Optimal Privacy-Constrained Mechanisms*

Ran Eilat†, Kfir Eliaz‡ and Xiaosheng Mu§

September 21, 2019

Abstract

Modern information technologies make it possible to store, analyze and trade unprecedented amounts of detailed information about individuals. This has led to public discussions on whether individuals' privacy should be better protected by restricting the amount or the precision of information that is collected by commercial institutions on its participants. We contribute to this discussion by proposing a Bayesian approach to measure loss of privacy and applying it to the design of optimal mechanisms. Specifically, we define the loss of privacy associated with a mechanism as the difference between the designer's prior and posterior beliefs about an agent's type, where this difference is calculated using Kullback-Leibler divergence, and where the change in beliefs is triggered by actions taken by the agent in the mechanism. We consider both ex-post (for every realized type, the maximal difference in beliefs cannot exceed some threshold κ) and ex-ante (the expected difference in beliefs over all type realizations cannot exceed some threshold κ) measures of privacy loss. Using these notions we study the properties of optimal privacy-constrained mechanisms and the relation between welfare/profits and privacy levels.

---

*We thank Benny Moldovanu and Doron Ravid for helpful conversations.

†Dept. of Economics, Ben Gurion University. eilatr@bgu.ac.il

‡School of Economics, Tel-Aviv University and David Eccles School of Business, the University of Utah. kfire@tauex.tau.ac.il.

§Columbia University Department of Economics. xm2230@columbia.edu. Xiaosheng Mu acknowledges the hospitality of the Cowles Foundation at Yale University, which hosted him during parts of this research.

1 Introduction

Modern information technologies make it possible to store, analyze and trade unprecedented amounts of detailed information about individuals. At the same time, the rapid growth of online markets has significantly increased the participation of individuals in decentralized pricing mechanisms that rely on personal information provided by the participants. Consequently, the organizers of these markets are able to gather vast amounts of data on individuals' characteristics such as their tastes and willingness-to-pay for products and services. These data are valuable to a variety of entities including commercial firms as well as political institutions. If leaked to these entities, this information may be used against the users' interests. In light of this, there has been a growing sentiment that governments should enact laws that regulate the ability of private entities to collect and use personal information. If the growing concerns for maintaining privacy were to lead to regulations that impose privacy constraints on pricing mechanisms, how would that affect the design of these mechanisms, and what is the trade-off between profits, welfare and privacy? This paper takes a step towards addressing these questions by proposing a Bayesian approach to the measurement of loss of privacy and applying this approach to the design of optimal mechanisms that are restricted in the amount/precision of private information that they can elicit from participants.

The cornerstone of our approach is the idea that loss of privacy is a *relative* notion: How much information is effectively extracted from an individual through his actions should be measured relative to what is already known about that individual. For example, if it is *publicly* known that some individual was convicted of some crime, then there is no loss of private information when that person voluntarily discloses his conviction. Similarly, how much privacy is lost when a reputed business manager (say, the CEO of a publicly traded firm) discloses her wealth should depend on existing public information (for example, about her annual income). In the context of private information that is gleaned through trade, a seller who operates in a particular zipcode may already know the wealth distribution of the residents in that zipcode. Similarly, if an individual who is already active on some platform (say, an online radio) decides to make a purchase on that platform, the information revealed from the purchase itself is in addition to the data that the platform has already collected prior to the purchase, such as that person's taste in music.

Taking the idea of relative privacy to mechanism design, we suppose that the designer already possesses some information about the participants in the form of a prior belief over their "types". He updates this belief as a result of the participants' interactions with the

mechanism, which reveal more information about them. For example, if the mechanism consists of a menu, then an individual's choice of a particular menu item enables the designer to learn additional information about this person's preferences. Consequently, the designer's posterior belief about the participant's type may be quite different from his prior belief. This suggests that the *difference* between the designer's prior and posterior beliefs should serve as the basis for measuring the loss of privacy associated with a particular mechanism.

Building on this observation, we propose a Bayesian measure of relative privacy loss for mechanisms and apply it to screening mechanisms (which shut down the strategic interaction among different participants). Specifically, we consider the classic Mussa-Rosen set-up in which a monopolist seller faces increasing marginal cost for producing a higher quantity (or quality) of a product, and wishes to offer the optimal menu of quantity-price pairs to consumers with private willingness-to-pay per unit. The standard solution implies that, under a regularity condition, all consumer types that opt in perfectly reveal their private types. Hence, the optimal solution entails *complete* loss of privacy: The designer has a degenerate posterior belief on the type of each participant.

To study the design of mechanisms that preserve some level of privacy, we follow the information theory literature and propose to measure a mechanism's inherent loss of privacy as the *maximal relative entropy* (i.e., the Kullback-Leibler divergence) between the seller's posterior and prior beliefs. The posterior belief is derived from the "message" received in the mechanism, and the maximum KL divergence is computed across all messages (i.e., across all consumer types). We then augment the standard mechanism design problem by restricting attention to mechanisms with privacy loss at most κ. The parameter κ, which takes values between 0 (full privacy) and infinity (no privacy), captures the strength of the privacy requirement. For any finite κ, our approach imposes an upper bound on the amount of new information that a designer may learn about any participant.

Before describing our results, we offer some discussion of this modeling framework. In considering posterior beliefs that are derived from the *messages*, rather than the *outcomes*, we study here a benchmark in which all the actions by participants are observed by the designer, who cannot credibly commit not to access the data received. Our approach to privacy is concerned with constraining the amount of information learned by the *mechanism designer*. Even though the outcome of the mechanism may reveal little information to an *outside observer*, the actions taken by participants are observed by the designer and may reveal much more. This is a key difference between our approach and the notion of differential privacy, which we discuss in detail in Section 2.

Our choice of an information-theoretic measure of privacy is motivated by the following reasons. First, we seek a "context-independent" measure that is portable and can be applied to any mechanism design setting regardless of the particular type space and details of the interaction. In particular, KL divergence has the desirable property of being *invariant to transformations of the type space*, so that how types are "labeled" does not matter.[1] Second, our measure requires making no specific assumptions about the exact form of future interactions between the data holder and the agent. This approach is consistent with the widely agreed view that the fast-paced technological, cultural and political developments of the information era make it impossible to predict how data collected today will be used in the future. Given this, we believe that assuming perfect knowledge of the continuation game is rather extreme.[2,3] Finally, our measure captures the idea that greater privacy loss means smaller uncertainty about a consumer's characteristics after he takes an action. Different functional forms can be used to measure the change in the designer's beliefs, and our choice of KL divergence is made for tractability; in fact, our qualitative results would be unchanged if we were to use any member in the general class of f-divergences.

By imposing an exogenous privacy constraint, we take a "paternalistic" approach to privacy in the sense that we do not explicitly model consumers' preferences over privacy (i.e., how consumers trade-off privacy, consumption and money), but rather assume that mechanisms are required to guarantee a certain level of privacy. This is motivated by research showing that most consumers are not fully aware of the implications of allowing commercial entities to record information about them. Indeed, many users make public postings on social media, log in to websites through their social media accounts and do not delete cookies (e.g., see Acquisti and Grossklags (2005), Barth and de Jong

[1] Formally, suppose that instead of measuring the change in beliefs about the type θ, we compute the KL divergence between posterior and prior distributions about $g(\theta)$ for some smooth injective function g. Then the resulting measure is unchanged.

[2] Quoting the renowned computer scientist and cryptography expert Shafi Goldwasser in Lan (2019, translated): "... You can't know what tomorrow will bring. The best example is genetic information. The genome has a lot of information today that no one even knows what it is, and maybe in 10 more years we'll know. For example, if you're prone to certain diseases. Today you release information that you don't know if you want it to be stored, not just about you, but about your family. It's quite worrying, because you can't even know what can be done with this information, and not just for commercial purposes. Today's government is not tomorrow's government."

[3] Oftentimes data that are collected for one purpose are used for a surprisingly different one. A notable example is the 2018 Cambridge Analytica political scandal in which it was revealed that Cambridge Analytica had harvested the personal data of millions of peoples' Facebook profiles without their consent and used it for political advertising purposes.

(2017) and Kokolakis (2017)).[4] Our paternalistic approach is analogous to the one that is frequently taken in the design of commitment mechanisms, which try to help agents with self-control problems by limiting their choice sets even when there is no demand for such commitment devices on part of the agents.[5,6] Alternatively, our approach can be interpreted as assuming that individuals have lexicographic preferences over material wealth and privacy that take a threshold form: They participate in a mechanism only if the associated privacy loss obeys an exogenous upper bound, but conditional on participating, their preferences over consumption and money are standard. While under this interpretation our model is stated for the case where all consumers have the same subjective privacy threshold, our results can also be applied to analyze the case where there is a distribution of privacy thresholds in the population, so long as this distribution is independent of the distribution of consumption valuations.

