who bought "200g*0.01g Mini Digital Pocket Scale for Jewelry Kitchen Gram Oz Ct" also bought these items.

Customers Also Bought...

100 Plastic Resealable Grip Seal Bags...

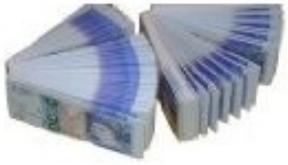
Sold by: AGOODBUYFROMME

100 Grip Seal Bags 2.25 x 3 Inch...

Sold by: Express Goods UK

250 Raw Filter TIPS card booklets...

Sold by: Martins Deals



S AND S Â£20 Pound Note Design...

Sold by: BARGAIN BASE

100 Grip Seal Bags 1.5" x 2.5"

Sold by: Swoosh Supplies

JUICY Juicy Jays Kingsize...

Sold by: Premier Life Store

S AND S 750 Roaches Roach Filter...

Sold by: BargainShop_London

More stuff in the paper I

- Extension: privacy as opt in
  - individuals can choose whether to keep p_i private or not
  - multiple equilibria
  - privacy is not a robust equilibrium: unraveling
- Extension: defensive action
  - Alice can hire a lawyer to help her with the employer
  - i can take costly defensive action that reduces the disutility of A and lowers the payoff of OP (regardless of his action)
  - equilibria where OP is strictly better off with privacy

More stuff in the paper II

- Alternative utility specification: private decision
  - stage 1: no information aggregation, i.e. no externalities from p_j on i
  - additional result: privacy is welfare optimal if correlation between θ_i and τ_i is not too high
- Alternative utility specification: state matching
  - stage 1: payoff of each i is $p\theta$ where θ is an unknown state and each i has a noisy signal θ_i
  - same results but privacy makes every i better off as chilling inhibits efficient information aggregation

Application and discussion: Credit scoring

repayment probability (τ_i) is not directly observable

- Consider two preferences (two different θ_i) that are predictive of τ_i: education and music taste
- Low education and a preference for rap music predict low repayment probability.
- There is a chilling effect in both cases, but in the first we might consider it desirable!?
- Should the bank be allowed to use data on music taste? ("Equal Credit Opportunity Act" outlaws "redlining" in the US)
- blacklist vs. whitelist

(plausibility: facebook owns patent on making credit scores from its user data)

Application and discussion: Working in committees

- committee is debating two policies (e.g. raise interest rates or not)
- debate and vote can either be in secret or in public
- correlation between policy preference θ_i and competence τ_i
- members worry about being perceived as incompetent; advocate less radical positions
- Fed is forced to publish minutes of FOMC meetings since 1993;
  studies show an increase in conformity and a decrease of disagreement with the chairman (Meade and Stasavage 2008)
- Thomas Hoenig, President of the Kansas City Fed: "The *tape has had some chilling effect on our discussions*. I see a lot more people reading their statements."

Conclusion

- we give a simple model to understand privacy from the perspective of information economics
  - no privacy leads to chilling effects
  - privacy can be welfare optimal
  - privacy is redistributive (similar to freedom of speech, Friedman 72)
- the model leads to interesting policy questions
  - blacklist vs. whitelist
  - how open should government be?
  - mandate vs. option of privacy
  - which kind of data should be accessible by who?
- we identify crucial elements to address these questions (correlation, behavior change, threat potential)