A Short Paper on the Incentives to Share Private Information for Population Estimates

Michela Chessa[1], Jens Grossklags[2], and Patrick Loiseau[1]

[1] EURECOM, France. Email: name.surname@eurecom.fr
[2] The Pennsylvania State University, USA.
Email: jensg@ist.psu.edu

Abstract. Consumers are often willing to contribute their personal data for analytics projects that may create new insights into societal problems. However, consumers also have justified privacy concerns about the release of their data.
We study the trade-off between privacy concerns related to data release and the incentives to contribute to the estimation of a population average of a private attribute. Consumers may decide whether to participate in the analytics project, and what level of data precision they are willing to provide. We show that setting a minimum precision level for participating users leads to a strict improvement of the estimation.

Keywords: Non-Cooperative Game Theory, Privacy, Estimation Cost, Data Analytics, Incentives for Participation

1 Introduction

Personal data has been heralded as the "New Oil" of the 21st Century [1]. The trend to economically utilize consumer data is facilitated by the growing importance and popularity of cloud computing services and social network sites.

On the one hand, the newly-won abundance of data allows for rigorous analytic treatment of many complex challenges related to social dynamics, public health considerations, market research, and political decision-making [2]. Many analytic results that are based on individuals' personal data can be interpreted as *public goods* with societal importance and consumers are willing to contribute their personal data for the purpose of creating new insights into societal problems. On the other hand, there are justified *privacy concerns* about the release of personal data, which may be used (or abused) for unsolicited advertisements, or social and economic discrimination (e.g., [3–5]). Individuals may also perceive the release and use of their data as an intrusion of their personal sphere [6, 7], or as a violation of their dignity [8, 9].

Understanding the trade-off between privacy, the quality of data analysis results, and willingness-to-participate in such projects is of current and growing importance [10]. Our research addresses this problem area. More precisely, we are investigating individuals' incentives to participate in data analysis projects

when they have (perceived or actual) privacy cost associated with their data release, but also derive (perceived or actual) benefits from the analysis' results.

Our research models, for example, a situation in which data about individuals is collected in a database (e.g., consumer data or clinical data). Control over the utilization of the data takes two forms: 1) individuals can authorize the access to their data at a self-chosen level of precision, and 2) individuals can decide whether they want to participate (or not), thereby authorizing (or declining) the release of their data irrespective of the level of precision. We further investigate the situation where the research analyst has flexibility to adjust requirements for data precision with the objectives that individuals are still willing to contribute to the project, and that the quality of the estimation improves.

We follow a game-theoretical approach to investigate this trade-off scenario. We conduct a rigorous analysis and derive concrete results about the precision of contributions, the quality of the population estimate, and the overall willingness to contribute to the project.

This paper is structured as follows. We review related work in Section 2. In Section 3, we develop and describe a canonical case of our model with homogeneous agents. We conduct our analysis in Section 4, and offer concluding remarks in Section 5.

2 Related Work

Research on the optimal design of experiments assumes that already the stage of data collection can be influenced by the analyst in order to improve the learning of a linear model [11, 12]. In this paper, we allow the analyst to require data contributions at a certain level of precision to improve the computation of a population estimate, which is a related concept. Optimal design of experiments has been studied from the perspective of incentives [13], or with the scope of obtaining an unbiased estimator [14]. We propose to improve the design of experiments focusing on the privacy concerns of the agents.

Privacy-preserving techniques in the context of data analytics have a long history. Some recent papers propose new approaches, which allow users to protect their privacy selling aggregates of their data [15, 16]. The more classical framework of ϵ-differential privacy [17, 18], assumes that data are perturbed after an analysis has been conducted on unmodified inputs. That is, the analyst is considered trustworthy. In this framework, researchers have also studied the role of incentives [19–21]. Our work differs, as we assume agents to be releasing their data independently, and an untrusted data analyst which motivates perturbations of data before submission. The idea of affecting the level of precision of released personal data, adding noise in advance of data analysis has been studied in the context of privacy-preserving data-mining (see, e.g., [22, 23]) and specific application scenarios such as building decision trees [24], clustering [25], and association rule mining [26]. From a mechanism design perspective, scenarios have been studied where survey subjects are assumed to potentially misreport their private values [27, 28], however, these behaviors are not studied

in the context of a non-cooperative scenario. A strategic approach is followed in [29], where an analyst performs a linear regression based on users' perturbed data (our starting point is a simplified version of this model). We continue this line of research by studying the benefits of restricting potential perturbation on the population estimate accuracy, and the incentives for participation in a game-theoretic framework.