Our main results highlight the following key properties of optimal privacy-preserving mechanisms. Within the class of static mechanisms (see Section 6 for discussion of this restriction), the optimal κ-constrained mechanism partitions the type space into *finitely* many intervals (whose number depends on κ and is bounded above by e^{κ}), such that consumers truthfully announce to which interval their type belongs. Thus, even though there is a continuum of types, and the privacy constraint allows for a continuum of noisy messages,[7] maximal profits are attained with only finitely many messages. When the production cost function and the type distribution satisfy some additional properties, any optimal κ-constrained mechanism consists of intervals that are *monotonically* ordered according to their mass from left to right, so that the gain from differentiating higher types

[4]Even if consumers were to fully endogenize the consequences of disclosing their private information, it is often the case that individuals neglect the effects of sharing their information on others. For example, when uploading a photo to Instagram, the privacy of all those who appear in the photo is affected; When disclosing health information online, the privacy of all genetically related relatives is affected; When paying a high price in an online market, some information about other consumers from the same geographic location is leaked. These and other examples suggest that it might sometimes be beneficial (in terms of social welfare) to exogenously restrict the amount of data that a mechanism can collect. For a discussion of the negative externalities and the "public bad" properties of (loss of) privacy, see e.g. Choi et al. (2019) and Fairfield and Engel (2015).

[5]David Laibson (2018) calls it the commitment paradox and gives the following example: "College instructors adopt course policies that force students to focus on their coursework: pop quizzes, classroom attendance requirements, cold calling, graded problem sets, deadlines, classroom wifi blocking, and classroom laptop bans. In my experience, most students don't welcome these paternalistic restrictions."

[6]Some studies, however, consider heterogeneous levels of self-control or heterogeneous levels of sophistication, in which case commitment devices serve as a tool for screening.

[7]E.g., when each type θ reports $\theta + \varepsilon$, where ε is a continuous random variable.

either diminishes or increases (we give sufficient conditions for each of these cases). If we impose more structure on the cost function, we can also give some welfare implications of the privacy constraint. In particular, when costs are quadratic, total welfare is maximized at $\kappa = 0$ and minimized at $\kappa = \infty$ when the prior density function is increasing, while the opposite is true if it is decreasing.

To illustrate a complete characterization of the optimal privacy-preserving mechanism, we analyze the uniform-quadratic case where types are drawn from a uniform distribution and costs are quadratic. In this case, for any $\kappa \in [\log(n), \log(n+1))$, the κ-optimal mechanism divides the type space into n equal intervals. Hence, when κ is in the interior of this interval, the privacy constraint is *slack* and profit is *constant* as κ increases (within this interval).

A consumer's action or message can be thought of as a noisy signal that the designer observes about the consumer's type. Ex ante, there is a distribution over these signals induced by the prior type distribution and each type's optimizing behavior. An alternative, more permissive notion of privacy would require that *on average*, the reduction in the designer's uncertainty regarding the consumer should not exceed some threshold. More formally, one could impose that the *expected* relative entropy between the designer's posterior and prior beliefs should be at most κ. This *ex-ante* notion of privacy, in contrast to the ex-post notion discussed previously, is lenient to events of high loss of privacy that occur with small probabilities. Thus, the designer can preserve privacy in a differential manner across consumer types, so long as on average a given level of privacy is maintained. This allows the designer to take advantage of the fact that some consumers' private information may be more valuable than others, e.g., uncovering "high valuation types" may be more profitable than uncovering "low valuation types."

We show that many of our findings continue to hold under the ex-ante notion of privacy (restricting again to static mechanisms). However, there are a number of major differences. First, even though it is again without loss to consider mechanisms that involve interval partitions, *existence* of an optimal mechanism is no longer straightforward. The difficulty arises since the ex-ante constraint, unlike our original ex-post version, allows for partitions with countably many intervals. Without a finite upper bound on the number of intervals, it is more challenging to use a compactness argument to prove existence. To show that an optimal mechanism indeed exists, we show that there can be at most *one* interval of arbitrarily small mass, i.e., at most one set of types about which the seller attains very precise information. This property restores compactness and settles the existence issue; it also implies that for κ small (near full privacy), the optimal ex-ante κ-constrained mechanism has exactly two intervals. Another distinction

of the ex-ante constraint is that the resulting optimal mechanism always exhausts the privacy constraint. Consequently, the optimal mechanism in the uniform-quadratic case is different from the one derived for ex-post privacy: For $\kappa \in (\log(n-1), \log(n)]$, there is exactly one "short" interval and $n-1$ "long" intervals of equal length; the lengths are uniquely determined by the binding privacy constraint, and the position of the "short" interval does not matter.

In the absence of a commonly-agreed-upon notion of privacy loss, our main contribution is to propose a Bayesian definition that builds upon a familiar concept from information theory, which has already been adopted by economists as a measure of the cost of information. This privacy notion can be easily incorporated into the standard mechanism design framework, and has the advantage of allowing to quantify the effect of a marginal change in the privacy threshold on profits and welfare. As our results suggest, the proposed privacy notion also provides a rationale for using "simple/coarse" mechanisms with restricted message spaces (we expand on this in the next section).

We proceed as follows. In Section 2 we survey the related literature. In Section 3 we formally describe the model and our ex-post Bayesian privacy notion, and in Section 4 we analyze optimal privacy-constrained mechanisms using this notion. Section 5 presents an alternative, ex-ante, Bayesian privacy measure and solves for the corresponding optimal mechanisms under this constraint. Section 6 discusses the potential loss of generality in restricting to static mechanisms, as well as the challenges in extending our framework to multi-agent interactions. Section 7 concludes.

2 Related literature

Our notion of privacy differs from the popular measure of "differential privacy" that is often used in the computer science literature.[8] Introduced by Dwork et al. (2006), differential privacy roughly means that changing the data of a single individual (or of a single attribute of an individual) should have a negligible effect on computations done on the entire data. In the context of mechanism design, Pai and Roth (2013) show that this notion can be formalized as follows. Suppose there are n individuals, who each draws a private type from some set T. Define a mechanism M as a mapping from profiles of types $t \in T^n$ to distributions over some set of outcomes X. Then M is ϵ-*differentially private* if for all pairs of type profiles (t, t') that differ only in t_i, and for any payoff function

[8] There are many papers in this literature, but we will be able to mention only a few of them. For more detailed surveys on privacy in computer science and economics, see Pai and Roth (2013), Heffetz and Ligett (2014) and Acquisti et al. (2016).

$u : X \to \mathbb{R}$, it holds that

$$\mathbb{E}_{M(t)} u(x) \leq \exp(\epsilon) \cdot \mathbb{E}_{M(t')} u(x).$$

This definition implies that the message of a single player has a negligible effect on the outcome, such that any action is "almost" weakly dominant (in the sense that it cannot affect a player's payoff by a multiplicative factor of more than 2ϵ, regardless of the other players' actions). In light of this, several studies in computer science have used the above notion to design mechanisms where truth-telling is either almost or exactly weakly dominant; see, e.g., McSherry and Talwar (2007), Kearns et al. (2012), and Nissim et al. (2012).

What distinguishes our approach is that we measure privacy loss *relative* to a prior, whereas differential privacy is independent of the prior. Additionally, note that the differential privacy constraint is stated solely in terms of the (random) outcome distribution, sidestepping any action taken in the mechanism. This captures privacy loss from the perspective of an *outside observer*, but does not correspond to what the *designer* may learn through observing participants' choices. As we discussed in the Introduction, the two can be significantly different, and our focus is on the latter (what designer learns).

A second line of literature in computer science deals with distortion and anonymization of databases and communication channels due to privacy concerns. This literature uses KL divergence to measure loss of privacy in a setting where a sender wants to send part of a dataset to a receiver in such a way that the receiver learns as little as possible about some other sensitive part of the data. The problem is formulated as finding the message that minimizes the KL divergence between the common prior on the sensitive data and the posterior, given the sender's message, subject to the constraint that the receiver attains some level of utility. Examples of works in this literature include Agrawal and Aggarwal (2001), Díaz et al. (2003), Rebollo-Monedero et al. (2010), Sankar et al. (2013) and Wang et al. (2016). However, these papers do not consider the strategic interaction between privacy, mechanism and agent behavior, as we do here.

Another line of research has proposed ways of incorporating agents with privacy concerns into a mechanism design framework. The literature has mostly assumed that each agent incurs an additive cost for loss of privacy, where this cost increases with the level of differential privacy (i.e., with the ϵ above). Some notable examples of these studies are Ghosh and Roth (2011), Ligett and Roth (2012), Fleischer and Lyu (2012). More recently, Gradwohl (2018) studies the problem of full implementation when agents prefer to protect their privacy, and Gilboa-Freedman and Smorodinsky (2018) show that

an agent's preferences over privacy satisfy some natural properties if and only if the agent ranks mechanisms according to a divergence measure between the distributions over signals that his different types induce.[9] In contrast, our goal is to study the design of profit-maximizing mechanisms under the privacy constraint.

Within economics, a number of papers have studied privacy in dynamic models, where the agent's preferences for privacy is derived from how the designer can use the information against her in the future. Such models appear in Taylor (2004), Calzolari and Pavan (2006), Conitzer et al. (2012) among others. In contrast, our approach based on the divergence between prior and posterior beliefs is *reduced-form* and applicable to settings where it is not a priori clear what the seller would/could do with the collected consumer data. We note that in our model, the agent reveals information about her type by responding to the mechanism. Recent works of Hidir and Vellodi (2018) and Ichihashi (2019) have also considered letting the consumer directly communicate to the seller via cheap talk or Bayesian persuasion.