Our research is also relevant to the context of the provisioning of public goods [30]. Our results show a new way of increasing the public good provision by restricting the agents' possible actions, as opposed to using monetary incentives.

3 Model Description

3.1 The Linear Model and the Estimation

Consider a set of n agents, denoted by $N = \{1, \ldots, n\}$. Each agent $i \in N$ is associated with a private variable $y_i \in \mathbb{R}$ which contains sensitive information. We suppose there exists $y_M \in \mathbb{R}$, s.t., the private variables are of the form

$$y_i = y_M + \epsilon_i, \quad \forall i \in N, \tag{1}$$

where ϵ_i are i.i.d., zero-mean random variables with finite variance $\sigma^2 < \infty$, which capture the inherent noise.

An analyst wishes to observe the private variables y_i and to estimate y_M (the mean of the y_i's). The agents, however, motivated by privacy concerns, do not allow the access to the actual values of their private variables, but to a perturbed value with added excess noise. More specifically, for each agent $i \in N$ the perturbed variable is given by $\tilde{y}_i = y_i + z_i$, where z_i is a zero-mean random variable with variance σ_i^2 chosen by her. We assume that the $\{z_i\}_{i \in N}$ are independent and are also independent of the inherent noise variables $\{\epsilon_i\}_{i \in N}$. The aggregate variance of the perturbed variable $\tilde{y}_i$ is $\sigma^2 + \sigma_i^2$. We define

$$\lambda_i = 1/(\sigma^2 + \sigma_i^2) \in [0, 1/\sigma^2], \quad \forall i \in N,$$

the *precision* of the perturbed variables $\tilde{y}_i$, i.e., the inverse of the aggregate variance. To simplify, we will assume that each agent chooses a level of precision $\lambda_i \in [0, 1/\sigma^2]$ (rather than its excess variance σ_i^2, as both are clearly equivalent, or even a more "user friendly" precision level normalized in $[0, 100]$). If agent $i \in N$ has very high privacy concerns, she can choose a precision $\lambda_i = 0$. In our model, this corresponds to adding noise of infinite variance or, equivalently, this represents the fact that the agent can choose not to participate (i.e., not to allow the access to her data). Define $\boldsymbol{\lambda} = [\lambda_i]_{i \in N}$ the vector of the precisions.

The analyst has access to the perturbed variable $\tilde{y}_i$ as well as its precision λ_i, for each $i \in N$. Then, we assume that the analyst estimates the mean as

$$\hat{y}_M(\boldsymbol{\lambda}) = \frac{\sum_{i \in N} \lambda_i \tilde{y}_i}{\sum_{i \in N} \lambda_i}, \tag{2}$$

where observations with smaller variance receive a larger weight. This estimator is the standard generalized least square estimator. The estimator is unbiased, i.e., $\mathbb{E}[\hat{y}_M] = y_M$, and has variance

$$\sigma_M^2(\boldsymbol{\lambda}) = \mathbb{E}[(\hat{y}_M(\boldsymbol{\lambda}) - y_M)^2] = \frac{1}{\sum_{i \in N} \lambda_i} \in [\sigma^2/n, +\infty]. \tag{3}$$

Observe that, when $\lambda_i = 0$ for each $i \in N$, the variance (3) is infinite. This corresponds to the situation in which no agent decided to participate and then the analyst cannot estimate y_M. On the opposite end, when $\lambda_i = 1/\sigma^2$ for each $i \in N$, the analyst estimates y_M with variance σ^2/n, resulting only from the inherent noise. This corresponds to the situation in which each agent is giving data with maximum precision, i.e., no agent is perturbing her private variable. The set of precision vectors for which the estimator has a finite variance is $[0, 1/\sigma^2]^n \setminus \{(0, \dots, 0)\}$.