The privacy constraint in our model entails that, *in equilibrium*, agents cannot communicate all their private information to the designer. Several papers have investigated a related question of optimal mechanism design with limited communication, by imposing *exogenous* restrictions on the cardinality of the action space available to agents. Notable examples are Green and Laffont (1987), Melumad et al. (1992), Blumrosen et al. (2007), Bergemann et al. (2012), Kos (2012) and Blumrosen and Feldman (2013). In a different setting, Mookherjee and Tsumagari (2014) study a dynamic mechanism design problem with costly communication and compare between centralized and decentralized production decisions. Van Zandt (2007), Fadel and Segal (2009) and Babaioff et al. (2013) study the interaction between communication capacity and incentive feasibility by quantifying the "cost of selfishness" – the amount of excess information (bits) that needs to be exchanged to implement a given social choice function, relative to the case in which agents honestly report their types. More generally, optimization problems over partitions have been considered in Alonso and Matousheck (2008) and Frankel (2014), who study delegation, as well as in Crémer et al. (2007), who study language design within an organization.

Finally, our work is related to the growing literature on information costs that take the form of relative entropy or mutual information. Pomatto et al. (2019) recently provide

[9] Also related is Ollár, Rostek and Yoon (2016), who analyze uniform pricing rules in markets with multiple traders facing Gaussian information structures. Their focus is on the information that traders may learn about *each other* from any feedback provided by the mechanism. They show that in their environment, privacy – in the sense that no trader learns anything about the other traders' valuations – is necessary for truthful bidding.

an axiomatization of relative entropy as a measure of information cost. In the rational inattention literature (see, e.g., Sims (2003), Matějka and McKay (2015), Maćkowiak and Wiederholt (2015) and Matějka (2016)), an uninformed decision maker chooses the structure of a signal he wants to observe, subject to the information constraint that the signal can only contain a limited amount of information about the state (as measured by mutual information). Note that mutual information is equal to the expected KL divergence between the decision maker's posterior and prior beliefs. Thus, the rational inattention approach relates to our mechanism designer's problem under the ex-ante privacy constraint. The key difference is that in our setting, the "signal" observed is the consumer's message, which is bound by an additional incentive constraint. Studying the interaction between the information constraint and the incentive constraint is the main objective of our work.

3 The framework

We consider the classic Mussa-Rosen (1978) set-up of monopolistic screening. A seller wishes to sell some quantity/quality $q \in \mathbb{R}^+$ to a buyer, in exchange for payment $p \in \mathbb{R}$. The seller's profit is given by:

$$\pi(p, q) = p - c(q)$$

where $c(\cdot)$ is a twice-continuously differentiable cost function that satisfies $c(0) = c'(0) = 0$ and $c''(q) > 0$ for all $q \geq 0$. The buyer's willingness to pay per unit is $\theta \in \Theta = \left[\underline{\theta}, \overline{\theta}\right]$, and is unknown to the seller. If the buyer consumes q and pays p, his utility is

$$u(p, q, \theta) = q \cdot \theta - p$$

The seller's prior probability distribution on θ is F, which has support Θ and density $f > 0$. We assume that the buyer's virtual valuation, $v(\theta) \equiv \theta - \frac{1-F(\theta)}{f(\theta)}$, is increasing in θ and satisfies $v(\underline{\theta}) \geq 0$. To facilitate some technical arguments, we make the slightly stronger assumption that v is continuously differentiable and $v' > 0$.

Positive virtual valuation allows us to focus on the case in which the seller wants to include all buyer types, and the only question is what quantity/quality and price should be offered to each buyer type. Nonetheless, some of our results extend to the case where $v(\underline{\theta})$ is negative – in particular, the optimal privacy-constrained mechanism again partitions the type space into intervals (as we describe below in Proposition 1),

even though the lowest interval may correspond to no sale.[10]

To sell the good the monopolist devises a static mechanism $\mathbb{M} = \langle M, p, q \rangle$, where M is an arbitrary set of messages, and $p : M \to \mathbb{R}^+$ and $q : M \to \mathbb{R}^+$ are functions that map each message in M to an outcome: Given a message $m \in M$, the seller provides the quantity $q(m)$ and charges the price $p(m)$.

The seller's objective is to maximize his expected profit Π:

$$\Pi(\mathbb{M}) = \mathbb{E}_m [p(m) - c(q(m))]$$

where $\mathbb{E}_m$ is evaluated according to the probability that, given $\mathbb{M}$, each message $m \in M$ is sent by a utility maximizing buyer in equilibrium. A strategy for the buyer is a function $\sigma : \Theta \to \Delta(M)$.

In the absence of privacy constraints, an optimal (revenue maximizing) mechanism in this set-up is a direct revelation mechanism in which: (i) The agent truthfully reports his type θ, (ii) The produced quantity $q(\theta)$ is determined such that $c'(q(\theta)) = v(\theta)$, and (iii) The requested price is $p(\theta) = q(\theta)\theta - \int_{\underline{\theta}}^{\theta} q(x)\,dx$.

Remark: In Section 6 we show that in the presence of privacy constraints, restricting attention to static mechanisms may potentially entail some loss of generality in the following sense: There are incentive-compatible dynamic mechanisms in which the designer randomizes (over the menus of actions available to the agent in each stage) that induce a distribution over the set of outcomes assigned to each agent type, and a distribution over posterior beliefs about the type, which cannot be generated *jointly* by any incentive-compatible *static* mechanism. However, it remains an open question whether the *optimal* dynamic mechanism cannot be mimicked by static mechanisms.

3.1 Bayesian privacy

At the outset, the seller already has some information about the buyer: He knows the buyer's type is distributed according to F. When a buyer decides to participate in the mechanism and sends a message $m \in M$, the seller updates his information according to the posterior belief distribution $F(\cdot|m)$. This change of beliefs entails loss of privacy

[10] The results in Proposition 2 and Proposition 3 also continue to hold, so long as we "exclude" the interval with no sale. Specifically, Proposition 2 would say the following: If $c''' \geq 0$ and $v(\theta)$ is a strictly concave function of $F(\theta)$, then any κ-optimal mechanism consists of intervals ordered in increasing mass from left to right, except possibly for the lowest interval when its expected virtual valuation is negative (which corresponds to no sale). Similarly, Proposition 3 would say that in the uniform-quadratic case, any κ-optimal mechanism consists of intervals with equal lengths, except possibly for the lowest interval.

for the buyer. For convenience, we define $F(\cdot|m) = F$ for any message m sent with zero probability.

We measure the loss of privacy entailed by a message $m \in M$ by the *relative entropy* between the posterior belief triggered by m and the prior belief: If the posterior distribution $F(\cdot|m)$ has density $f(\cdot|m)$, the relative entropy (or Kullback-Leibler Divergence) from $F(\cdot|m)$ to F is defined by:[11]

$$D_{KL}(F(\cdot|m)\,||F) = \int_{\underline{\theta}}^{\overline{\theta}} f(\theta|m) \cdot \log \frac{f(\theta|m)}{f(\theta)}\, d\theta, \tag{1}$$

whenever $F(\cdot|m)$ admits a density w.r.t. Lebesgue measure.[12] If instead $F(\cdot|m)$ contains atoms, we define $D_{KL}(F(\cdot|m)\,||F) = +\infty$ to preserve continuity (in the weak topology). Throughout the paper, "log" represents the natural logarithm.

We define the *ex-post loss of privacy* entailed by a mechanism to be the *maximum* divergence between the possible posteriors and the prior:

Definition 1 *The ex-post loss of privacy entailed by mechanism* $\mathbb{M} = \langle M, p, q\rangle$ *is given by:*[13]

$$I(\mathbb{M}) = \sup_m \left[D_{KL}(F(\cdot|m) \;||\; F)\right]$$

4 Optimal privacy-constrained mechanisms

Suppose the seller has to design a mechanism that does not exceed some privacy capacity $\kappa > 0$. His problem can then be described as follows: Find a mechanism $\mathbb{M} = \langle M, p, q\rangle$ and a strategy σ for the buyer that maximize the expected profit $\Pi = \mathbb{E}_m\left[p(m) - c(q(m))\right]$ subject to three constraints:

---

[11] The relative entropy exhibits a number of key properties: $D_{KL}(G||F) \geq 0$ for all G and F with equality if and only if $G = F$, and $D_{KL}(G||F)$ is convex in both G and F. It is however not a metric due to the failure of symmetry and of the triangle inequality.

[12] Since we have assumed that F admits a density w.r.t. Lebesgue measure, the integral on the RHS of (1) can be evaluated whenever $F(\cdot|m)$ also admits a density.

[13] To be fully rigorous, we note that the loss of privacy as defined here may in general depend on equilibrium selection; so $I(\mathbb{M})$ should better be written as $I(\mathbb{M}, \sigma)$. However, multiple equilibria only arise when a positive measure of buyer types are indifferent between two messages, which must then lead to the same quantity-price pair. As we discuss below, such messages are "wasteful" and without loss excluded from the optimal mechanism. Hence we will omit the issue of multiplicity.

1. *Incentive-compatibility* - given $\mathbb{M}$, the strategy σ is optimal for the buyer:

$$u\left(p\left(m\right),q\left(m\right),\theta\right)\geq u\left(p\left(m'\right),q\left(m'\right),\theta\right) \tag{IC}$$

for all $\theta\in\Theta$, all $m\in supp\left(\sigma\left(\theta\right)\right)$ and all $m'\in M$,

2. *Individual-rationality* - given $\mathbb{M}$, a buyer who follows σ is not worse off than if he did not participate in $\mathbb{M}$:

$$u\left(p\left(m\right),q\left(m\right),\theta\right)\geq 0 \tag{IR}$$

for all $\theta\in\Theta$ and all $m\in supp\left(\sigma\left(\theta\right)\right)$,

3. *Privacy constraint* -

$$I\left(\mathbb{M}\right)\leq\kappa. \tag{P}$$

We refer to any mechanism that satisfies the above constraints as a *κ-feasible mechanism*.[14] Any mechanism that is profit-maximizing among all κ-feasible mechanisms is called a *κ-optimal mechanism*. Our objective is to derive key properties of this constrained-optimal mechanism. In particular, we are interested in addressing the following questions: What information does each buyer type disclose to the mechanism (Proposition 1)? Do some buyer types disclose more information than others (Proposition 2)? Is the privacy constraint even binding (Proposition 3 and ensuing discussion)?

4.1 Interval mechanisms

Note that in standard mechanism design the monopolist maximizes his expected profit subject only to the incentive-compatibility and individual-rationality constraints. Under our assumptions on the virtual valuation, the optimal mechanism in this case perfectly screens every buyer type, and each of the posterior beliefs is a degenerate distribution with a single atom on the buyer's exact type. The loss of privacy entailed by such a mechanism is infinite according to our definition, and is therefore infeasible for any finite κ. This means that in a κ-optimal mechanism the monopolist obtains only a *noisy* signal about the buyer's type. Our first result establishes that this noise has a particular structure, which can be interpreted as a *coarse* revelation principle: There is no loss of generality in focusing on mechanisms that partition the type space into intervals and each type reports the interval he belongs to.

[14] How relevant the value of κ is for feasibility depends on the prior, as does the KL divergence measure.

Lemma 1 *For any κ-feasible mechanism, there exists another κ-feasible mechanism $\mathbb{M} = \langle M, p, q\rangle$ with the same profit level, such that M consists of intervals that partition $[\underline{\theta}, \overline{\theta}]$, and each type $\theta \in \Theta$ reports the message $m \in M$ for which $\theta \in m$.*

For future reference, we call such mechanisms as described in the lemma "interval mechanisms."

The intuition for this result is as follows. First, messages that lead to the same quantity-price pair can be combined without affecting the outcome of the mechanism. Due to convexity of the relative entropy function, this also relaxes the privacy constraint. We can thus without loss assume that different messages in the mechanism are strictly ranked by the quantity levels they are mapped into. Next, we argue that the set of types selecting the same message in equilibrium must form an interval; this is because buyer preference is single crossing in the type and quantity served. Finally, any two of these intervals do not intersect each other except at the boundary, since the indifference condition holds for at most one type. The lemma then follows (see the appendix for the formal proof). We mention that the same argument (and result) holds even if the designer's objective is to maximize a weighted sum of profit and consumer surplus.

Next, we derive the privacy loss of any interval mechanism. Note that when the seller sees a message $m = [\theta', \theta'']$ in equilibrium, his posterior density updates to $f(\theta \mid m) = \frac{f(\theta)}{F(\theta'') - F(\theta')}$ for $\theta \in [\theta', \theta'']$, and $f(\theta \mid m) = 0$ otherwise. The relative entropy between this posterior belief and the prior is computed as

$$\int_{\theta'}^{\theta''} f(\theta \mid m) \log \frac{f(\theta \mid m)}{f(\theta)} d\theta = -\log\left[F(\theta'') - F(\theta')\right]. \tag{2}$$

Since the ex-post privacy constraint requires this to be at most κ, we derive the following result:[15]

Lemma 2 *In any κ-feasible interval mechanism, each interval has mass (according to the prior F) at least $e^{-\kappa}$, and thus there are at most e^{κ} intervals.*

Assuming now that the intervals in M are $m_1 = [\theta_0, \theta_1], m_2 = [\theta_1, \theta_2], \ldots, m_n = [\theta_{n-1}, \theta_n]$, with $\underline{\theta} = \theta_0 < \theta_1 < \cdots < \theta_n = \overline{\theta}$. Given the number n and cutoffs $\theta_1, \ldots, \theta_{n-1}$, we can derive the optimal quantity and price that a κ-optimal mechanism assigns to

[15] A similar result holds even if KL divergence is replaced by a more general divergence of the form $\int_\theta \phi(\frac{f(\theta|m)}{f(\theta)}) dF(\theta)$ for some convex function ϕ, with KL being the special case where $\phi(x) = x \log x$. Our subsequent results also generalize to this case, so long as the capacity constraint κ is suitably scaled according to the function ϕ.

each message. Specifically, note that the expected profit for the seller from employing this interval mechanism is given by:[16]

$$\Pi(\mathbb{M}) = \sum_{i=1}^{n} \left[q(m_i) \int_{\theta_{i-1}}^{\theta_i} v(\theta) f(\theta) d\theta - c(q(m_i)) \cdot [F(\theta_i) - F(\theta_{i-1})] \right], \tag{3}$$

where $v(\theta)$ is the virtual valuation of type θ. Therefore, the quantity that maximizes the expected profit while maintaining IC and IR is uniquely determined by:[17]

$$c'(q(m_i)) = \mathbb{E}_F [v(x) \mid x \in [\theta_{i-1}, \theta_i]] \qquad \text{for any } m_i = [\theta_{i-1}, \theta_i] \in M. \tag{4}$$

The standard envelope condition (derived from local IC) for buyer surplus also pins down the requested price:

$$p(m_i) = q(m_i) \cdot \theta_{i-1} - \sum_{j=0}^{i-1} (\theta_j - \theta_{j-1}) \cdot q(m_j) \qquad \text{for any } m_i = [\theta_{i-1}, \theta_i] \in M. \tag{5}$$

It follows that the assignment of types to quantity-price pairs in any κ-optimal mechanism is completely determined by the interval partition. To save notation, we will write q_i and p_i in place of $q(m_i)$ and $p(m_i)$ whenever the partition is fixed by the context.

Summarizing the above discussion, we have reduced the problem to finding the profit-maximizing interval partition (where profit follows from Equations (3) and (4)) subject to each interval having mass at least $e^{-\kappa}$.

4.2 Characterization

With the above preliminary analysis, we can show that a κ-optimal (interval) mechanism exists. Indeed, by Lemma 1, it is sufficient to show there exists a profit-maximizing interval partition. For this we apply a compactness argument. Consider the following metric on the space of finite partitions: If M consists of cutoffs $\{\theta_1, \ldots, \theta_{n-1}\}$ and M'

[16] To see this, recall that in every mechanism that satisfies (local) IC and binds IR at the lowest type, the seller's profit is given by $\Pi(\mathbb{M}) = \int_{\underline{\theta}}^{\overline{\theta}} \left[\tilde{q}(\theta)\theta - \int_{\underline{\theta}}^{\theta} \tilde{q}(x)\, dx - c(\tilde{q}(\theta)) \right] f(\theta)\, d\theta$, where $\tilde{q}(\theta)$ is the quantity provided to type θ. The first term in the integrand is the social surplus generated by selling quantity $\tilde{q}(\theta)$ to type θ, the second term is the minimal information rent that is left with type θ in every IC mechanism, and the third term is the cost of producing $\tilde{q}(\theta)$. The seller is the residual claimant of welfare. Equation (3) is obtained from this formula using integration by parts.

[17] Since c is strictly convex and $c'(0) = 0$, the first order condition (4) uniquely determines the value of the optimal $q(m_i)$. The fact that virtual valuation is increasing ensures that higher types receive higher quantity in equilibrium, and thus local IC implies global IC.

consists of cutoffs $\{\theta'_1, \ldots, \theta'_{m-1}\}$, then define $d(M, M')$ to be the smallest $\delta \geq 0$ such that for each θ_i there exists θ'_j within δ distance from it, and vice versa. It is straightforward to check this is indeed a metric, and that the resulting quantities q_i and profit $\Pi(\mathbb{M})$ are continuous with respect to this metric.

When the number of cutoffs is bounded above by e^κ (as ensured by Lemma 2), the space of partitions is compact in the topology induced by this metric. Thus, if we consider any sequence of feasible partitions (with each interval having mass at least $e^{-\kappa}$) that *approximates* the supremum profit, there exists a convergence subsequence. The limit partition is also feasible under the ex-post privacy constraint, and by continuity it must *achieve* the supremum profit. Hence we have found an optimal interval mechanism.

Proposition 1 *There exists a κ-optimal mechanism $\mathbb{M} = \langle M, p, q \rangle$, such that M consists of at most e^κ intervals (each with mass $\geq e^{-\kappa}$) that partition $[\underline{\theta}, \overline{\theta}]$, and each type $\theta \in \Theta$ reports the interval to which it belongs.*

An immediate corollary is the following:

Corollary 1 *For $0 < \kappa < \log(2)$, any κ-optimal interval mechanism involves perfect pooling. For $\log(2) \leq \kappa < \log(3)$, any κ-optimal interval mechanism consists of exactly two intervals.*

The first part follows from observing that for a vert small κ, the number of intervals is at most $e^\kappa < 2$, hence, only a single interval can be accommodated. As for the second part, $\kappa \in [\log(2), \log(3))$ means there can be at most two intervals. Moreover, by Lemma 4 in the appendix, the seller always benefits from having more information in the form of dividing an interval into two subintervals (so long as the resulting partition is still feasible). In particular, any feasible two-interval partition leads to higher profit compared to the trivial partition, leading to the above result.