3.2 The Game Γ *without* Minimal Precision Level

We next describe the interaction between the agents that results in their choices of precision levels. We assume that each agent $i \in N$ wishes to minimize a cost function $J_i : \mathbb{R}^n \to \bar{\mathbb{R}}$, s.t., for each $\boldsymbol{\lambda} \in [0, 1/\sigma^2]^n$,

$$J_i(\boldsymbol{\lambda}) = c\lambda_i^k + \sigma_M^2(\boldsymbol{\lambda}), \tag{4}$$

with $c > 0$ and $k \geq 2$. The first component is the *privacy cost*: $c\lambda_i^k$ is the cost that agent i incurs on account of the privacy violation sustained by revealing the perturbed variable. We assume it to be monomial and depending only on the precision λ_i, hence it is (strictly) convex. The second component, given by the variance of the estimation $\sigma_M^2(\boldsymbol{\lambda})$, is the *estimation cost*: it captures the cost of an inaccurate estimation of the mean. This cost translates the idea that agents benefit from an accurate estimate of the population average y_M. In that sense, the accuracy of the estimate can be seen as a public good to which each agent contributes by its choice of precision λ_i.

To describe the strategic interaction between the agents, we define the game $\Gamma = \langle N, [0, 1/\sigma^2]^n, (J_i)_{i \in N} \rangle$ with set of agents N, strategy space $[0, 1/\sigma^2]$ for each agent $i \in N$ and cost function J_i given by (4).

3.3 The Game $\Gamma(\eta)$ *with* Minimum Precision Level η

As we shall see (Section 4.1), the game Γ has a unique Nash equilibrium $\boldsymbol{\lambda}^*(n)$ for which the estimation cost $\sigma_M^2(\boldsymbol{\lambda}^*(n))$ is larger than the optimal cost σ^2/n due to the excess noise added by agents to protect their privacy. In this paper, we investigate a novel way in which the analyst can mitigate the effect of agents' privacy concerns and to improve the accuracy of the estimation obtained. Specifically, we propose to let the analyst fix a *minimum precision level* $\eta \in [0, 1/\sigma^2]$, which is equivalent to fixing a maximum variance for the noise agents can add

to perturb their data. Obviously, it is not possible to force agents to reveal their data with a given precision (otherwise, the estimation problem would be trivial). Accordingly, we still assume that agents can choose not to participate, choosing a precision level zero. This idea of imposing a minimum precision level allows the analyst to improve the estimation using an adjustment of the initial scheme.

In the variant, we assume that agents are informed of the minimum precision level η and choose their precision λ_i in the range imposed by the analyst or decide not to participate, i.e., choose their precision in $\{0\} \cup [\eta, 1/\sigma^2]$. To analyze this variant, we define a modified game $\Gamma(\eta) = \left\langle N, \left[\{0\} \cup [\eta, 1/\sigma^2]\right]^n, (J_i)_{i \in N} \right\rangle$ (where the cost function J_i is still given by (4)), which is identical to Γ except for the restricted strategy space.

Observe that $\Gamma(0) = \Gamma$. For $\eta > 0$, the two games $\Gamma(\eta)$ and Γ differ as in the original one Γ, the agents can choose any precision, while in the variant $\Gamma(\eta)$, the participating agents have to respect a minimum precision level η. We analyze the two games Γ and $\Gamma(\eta)$ as *complete information games* between the agents, i.e., we assume that the set of agents, the action sets (in particular, when present, the value of the parameter η) and the costs are known by all the agents.

4 The Estimation

4.1 The Estimation in the Game Γ

We first analyze the estimation game Γ, in which the analyst allows the agents to choose any precision level between 0 and $1/\sigma^2$. A Nash equilibrium (in pure strategy) of this game is a strategy profile $\boldsymbol{\lambda}^* \in [0, 1/\sigma^2]^n$ satisfying

$$\lambda_i^* \in \underset{\lambda_i \in [0, 1/\sigma^2]}{\arg\min} \; J_i(\lambda_i, \boldsymbol{\lambda}_{-i}^*), \quad \forall i \in N. \tag{5}$$

The game Γ with strategy space $[0, 1/\sigma^2]$ is a special case of the game in [29], where the existence of a unique Nash equilibrium is established. However, our specific assumptions allow us to characterize the equilibrium in more detail:

Theorem 1. *The game Γ has a unique Nash equilibrium $\boldsymbol{\lambda}^*(n)$ s.t. $\lambda_i^*(n) = \lambda^*(n) > 0$ for each $i \in N$, where $\lambda^*(n)$ is defined by*

$$\lambda^*(n) = \begin{cases} \left(\dfrac{1}{ckn^2}\right)^{\frac{1}{k+1}} & \text{if } \left(\dfrac{1}{ckn^2}\right)^{\frac{1}{k+1}} \leq 1/\sigma^2 \\ 1/\sigma^2 & \text{otherwise .} \end{cases} \tag{6}$$

Proof. Γ is a symmetric potential game, with potential function $\Phi : [0, 1/\sigma^2]^n \to \bar{\mathbb{R}}$, s.t., for each $\boldsymbol{\lambda} \in [0, 1/\sigma^2]^n$

$$\Phi(\boldsymbol{\lambda}) = \sum_{j \in N} c\lambda_j^k + \sigma_M^2(\boldsymbol{\lambda}). \tag{7}$$

By the definition of a potential game, the set of Nash equilibria of Γ is contained in the set of local minima of function Φ. Function Φ has a unique local minimum

$\boldsymbol{\lambda}^* \in [0, 1/\sigma^2]^n$, which is also the unique Nash equilibrium of Γ. The optimum $\boldsymbol{\lambda}^*$ is such that for each $i \in N$, λ_i^* satisfies the following KKT conditions

$$\begin{cases} -\dfrac{1}{(\sum_{j\in N}\lambda_j^*)^2} + ck\lambda_i^{*k-1} - \psi_i^* + \phi_i^* = 0 \\ \psi_i^*\lambda_i^* = 0 \quad \phi_i^*(\lambda_i^* - 1/\sigma^2) = 0, \quad \psi_i^*, \phi_i^* \geq 0. \end{cases} \tag{8}$$

Observe that, as a consequence of the assumption of monomial privacy cost, $\lambda_i^* > 0$ for each $i \in N$. Moreover, as Φ is a symmetric function on a symmetric domain, the only minimum has to be symmetric, i.e., $\lambda_i^* = \lambda^*$ for each $i \in N$. Then, solving the system in (8), we obtain that λ^* is given by (6). □

Theorem 1 states that the unique equilibrium of Γ is symmetric and gives analytically the precision $\lambda^*(n)$ chosen by each agent at equilibrium. Remarkably, we observe that, for any n, $\lambda^*(n) > 0$, i.e., no agent decides not to participate. The equilibrium precision $\lambda^*(n)$ is a function of the number of agents n (unless n is so small that each agent provides data with maximum precision, i.e., no agent distorts her data). From (6), we derive the properties of $\lambda^*(n)$ which are summarized in the following corollary.

Corollary 1. *The equilibrium precision level $\lambda^*(n)$ satisfies*

(i) $\lambda^*(n)$ is a non-increasing function of the number of agents, and
(ii) $\lim_{n\to+\infty} \lambda^*(n) = 0$.

This corollary states that the equilibrium contribution of each agent decreases as the number of agents increases. This is a standard property in public good problems as agents choose their equilibrium contribution such that the marginal increase in the contribution cost equates the marginal decrease in the estimation cost, and the marginal effect of a single agent decreases when the number of agents increases. In the limit when n becomes very large, the contribution of each agent tends to zero (i.e., each agents adds a variance tending to infinity).

The variance of the estimate of y_M obtained by the analyst at equilibrium is

$$\sigma_M^2(\boldsymbol{\lambda}^*(n)) = \frac{1}{n\lambda^*(n)}. \tag{9}$$

The properties of the variance of the population estimate at equilibrium, as a function of the number of agents, are summarized in the following corollary.

Corollary 2. *The equilibrium variance of the estimate of y_M satisfies*

(i) $\sigma_M^2(\boldsymbol{\lambda}^*(n))$ is a non-increasing function of the number of agents n, and
(ii) $\sigma_M^2(\boldsymbol{\lambda}^*(n)) \sim_{n\to\infty} n^{\frac{2}{k+1}-1}$ and $\lim_{n\to+\infty} \sigma_M^2(\boldsymbol{\lambda}^*(n)) = 0$.