In general, it is difficult to fully characterize the optimal number n of intervals and their cutoffs for an arbitrary distribution F and cost function c, without imposing additional structure on these primitives. The following result provides a step toward that characterization. It shows that under certain conditions on F and c, the optimal mechanism involves more pooling for higher (or lower) types:

Proposition 2 *Suppose that the cost function $c(\cdot)$ has non-negative third derivative, and that the virtual valuation $v(\theta)$ is a strictly concave function of $F(\theta)$.*[18] *Then any κ-optimal mechanism consists of intervals that are ordered in increasing mass from left to*

[18] The result also holds if $c(\cdot)$ has strictly positive third derivative and $v(\theta)$ is a concave function of $F(\theta)$. That is, we only need one of the two conditions to be strict.

right; that is, $F(\theta_{i+1})-F(\theta_i) \geq F(\theta_i)-F(\theta_{i-1})$ *for all* $i \in \{1,\ldots,n-1\}$*. Symmetrically, the intervals in the optimal mechanism would be ordered in decreasing mass if* $c''' \leq 0$ *and* $v(\theta)$ *is a strictly convex function of* $F(\theta)$*.*

We mention that a sufficient condition for $v(\theta)$ to be a concave function of $F(\theta)$ is that the prior density function $f(\theta)$ is increasing and log-concave; in fact, $F(\theta)$ would be convex and $v(\theta)$ concave.[19] For example, this is satisfied when $F(\theta) = \theta^r - s$ with $\underline{\theta} = s^{\frac{1}{r}}$ and $\overline{\theta} = (s+1)^{\frac{1}{r}}$, where r, s are parameters satisfying $r > 1$ and $s > \frac{1}{r}$ (the latter ensuring that virtual valuations are positive).[20]

The intuition for Proposition 2 is as follows. From $c'(q_i) = \mathbb{E}_F[v(x) \mid x \in [\theta_{i-1}, \theta_i]]$, we see that the second derivative c'' measures the extra production cost incurred when the seller divides an interval into two subintervals and adjusts quantity/price accordingly. So $c''' > 0$ means that the seller faces greater production cost in trying to screen the higher types. On the other hand, observe that the derivative of the virtual valuation captures the revenue gain when dividing an interval into two.[21] Thus $v(\theta)$ being a concave function of F implies that the seller benefits less from differentiating high types. Both effects combine to yield intervals that are optimally ordered in increasing mass.

4.3 Uniform-quadratic case

To give a complete and detailed characterization of an optimal κ-contrained mechanism, we need to select a specification of the distribution of types and of the cost function. Although such a characterization is only a particular example, it is still insightful as it gives a vivid illutration of the properties of optimal mechanism that we derived in the previous subsection. One specification that admits an elegant analytical characterization is the "uniform-quadratic" case, where F is uniform and c is quadratic. Since in this case $c''' = 0$ and $v(\theta)$ is *as convex as* $F(\theta)$ (both are linear in θ), the results in Proposition 2 suggest that the ordering of intervals does not matter for profit. As we show below, this property enables us to focus on the lengths of the intervals and obtain a full characterization of the optimal partition in this special case.

---

[19] We have $v''(\theta) = \left(\frac{f'(\theta)}{f(\theta)}\right)' \cdot \frac{1-F(\theta)}{f(\theta)} - \frac{f'(\theta)}{f(\theta)} \cdot \left(1 + \frac{f'(\theta)(1-F(\theta))}{f(\theta)^2}\right) < 0.$

[20] On the other hand, $v(\theta)$ would be more convex than $F(\theta)$ when $r \in [\frac{1}{2}, 1)$ in this example.

[21] This can be understood by considering the extreme case: When the virtual valuation $v(\theta)$ is constant, it is (unconstrained) optimal to offer the same quantity to every buyer type, and there is no gain from creating subintervals.

Proposition 3 *Suppose $F \sim U[\underline{\theta}, \overline{\theta}]$ with $\overline{\theta} \geq 2\underline{\theta}$ (so that $v(\underline{\theta}) \geq 0$), and $c(q) = \frac{q^2}{2}$. Then, given any positive integer N and any $\kappa \in [\log(N), \log(N+1))$, the κ-optimal mechanism divides the type space into N equal intervals.*

To prove this result, we show that maximizing profit in the uniform-quadratic case is equivalent to minimizing the *cubic sum* of the lengths of the intervals. Subject to the feasibility constraint that these lengths are at least $e^{-\kappa}$, we argue that an equal division (into as many intervals as can be accommodated) is optimal.

We highlight that the optimal mechanism in this case is locally the same as κ increases. Thus, for κ in the interior of the interval $(\log(N), \log(N+1))$, the optimal mechanism entails a privacy loss of $\log(N)$ which is strictly less than κ. It follows that the ex-post privacy constraint is often *slack*. This is one of the challenges in obtaining a full characterization of the optimal mechanism more generally.

4.4 Welfare analysis

Varying the privacy capacity of a mechanism affects the seller's profit, the buyer surplus and the total welfare (sum of profit and buyer surplus). In this section we analyze how κ changes these quantities. Throughout this section we assume F has monotone hazard rate, that is $\frac{f(\theta)}{1-F(\theta)}$ increases in θ. This property implies that the virtual valuation $v(\theta)$ is increasing.

Because higher κ relaxes the privacy constraint (P) in the seller's problem, it is immediate to notice that the expected profit Π is (weakly) increasing in κ.

Corollary 2 *Profit from a κ-optimal mechanism is increasing in κ.*

As the uniform-quadratic case shows, this monotonicity is not always (everywhere) strict. However, we do know that profit is maximized at $\kappa = \infty$, when the monopolist is allowed to fully separate all buyer types. Profit is minimized at $\kappa = 0$, with full privacy.[22]

For the buyer, the opposite is true since she is worse-off whenever the seller obtains finer information about her in the form of dividing an interval into two:

Proposition 4 *Suppose the cost function c satisfies $c''' \geq 0$. Then buyer surplus from a κ-optimal mechanism is maximized at $\kappa = 0$, where every type receives the same quantity, and it is minimized at $\kappa = \infty$, where types are fully separated.*

[22] In fact, profit is strictly maximized/minimized at the extreme κ, because profit strictly increases when an interval is divided into two subintervals (and quantities/prices are adjusted accordingly). This is proved in Lemma 4 in the appendix.

Finally, a regulator might be interested in finding the level of κ that maximizes total welfare. The answer turns out to be more subtle, as the following partial characterization suggests:

Proposition 5 *Assume quadratic costs $c(q) = \frac{q^2}{2}$. If the density $f(\theta)$ increases in θ. Then total welfare is maximized at $\kappa = 0$ and minimized at $\kappa = \infty$. Conversely, if $f(\theta)$ decreases in θ, then total welfare is minimized at $\kappa = 0$ and maximized at $\kappa = \infty$.*

As a corollary, it is interesting to note that in the uniform-quadratic case, total welfare in the κ-optimal mechanism is independent of the privacy capacity κ.

5 Ex-ante privacy-constrained mechanisms

So far we have analyzed an ex-post notion of privacy loss. Under this criterion, the monopolist cannot learn too much about *any* buyer type. In this section we explore an alternative, less stringent, notion of privacy, requiring the designer not to learn too much about the buyer type *on average*. Formally, we weaken the ex-post privacy constraint $I(\mathbb{M}) \leq \kappa$ to its ex-ante version:

Definition 2 *The ex-ante loss of privacy entailed by mechanism $\mathbb{M} = \langle M, p, q \rangle$ is given by*

$$I^{ea}(\mathbb{M}) = \mathbb{E}_m \left[D_{KL} \left(F(\cdot|m) \;||\; F \right) \right]$$

where $\mathbb{E}_m$ is evaluated according to the probability that each message $m \in M$ is sent in an equilibrium of $\mathbb{M}$.[23]

That is, we impose an upper bound on the *average* change in the seller's beliefs, as measured by relative entropy. Given $\kappa > 0$, we say a mechanism is *ex-ante κ-feasible* if $I^{ea}(\mathbb{M}) \leq \kappa$. It is *ex-ante κ-optimal* if it is profit-maximizing among all mechanisms $\mathbb{M}$ that satisfy IC, IR and the ex-ante privacy constraint.

5.1 Existence

To begin the analysis, we can use essentially the same argument to show that Lemma 1 also holds under the ex-ante criterion. Thus it is without loss to consider interval

[23] In calculating $I^{ea}(\mathbb{M})$ we adopt the convention that $0 \cdot \infty = 0$, and therefore $I^{ea}(\mathbb{M})$ can still be finite if there is a measure-zero set of messages (sent in equilibrium) that induce posterior distributions $F(\cdot|m)$ whose divergence from the prior F is infinite. But if the set of such messages has positive measure, then $I^{ea}(\mathbb{M}) = +\infty$ according to our definition.

mechanisms. Using Equation (2), we can compute the ex-ante loss of privacy entailed by a (finite) interval mechanism to be:

$$I^{ea}(\mathbb{M}) = \sum_{i=1}^{n} -\left[F(\theta_i) - F(\theta_{i-1})\right] \cdot \log\left[F(\theta_i) - F(\theta_{i-1})\right] = H(g_M), \tag{6}$$

where $g_M(m_i) = F(\theta_i) - F(\theta_{i-1})$ is a discrete distribution over the elements of M induced by the prior, and $H(\cdot)$ is the Shannon entropy function. At this moment we cannot rule out the possibility that there are countably many intervals. In that case, the discrete distribution g_M is defined in the same way, and we again have $I^{ea}(\mathbb{M}) = H(g_M)$.

This discussion suggests that under the ex-ante privacy constraint, the profit maximization problem reduces to finding an interval partition M subject to $H(g_M) \leq \kappa$ that maximizes the profit in Equation (3) with quantities given by (4).