Proof. When $n_1 \geq n_2 > 0$, then $\lambda^*(n_1) \leq \lambda^*(n_2)$, because of Corollary 1. In particular, there exists $m > 0$ s.t., for each $n \geq m$, $\lambda^*(n) = \left(\frac{1}{ckn^2}\right)^{\frac{1}{k+1}}$, and we may write the estimation cost as

$$\sigma_M^2(\boldsymbol{\lambda}^*(n)) = c^{\frac{1}{k+1}} k^{\frac{1}{k+1}} n^{\frac{2}{k+1}-1}.$$

Then, $\sigma_M^2(\boldsymbol{\lambda}^*(n)) \sim_{n\to\infty} n^{\frac{2}{k+1}-1}$. For $k > 1$, this is a decreasing function which goes to zero when n goes to infinite. This proves *(ii)* and *(i)* in the case $n_1 \geq n_2 \geq m$. When $m \geq n_1 \geq n_2$, then

$$\sigma_M^2(\lambda^*(n_1)) = n_1^{-1}\sigma^2 \leq n_2^{-1}\sigma^2 \leq \sigma_M^2(\lambda^*(n_2)),$$

and when $n_1 \geq m \geq n_2$, then

$$\sigma_M^2(\lambda^*(n_1)) = n_1^{-1}\lambda^*(n_1)^{-1} \leq n_2^{-1}\left(\frac{1}{ckn_2^2}\right)^{\frac{1}{k+1}} \leq n_2^{-2}\sigma^2 = \sigma_M^2(\lambda^*(n_2)).$$

□

Corollary 2-*(i)* shows that, for the analyst, it is always better to have a larger number of agents giving data despite the fact that, when the number of agents increases, each agent gives data with smaller precision (see Corollary 1). Part *(ii)* of Corollary 2 analyzes the case for a large number of agents n. Interestingly, when n gets large, the variance decreases at a rate smaller from the standard $1/n$. In particular, if k is small, the rate of decrease can be very slow. On the other end of the spectrum, if n is low (such that $\left(\frac{1}{ckn^2}\right)^{\frac{1}{k+1}} > 1/\sigma^2$), then at equilibrium every agent chooses the maximum precision level, and the estimation of y_M has minimum variance equal to σ^2/n.

4.2 The Estimation in the Game $\Gamma(\eta)$

We now move to the case where the analyst introduces a minimum precision level $\eta \in [0, 1/\sigma^2]$ with the goal of improving the accuracy of the estimate. We assume that $\lambda^*(n) \neq 1/\sigma^2$, since in that case, the estimation is already optimal with $\eta = 0$. A Nash equilibrium (in pure strategy) of the game $\Gamma(\eta)$ is a strategy profile $\boldsymbol{\lambda}^* \in \left[\{0\} \cup [\eta, 1/\sigma^2]\right]^n$ satisfying

$$\lambda_i^* \in \underset{\lambda_i \in \{0\}\cup[\eta,1/\sigma^2]}{\arg\min} J_i(\lambda_i, \boldsymbol{\lambda}_{-i}^*), \quad \forall i \in N. \tag{10}$$

In the following lemma, we state that it is possible for the analyst to improve the estimation by setting a strictly positive minimum precision level.

Theorem 2. *Given Γ with equilibrium precision level $\lambda^*(n) \neq 1/\sigma^2$, there exists $\eta \in (\lambda^*(n), 1/\sigma^2]$ s.t. $\Gamma(\eta)$ has a unique Nash equilibrium $\lambda^*(n, \eta)$ and the estimation cost at equilibrium is strictly smaller, i.e., $\sigma_M^2(\lambda^*(n, \eta)) < \sigma_M^2(\lambda^*(n))$.*

Proof. The game $\Gamma(\eta)$ is a potential game, with potential function as in (7), but restricted to the smaller domain $\left[\{0\} \cup [\eta, 1/\sigma^2]\right]^n$. At first, we focus on the local minima in $[\eta, 1/\sigma^2]^n$. When $\eta = \lambda^*(n) + \epsilon$, with $\epsilon > 0$, the vector $\boldsymbol{\eta} = [\eta]_{i\in N}$ is the only local minimum of Φ on $[\eta, 1/\sigma^2]^n$. Because of the convexity of Φ, any deviation of agent $i \in N$ to a precision level in $(\eta, 1/\sigma^2]$ would make her cost function bigger. Moreover, if agent $i \in N$ deviates to 0, her cost function

does not become smaller if $\lambda^*(n) \leq \left(\frac{1}{cn(n-1)}\right)^{\frac{1}{k+1}} - \epsilon$, and there always exists $\epsilon > 0$ s.t. this inequality holds and the corresponding $\boldsymbol{\eta}$ is a Nash equilibrium. We show that there exists ϵ s.t. $\Gamma(\eta)$ has unique equilibrium $\boldsymbol{\eta}$. First, we state the following lemma.

Lemma 1. *Suppose that $\boldsymbol{\lambda}' = (\lambda'_1, \ldots, \lambda'_n)$ is a local minimum of the potential function Φ on $\left[\{0\} \cup [\eta, 1/\sigma^2]\right]^n$, with $\eta \in (\lambda^*(n), \lambda^*(n-t)]$, $T = \{i \in N : \lambda'_i = 0\}$ and $t = |T|$. Then, $\boldsymbol{\lambda}'$ is a local minimum on $\{0\}^t \times [\eta, 1/\sigma^2]^{n-t}$ and it is s.t. $\lambda'_i = \lambda^*(n-t)$ for each $i \in N \setminus T$.*

Let ϵ be s.t. $\eta = \lambda^*(n) + \epsilon \leq \lambda^*(n-t)$. Suppose that there exists a local minimum $\boldsymbol{\lambda}'$ s.t. calling $T = \{i \in N : \lambda'_i = 0\}$, then $t = |T| \geq 1$, i.e., the set of agents who are at a zero precision level is nonempty. Then, because of Lemma 1, $\lambda'_i = \lambda^*(n-t)$ for each $i \in N \setminus T$. This cannot be a Nash equilibrium. In fact,

$$\frac{1}{(n-t)\lambda^*(n-t)} > \frac{1}{(n-t+1)\lambda^*(n-t)} + c\lambda^*(n-t)^k,$$

when $k \geq 2$, meaning that if an agent in T deviates moving from the precision level 0 to the precision level $\lambda^*(n-t)$, she can strictly decrease her cost function. Then, $\boldsymbol{\eta}$ is the only Nash equilibrium and it is s.t. $\sigma_M^2(\boldsymbol{\lambda}^*(n, \eta)) = 1/(n\eta) < 1/(n\lambda^*(n)) = \sigma_M^2(\boldsymbol{\lambda}^*(n))$. □

Theorem 1 shows that the analyst can indeed improve the quality of the estimation simply by setting a minimum precision level.

5 Concluding Remarks

In this paper, we investigated the problem of estimating population quantities with privacy-sensitive agents who add noise to their data before revealing it to the analyst. The agents choose the precision of the data they reveal to balance their privacy cost and the benefit they derive from a more accurate population estimate. We show that the analyst can improve the population estimate's accuracy by restricting the variance of the noise users can add while maintaining incentive compatibility (i.e., users are still willing to give their data with limited noise rather than dropping out). Our results posit a new way of increasing the provision of a public good (here, the population estimate's accuracy) beyond the level of voluntary contributions by restricting the agents' strategy space. This scheme is attractive by its simplicity, as it does not involve for instance transfers of money that are used in more classical schemes.

In this short paper, we proposed a first analysis of the model, making a number of restrictive assumptions. However, the interesting results we obtained make really appealing an extension of this work. In particular, we suggest three possible lines of future research. First, our model assumes a perfectly symmetric scenario. Understanding how our results can be extended to the heterogeneous agent case is, in our opinion, the first possible future work. See, for example, [31]

for a distribution of privacy valuations across data types. Second, our system is very specific in the choice of the definition of the estimation cost and of the privacy cost. It would be interesting to investigate how the model behaves when assuming more abstract cost functions, to verify its applicability to more general scenarios. Third, we assumed that the analyst can collect data from n agents at no cost and we showed that the accuracy of the population estimate increases with n (although each individual contributes less). However, there could be a cost of collecting the data per agent (e.g., cost of asking for consent). A better understanding of this factor is of high practical relevance.

Acknowledgements

This work was funded by the French Government (National Research Agency, ANR) through the "Investments for the Future" Program reference # ANR-11-LABX-0031-01. We would like to thank the anonymous reviewers and Alvaro Cardenas for their helpful comments.