However, existence of an optimal mechanism is not straightforward in this case. The reason is that the feasibility constraint $H(g_M) \leq \kappa$ allows for countably many intervals, which lose compactness (unlike with the ex-post constraint). We settle the existence issue in the following result:

Proposition 6 *There exists an ex-ante κ-optimal mechanism $\mathbb{M} = \langle M, p, q \rangle$, such that M consists of finitely many intervals that partition $[\underline{\theta}, \overline{\theta}]$, and each type $\theta \in \Theta$ reports the interval to which it belongs.*[24]

We provide a sketch of the proof here, leaving further details to Section 9.1. Consider a sequence of ex-ante κ-feasible interval mechanisms $\mathbb{M}_j = \langle M_j, p_j, q_j \rangle$ such that $\Pi(\mathbb{M}_j)$ converges to the *supremum* profit Π^* across feasible mechanisms. We will replace each mechanism $\mathbb{M}_j$ by another feasible interval mechanism $\tilde{\mathbb{M}}_j = \left\langle \tilde{M}_j, \tilde{p}_j, \tilde{q}_j \right\rangle$, such that the new message set $\tilde{M}_j$ consists of at most N intervals, where N *is a constant that depends only on F and κ*. This upper bound N restores compactness and allows us to find a subsequence of the partitions $\{\tilde{M}_j\}$ that converges to some limit partition $\tilde{M}_\infty$, under the metric defined in Section 4.2. By continuity, $\tilde{M}_\infty$ is also a feasible mechanism, and it achieves the limit profit along the convergent subsequence. Therefore, if we could carry out the replacement in such a way that $\Pi(\tilde{\mathbb{M}}_j) \geq \Pi(\mathbb{M}_j)$, then $\Pi(\tilde{M}_\infty) \geq \limsup_j \Pi(\tilde{M}_j) = \Pi^*$ and $\tilde{M}_\infty$ would be ex-ante κ-optimal.

[24] It is instructive to compare this result to an analogous result in the rational inattention literature. Matějka (2016) shows that a rationally inattentive seller would charge only finitely many prices even though there is a continuum of states. The argument used to prove that result relies on properties of Hermite polynomials. In contrast, the proof in our environment is simpler and only makes use of the tradeoff between privacy and profit when merging/dividing intervals.

It remains to find the appropriate replacements $\tilde{\mathbb{M}}_j$. As discussed, starting from any mechanism $\mathbb{M}_j$, merging two adjacent intervals in M_j into a single interval (and adjusting the quantities/prices accordingly) strictly decreases the profit. However, by doing so the seller is able to save on the ex-ante privacy measure, which enables him to divide any other interval in M_j into two subintervals, increasing the profit. The key argument, then, is to *compare the profit gain in the latter step to the profit loss in the former.* We show that whenever two adjacent intervals are both of mass smaller than some constant ϵ, they can be combined to create enough slackness in the privacy constraint; and if the slackness is used to break another (big) interval into two, the seller achieves a net profit gain. Intuitively, this profit comparison holds because the entropy function severely punishes against precise knowledge about any small set of types. So when the seller combines two "small" intervals into a single one, the saved privacy measure is significant relative to the reduction in profit.

By repeatedly combining adjacent "small" intervals, we are able to transform $\mathbb{M}_j$ into a mechanism $\tilde{\mathbb{M}}_j$ with weakly higher profit, and with no adjacent intervals both having mass $< \epsilon$. The upshot is that $\tilde{M}_j$ has at most $N := \frac{2}{\epsilon} + 1$ intervals, completing the proof.[25]

5.2 Extension/modification of other results

In contrast to the ex-post problem, the ex-ante privacy constraint always binds at the optimal mechanism. Moreover, unlike Lemma 2, the number of intervals now admits a *lower bound* of e^{κ}.

Proposition 7 *The privacy constraint is exhausted in any ex-ante κ-optimal mechanism $\mathbb{M}$; that is, $I^{ea}(\mathbb{M}) = \kappa$. Moreover, any optimal mechanism involves at least e^{κ} intervals.*

The first part holds because the seller always benefits from dividing an interval into two. By choosing one of the subintervals to be "small," the *average* privacy constraint is still satisfied. The second part is a result of the first part, Equation (6) and the following well-known estimate of the Shannon entropy (applied to g_M):

[25] To be fully rigorous, in the proof we first find a replacement with *finitely* many intervals. This can be done because for any limit point M_j (more precisely, the bounds of intervals in M_j) may have, the seller incurs little profit loss if he combines *all* the small intervals near this limit point. Such loss is covered by the net profit gain in merging two small intervals and dividing a long one. Once we have a finite M_j to begin with, we still need to guarantee that the process of "combining small intervals" will come to an end. We do this by combining *two pairs of* adjacent small intervals at once and breaking a big interval into two. There is still net profit gain, and in addition the total number of intervals strictly decreases. The final $\tilde{M}_j$ involves at most one pair of adjacent small intervals, so its size is again bounded uniformly across j.

(Cover and Thomas, Theorem 2.6.4) *If a discrete random variable X takes n values, then its Shannon entropy satisfies $H(X) \leq \log(n)$, with equality if and only if X has a uniform distribution.*

On the other hand, we show that when the ex-ante privacy constraint is stringent, two messages are sufficient to implement the optimal mechanism. This complements the previous Corollary 1.

Proposition 8 *There exists $\underline{\kappa} > 0$ such that any ex-ante κ-optimal interval mechanism consists of exactly two intervals.*

We prove the result via a lemma stating that there can be at most one interval with arbitrarily small mass. Thus, even though the average privacy constraint allows the seller to learn almost perfectly about sets of small types, the optimal solution involves at most one such set. In the above proof sketch for Proposition 6, we have already discussed why having two *adjacent* "small" intervals is suboptimal. The next lemma additional rules out the presence of two "small" intervals that are non-adjacent.

Lemma 3 *For every $k > 0$, there exists $\epsilon > 0$ such that any ex-ante κ-optimal interval mechanism with $\kappa \leq k$ has at most one interval with mass $< \epsilon$.*

Finally, Proposition 2, which gives sufficient conditions for the optimality of intervals that are increasing/decreasing in mass, as well as the welfare results Propositions 4 and 5, continue to hold under the ex-ante privacy measure. The proofs of these results extend without change.

5.3 Uniform-quadratic case revisited

Recall that in the uniform-quadratic case, the solution under the *ex-post* constraint divides the type space evenly. Since this solution does not always exhaust the (ex-ante or ex-post) privacy constraint, we know from Proposition 7 that profit is not maximized under the ex-ante constraint. Indeed, the optimal mechanism under the ex-ante privacy constraint differs from the optimal mechanism under the ex-post constraint. Moreover, it has novel implications regarding the trade-off between privacy and profit, which does not arise under the ex-post constraint. In light of this, we revisit the uniform-quadratic example to further illustrate the difference between the two privacy notions.

Proposition 9 *Suppose $F \sim U[\underline{\theta}, \overline{\theta}]$ with $\overline{\theta} \geq 2\underline{\theta}$, and $c(q) = \frac{q^2}{2}$. Then, given any positive integer $N > 1$ and any $\kappa \in (\log(N-1), \log(N)]$, the ex-ante κ-optimal mechanism divides the type space into N intervals.*

Specifically, $N-1$ of these intervals have the same lengths ℓ_b, while the last interval has weakly smaller length ℓ_s. These two lengths are uniquely determined by the total length $\overline{\theta}-\underline{\theta}$ and the binding privacy constraint:

$$\begin{aligned}\overline{\theta}-\underline{\theta} &= \ell_s + (N-1)\,\ell_b \\ \kappa &= -\frac{\ell_s}{\left(\overline{\theta}-\underline{\theta}\right)}\log\frac{\ell_s}{\left(\overline{\theta}-\underline{\theta}\right)} - (N-1)\cdot\frac{\ell_b}{\left(\overline{\theta}-\underline{\theta}\right)}\log\frac{\ell_b}{\left(\overline{\theta}-\underline{\theta}\right)}\end{aligned}$$

The optimal interval partition is unique up to reordering of the intervals.

The proof consists of three steps. First, we show that the expected profit (and privacy measure) only depends on the lengths of the intervals in the partition, and not on the ordering of these intervals. Next, we argue that the first- and second-order conditions for the constrained maximization problem can only be satisfied if the intervals have at most two lengths, with exactly one interval having shorter length. Lastly, we determine the optimal *number* of intervals from the binding privacy constraint.

The structure of this optimal mechanism has an interesting implication for the trade-off between privacy and profit. For $\underline{\theta}=1$ and $\overline{\theta}=2$ in the uniform-quadratic case, Figure 1 depicts the expected profit of the monopolist in the ex-ante κ-optimal mechanism as a function of κ.

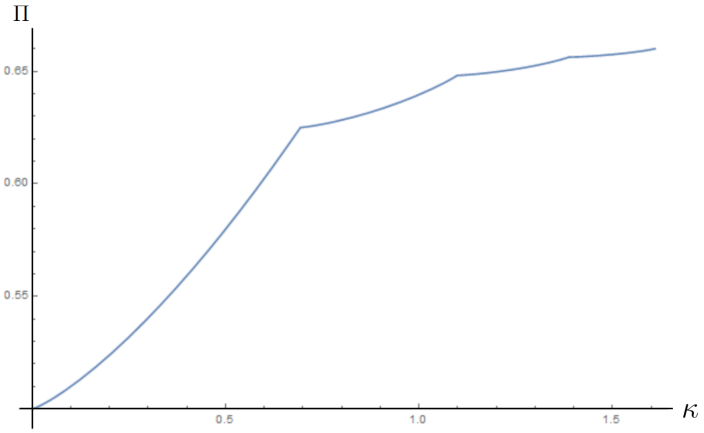

Figure 1. The privacy-profit frontier in the uniform-quadratic case

The kinks in Figure 1 represent values of κ where the number of intervals in the κ-optimal mechanism increases. Between kink points the number of intervals remains fixed but the intervals change. Notice that while there are *diminishing* returns to loss of privacy when the number of intervals increase, there are *increasing* returns to loss of privacy when

κ increases but the number of intervals remains fixed (that is, the curve between kink points is convex). This means that when we introduce a new (small) interval, the initial change in expected profit is small relative to the loss of privacy; intuitively, the ex-ante privacy measure punishes against precise information about a small set of types. But as we continue to lower privacy, expected profit rises at an increasing rate until a new interval is added.

We mention that the qualitative features of the privacy-profit frontier are robust to small changes in the prior distribution (away from uniform). This is because the set of ex-ante κ-optimal mechanisms, when viewed as a correspondence from the distribution F to the space of interval partitions, is *upper-hemicontinuous*.[26]

6 Discussion

6.1 Revelation principle

The revelation principle refers to the idea that the mechanism design problem can often be simplified without losing any generality by restricting attention to mechanisms with two properties: (1) each agent reports his type, and (2) the mechanism is one-shot. The first property clearly fails in the presence of privacy concerns, but we partially restore it with our notion of "coarse revelation" (reporting the interval that contains the agent's type). The second property holds when the designer is restricted to sequential mechanisms in which he cannot over the menus of actions available to the agent.[27] However, if such randomization is allowed, then there are incentive-compatible dynamic mechanisms that induce a distribution over the set of outcomes assigned to each agent type, and a distribution over posterior beliefs about the type, which cannot be generated *jointly* by any incentive-compatible *static* mechanism.

To illustrate this, suppose types are uniformly distributed on $[0, 1]$. Consider a dynamic mechanism in which nature first flips a coin. If heads comes up, the agent is asked to choose between no transaction, or purchasing one unit at price 0.4. If tails comes up, the agent faces a similar choice, but with price equal to 0.6. The type-dependent outcome distribution under this mechanism is as follows: Types in $[0, 0.4]$ get $(q, p) = (0, 0)$ with certainty. Types in $[0.4, 0.6]$ get $(0, 0)$ with probability $1/2$ and $(1, 0.4)$ with probability

[26] This follows from the Maximum Theorem, since the seller solves a constrained optimization problem in which the objective function (i.e., profit) is continuous in the interval partition (with respect to the metric defined in Section 4.2), and the constraint $H(g_M) \leq \kappa$ is both upper- and lower-hemicontinuous.

[27] But the agent is still allowed to randomize over actions that he deems indifferent.

1/2. Finally, types in $[0.6, 1]$ get a 50-50 lottery over $(1, 0.4)$ and $(1, 0.6)$.

Suppose we wanted to replicate this outcome using a static mechanism. Clearly any message sent by types below 0.4 must lead to the outcome $(0, 0)$. In addition, the possibly mixed strategies of types in $[0.4, 0.6]$ need to generate a 50-50 lottery over $(0, 0)$ and $(1, 0.4)$. Thus, any message sent by a type in $[0.4, 0.6]$ leads to some lottery over $(0, 0)$ and $(1, 0.4)$. But note that every type in $(0.4, 0.6]$ strictly prefers $(1, 0.4)$ to $(0, 0)$. So each of these types would choose the message that places the highest probability on $(1, 0.4)$. Therefore, the only way to generate the above type-dependent distribution over outcomes is for each type in $[0.4, 0.6]$ to send a message (or mix over identical messages) that gives a 50-50 lottery over $(0, 0)$ and $(1, 0.4)$. Similarly, each type above 0.6 must send a message that gives a 50-50 lottery over $(1, 0.4)$ and $(1, 0.6)$.

This analysis implies that in the static mechanism, types in the three different intervals $[0, 0.4]$, $[0.4, 0.6]$ and $[0.6, 1]$ send *distinct* messages. As a corollary, the ex-post privacy loss of the static mechanism is at least the privacy loss from knowing the interval $[0.4, 0.6]$, which is $\log(5)$. In contrast, the above dynamic mechanism has the property that with probability 1/2 the designer learns whether type is above 0.4, and with probability 1/2 he learns whether type is above 0.6. The resulting ex-post privacy loss is $\log(2.5)$, as derived from the interval $[0, 0.4]$ (and from $[0.6, 1]$). Hence, any static mechanism that replicates the type-dependent outcome distribution of the dynamic mechanism must entail greater privacy loss.

We comment that although this example illustrates the failure of the revelation principle, it does not imply that using dynamic mechanisms is necessarily profitable for the seller. In particular, the above dynamic mechanism can be viewed as a randomization over two static mechanisms, which means that its profit is no higher than the best static mechanism. It remains an open question whether *optimal* dynamic mechanisms can be mimicked by static mechanisms in terms of the induced type-dependent outcome distribution as well as privacy loss.

6.2 Multiple agents

Extending our analysis to mechanisms with more than one agent presents a number of challenges. First, the notion of privacy loss needs to be extended to accommodate the possibility that different participants are exposed to different losses of privacy.[28] One approach is to require that the maximal loss of privacy for any agent is at most κ; an

[28] This is particularly important since the literature on optimal mechanisms with restricted message spaces has highlighted the usefulness of asymmetric mechanisms; see Kos (2012).

alternative is to measure the average loss of privacy across all agents. As in our single-agent model, the privacy notion also has to address the fact that loss of privacy may differ across types of the same agent.

The second challenge concerns the failure of the revelation principle. Such failures are more significant with multiple agents, since even without randomization, a sequential mechanism may preserve more privacy by collecting information from a small number of agents. For example, a descending price auction is strategically equivalent to a sealed-bid first price auction, but collects information only about the winning bidder.

Aside from these challenges, our framework can be extended to allow for multiple agents. To illustrate, we analyze here the simple case of a seller with a single unit of a good and no production costs, and m buyers who independently draw private valuations for the good from a uniform distribution over $\left[\underline{\theta}, \overline{\theta}\right]$, where $\underline{\theta} \geq \frac{1}{2}\overline{\theta}$ ensuring that virtual valuations are non-negative. We restrict attention to *symmetric, static* mechanisms and require that each agent's ex-post (with respect to types) privacy loss be at most κ.

By essentially the same arguments as in our single-agent model, it can be shown that the optimal privacy-constrained mechanism partitions the set of types into finitely many intervals. In light of this, we consider the class of mechanisms where the types are partitioned into intervals, each buyer reports the interval to which his type belongs, and the bidder with the highest report is awarded the good (with ties broken evenly). The optimal mechanism within this class is given in the following result:

Proposition 10 *Suppose there are $m \geq 5$ buyers with uniformly distributed values, and no production costs. Then for any $\kappa \in [\log(n), \log(n+1))$, the optimal ex-post privacy-constrained symmetric auction partitions the type space into n intervals. Each of the upper $n-1$ intervals has equal mass of $e^{-\kappa}$ (which is privacy-binding), and the lowest interval has the remaining mass of $1-(n-1)e^{-\kappa}$ (which is weakly greater).*

Our proof in the appendix also shows that with 2 buyers, the optimal auction partitions the set of types into n equally long intervals, which coincides with the solution for a single buyer under quadratic costs. The solutions to 3 or 4 buyers can be derived by similar analysis, but the statements are more complex and thus omitted.

7 Concluding remarks

This paper proposed a Bayesian approach to incorporating privacy constraints into mechanism design. The underlying idea is that the designer *already* has some prior information about the participants, and the loss of privacy induced by a mechanism should be

measured as the *difference* between this prior information and the updated information that can be inferred from the agents' interaction with the mechanism. This entails an additional constraint - on top of the standard incentive-compatibility and individual-rationality constraints - that needs to be satisfied by a mechanism: The difference between the prior and posterior information must be below some threshold.

We illustrate this approach by using relative entropy to compute the difference between the prior and posterior beliefs and applying this measure to a canonical monopolistic screening problem. We show the implications of imposing the privacy constraint at the ex-post stage (i.e., for every realized consumer type, the loss of privacy must be below some bound), and at the ex-ante stage (i.e., the loss of privacy is bounded when averaging over possible type realizations). We also demonstrate how our framework can be helpful in understanding the effect of privacy constraints on consumer and seller welfare.

Our approach opens the door to many interesting questions about mechanism design and privacy. In particular, since the revelation principle can fail, what is the optimal mechanism when we allow for sequential mechanisms with randomization? What are optimal privacy-preserving auctions? Finally, since we follow the standard Bayesian approach to mechanism design, we inherit its assumption that the prior distribution is commonly known. How should a regulator elicit the prior beliefs of the designer in practice? We hope that future research will provide answers to these and related questions.

8 Appendix 1 - Proofs for the main model

8.1 Preliminaries

We define two auxiliary functions. First, we define $\phi(x)$ as the inverse of c', i.e., $c'(\phi(x)) = x$ for all $x \geq 0$. Convexity of the cost function and $c'(0) = 0$ ensures that ϕ is uniquely defined and increasing. In fact, by the chain rule we have

$$\phi'(x) = \frac{1}{c''(\phi(x))}.$$

Since $c''(q)$ is positive and continuous for $q \geq 0$, we deduce that $c''(\phi(x))$ is bounded above and away from zero for $x \in [v(\underline{\theta}), v(\overline{\theta})]$.