References

1. World Economic Forum: Personal Data: The Emergence of a New Asset Class. (2011)
2. Varian, H.: Beyond big data. Business Economics **49**(1) (2014) 27–31
3. Acquisti, A., Fong, C.: An experiment in hiring discrimination via online social networks. Technical report (2013) Available at SSRN: http://ssrn.com/abstract=2031979.
4. Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Crowd-assisted search for price discrimination in e-commerce: First results. In: Proceedings of the Conference on Emerging Networking Experiments and Technologies (CoNEXT). (2013) 1–6
5. Spiekermann, S., Grossklags, J., Berendt, B.: E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior. In: Proceedings of the 3rd ACM Conference on Electronic Commerce. (2001) 38–47
6. Altman, I.: The Environment and Social Behavior. Belmont (1975)
7. Warren, S., Brandeis, L.: The Right to Privacy. Harvard Law Review (1890) 193–220
8. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Security & Privacy **3**(1) (2005) 26–33
9. Westin, A.: Privacy and freedom. Atheneum, New York (1970)
10. Lane, J., Stodden, V., Bender, S., Nissenbaum, H.: Privacy, Big Data, and the Public Good: Frameworks for Engagement. Cambridge University Press (2014)
11. Pukelsheim, F.: Optimal design of experiments. Volume 50. Society for Industrial Mathematics (2006)
12. Atkinson, A., Donev, A., Tobias, R.: Optimum experimental designs, with SAS. Oxford University Press New York (2007)
13. Horel, T., Ioannidis, S., Muthukrishnan, S.: Budget feasible mechanisms for experimental design. In Pardo, A., Viola, A., eds.: LATIN 2014: Theoretical Informatics. Volume 8392 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2014) 719–730

14. Roth, A., Schoenebeck, G.: Conducting truthful surveys, cheaply. In: Proceedings of the 13th ACM Conference on Electronic Commerce. (2012) 826–843
15. Riederer, C., Erramilli, V., Chaintreau, A., Krishnamurthy, B., Rodriguez, P.: For sale : Your data: By : You. In: Proceedings of the 10th ACM Workshop on Hot Topics in Networks. (2011) 13:1–13:6
16. Bilogrevic, I., Freudiger, J., De Cristofaro, E., Uzun, E.: What's the gist? Privacy-preserving aggregation of user profiles. In Kutylowski, M., Vaidya, J., eds.: Computer Security - ESORICS 2014. Volume 8713 of Lecture Notes in Computer Science. Springer International Publishing (2014) 128–145
17. Dwork, C.: Differential privacy. In Bugliesi, M., Preneel, B., Sassone, V., Wegener, I., eds.: Automata, Languages and Programming. Volume 4052 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2006) 1–12
18. Kifer, D., Smith, A., Thakurta, A.: Private convex empirical risk minimization and high-dimensional regression. JMLR W&CP (Proceedings of COLT 2012) **23** (2012) 25.1–25.40
19. Ghosh, A., Roth, A.: Selling privacy at auction. In: Proceedings of the 12th ACM Conference on Electronic Commerce. (2011) 199–208
20. Nissim, K., Smorodinsky, R., Tennenholtz, M.: Approximately optimal mechanism design via differential privacy. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. (2012) 203–213
21. Ligett, K., Roth, A.: Take It or Leave It: Running a Survey When Privacy Comes at a Cost. In Goldberg, P., ed.: Internet and Network Economics. Volume 7695 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2012) 378–391
22. Vaidya, J., Clifton, C., Zhu, Y.: Privacy Preserving Data Mining. Springer (2006)
23. Domingo-Ferrer, J.: A survey of inference control methods for privacy-preserving data mining. In Aggarwal, C., Yu, P., eds.: Privacy-Preserving Data Mining. Volume 34 of Advances in Database Systems. Springer (2008) 53–80
24. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. (2000) 439–450
25. Oliveira, S., Zaiane, O.: Privacy preserving clustering by data transformation. In: Proceedings of the XVIII Simposio Brasileiro de Bancos de Dados. (2003) 304–318
26. Atallah, M., Bertino, E., Elmagarmid, A., Ibrahim, M., Verykios, V.: Disclosure limitation of sensitive rules. In: Proceedings of the Workshop on Knowledge and Data Engineering Exchange (KDEX'99). (1999) 45–52
27. Dekel, O., Fischer, F., Procaccia, A.D.: Incentive compatible regression learning. Journal of Computer and System Sciences **76**(8) (2010) 759–777
28. Perote, J., Perote-Pena, J.: Strategy-proof estimators for simple regression. Mathematical Social Sciences **47**(2) (2004) 153–176
29. Ioannidis, S., Loiseau, P.: Linear regression as a non-cooperative game. In Chen, Y., Immorlica, N., eds.: Web and Internet Economics. Volume 8289 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013) 277–290
30. Morgan, J.: Financing public goods by means of lotteries. Review of Economic Studies **67**(4) (2000) 761–84
31. Acquisti, A., Grossklags, J.: An online survey experiment on ambiguity and privacy. Communications & Strategies **49**(4) (2012) 19–39