Next, we define the function $h(x)$ as follows:

$$h(x) = \phi(x) \cdot x - c(\phi(x))$$

The first derivative of $h(x)$ is given by:

$$h'(x) = \phi'(x) \cdot x + \phi(x) - c'(\phi(x)) \cdot \phi'(x) = \phi'(x) \cdot x + \phi(x) - x \cdot \phi'(x) = \phi(x)$$

Thus the second derivative h'' is bounded above and away from zero for $x \in [v(\underline{\theta}), v(\overline{\theta})]$.

The following lemma provides an estimate of the seller's profit gain when an interval is divided into two subintervals. It will be used in subsequent proofs.

Lemma 4 *There exists a small positive constant η depending on F and $c(\cdot)$, such that for any triple of cutoffs $a < b < c$, the profit loss Δ incurred when merging the two intervals $[a, b]$ and $[b, c]$ into a single interval $[a, c]$ (and adjusting quantities/prices accordingly) satisfies*

$$\eta \leq \frac{\Delta}{(F(b) - F(a))(F(c) - F(b))(F(c) - F(a))} \leq \frac{1}{\eta}.$$

Note from Equations (3) and (4) that Δ only depends on a, b, c and is independent of the remaining cutoffs.

Proof: Let $\mathbb{E}v(m_i)$ denote $\mathbb{E}_F[v(\theta) \mid \theta \in [\theta_{i-1}, \theta_i]]$. Then the profit of mechanism $\mathbb{M} = \langle M, p, q \rangle$ as given by Equations (3) and (4) can be rewritten as

$$\Pi(\mathbb{M}) = \sum_i h(\mathbb{E}v(m_i)) \cdot [F(\theta_i) - F(\theta_{i-1})]$$

When two intervals $[a, b]$ and $[b, c]$ are combined, the profit loss is therefore

$$\begin{aligned}\Delta = \quad & h(\mathbb{E}[v(\theta) \mid a \le \theta \le b]) \cdot [F(b) - F(a)] + h(\mathbb{E}[v(\theta) \mid b \le \theta \le c]) \cdot [F(c) - F(b)] \\ & - h(\mathbb{E}[v(\theta) \mid a \le \theta \le c]) \cdot [F(c) - F(a)].\end{aligned} \tag{7}$$

For notational convenience, let $v_1 = \mathbb{E}[v(\theta) \mid a \le \theta \le b]$, $v_2 = \mathbb{E}[v(\theta) \mid b \le \theta \le c]$ and $v = \mathbb{E}[v(\theta) \mid a \le \theta \le c]$. Observe that $v_1 < v < v_2$ and

$$v_1 \cdot [F(b) - F(a)] + v_2 \cdot [F(c) - F(b)] = \int_a^b v(\theta) f(\theta) d\theta + \int_b^c v(\theta) f(\theta) d\theta = v \cdot [F(c) - F(a)]. \tag{8}$$

Thus from Equation (7) and the strict convexity of h, it is clear that $\Delta > 0$.

To obtain a sharper estimate as required by the lemma, we apply second-order Taylor expansion to write

$$\begin{aligned} h(v_1) &= h(v) + (v_1 - v)h'(v) + \frac{(v_1 - v)^2}{2} h''(\xi) \\ h(v_2) &= h(v) + (v_2 - v)h'(v) + \frac{(v_2 - v)^2}{2} h''(\zeta) \end{aligned}$$

for some $\xi \in (v_1, v)$ and $\zeta \in (v, v_2)$. Plugging these into Equation (7) and using (8), we have

$$\begin{aligned} \Delta &= h(v_1) \cdot [F(b) - F(a)] + h(v_2) \cdot [F(c) - F(b)] - h(v) \cdot [F(c) - F(a)] \\ &= \frac{(v_1 - v)^2}{2} h''(\xi) \cdot [F(b) - F(a)] + \frac{(v_2 - v)^2}{2} h''(\zeta) \cdot [F(c) - F(b)]. \end{aligned}$$

Recall that h'' is bounded above and away from zero, and $F(b) - F(a)$ is on the same order as $b - a$ (since the density f is bounded above and away from zero). Thus the lemma would follow once we show that $v - v_1$ is on the same order as $c - b$ (and similarly $v_2 - v$ is on the same order as $b - a$).

Indeed, we can rewrite Equation (8) as $(v_2 - v_1) \cdot [F(c) - F(b)] = (v - v_1) \cdot [F(c) - F(a)]$. Thus it remains to show $v_2 - v_1$ is on the same order as $c - a$. Note that

$$v_2 - v(b) = \frac{\int_b^c [v(\theta) - v(b)] f(\theta) d\theta}{F(c) - F(b)} = \frac{\int_b^c \int_b^\theta v'(y) f(\theta) \, dy d\theta}{F(c) - F(b)}.$$

As $v'(y) f(\theta)$ is bounded above and away from zero, the numerator above is on the same order as $\int_b^c \int_b^\theta 1 \, dy d\theta = \frac{(c-b)^2}{2}$. So $v_2 - v(b)$ is on the same order as $c - b$. Similarly

$v(b) - v_1$ is on the same order as $b - a$. This proves that $v_2 - v_1$ is on the same order as $c - a$, and hence the lemma.

8.2 Proof of Lemma 1

We will show that any mechanism $\mathbb{M} = \langle M, p, q \rangle$ that satisfies $I(\mathbb{M}) = \kappa$, for some finite $\kappa > 0$, can be transformed into an interval mechanism in a way that does not change the expected profit of the monopolist, and weakly decreases the loss of privacy.

Given $\mathbb{M} = \langle M, p, q \rangle$ and a best-response strategy $\sigma(\cdot)$ for the agent under $\mathbb{M}$, we first drop duplicate messages: We say that message m' is a duplicate of message m if $p(m) = p(m')$ and $q(m) = q(m')$. Clearly, if m' is a duplicate of m, then removing m' from M and adjusting σ such that all types who sent m' would now send m, does not change the seller's expected profit. Moreover, the posterior belief given the message m in the new mechanism is an average of the posterior beliefs given the messages m and m' in the original mechanism. Due to the convexity of the divergence function $D_{KL}(F(\cdot|m) \;||\; F)$ in its first argument, the entailed loss of privacy $I(\mathbb{M})$ is decreased. *This is true under both the ex-post and ex-ante measures of privacy.*[29]

Next, denote by $\mu(m)$ the set of all types who report the message $m \in M$ with positive probability under σ:

$$\mu(m) = \{\theta \in \Theta \;\mid\; m \in supp(\sigma(\theta))\}$$

Since the agent's preference satisfies increasing differences in (θ, q), the set $\mu(m)$ is either an interval or a singleton.[30] However, since κ is finite, there can be only a zero-measure

---

[29] Given σ and F, denote by $\Pr(m \mid \sigma, F)$ and $\Pr(m' \mid \sigma, F)$ the probabilities that messages m and m' are reported under σ, respectively. Then the convexity of $D_{KL}(F(\cdot|m)||F)$ in its first argument implies that:

$$\begin{aligned}&\Pr(m \mid \sigma, F) \cdot D_{KL}(F(\cdot|m)||F) + \Pr(m' \mid \sigma, F) \cdot D_{KL}(F(\cdot|m')||F)\\ &\geq [\Pr(m \mid \sigma, F) + \Pr(m' \mid \sigma, F)] \cdot \left[D_{KL}\left(\frac{\Pr(m \mid \sigma, F) \cdot F(\cdot|m) + \Pr(m' \mid \sigma, F) \cdot F(\cdot|m')}{\Pr(m \mid \sigma, F) + \Pr(m' \mid \sigma, F)} \;||\; F\right)\right]\end{aligned}$$

where $\frac{\Pr(m \mid \sigma, F) \cdot F(\cdot|m) + \Pr(m' \mid \sigma, F) \cdot F(\cdot|m')}{\Pr(m \mid \sigma, F) + \Pr(m' \mid \sigma, F)}$ is the posterior belief that is induced when all the types who sent m' in equilibrium would now send m.

This inequality precisely says that ex-ante privacy loss is decreased. It further implies that the relative entropy induced by the new message m is no greater than the maximum of the relative entropies induced by the two old messages m and m'. Hence ex-post privacy loss is also decreased.

[30] Formally, if $\theta' \in \mu(m)$ and $\theta'' \in \mu(m)$ for some $m \in M$, then $\theta \in \mu(m)$ for all $\theta \in [\theta', \theta'']$. To see this, observe that $\theta' \in \mu(m)$ implies $q(m)\theta' - p(m) \geq q(m')\theta' - p(m')$ for every message m'. Similarly

subset of messages $m \in M$ for which $\mu(m)$ is a singleton.[31] We can therefore drop these messages from M, and pick a new best response for each type whose message was dropped. Since the behavior of only a zero-measure set of types was affected, the expected profit $\Pi(\mathbb{M})$ and the entailed loss of privacy $I(\mathbb{M})$ are both unchanged.

Henceforth we may assume that $\mu(m)$ is an interval for each m. Since there are no duplicates, for every pair of messages m and m' the intersection $\mu(m) \cap \mu(m')$ is either empty or a singleton (in other words, almost all types do not randomize between messages as part of their best-response).

To complete the transformation of $\mathbb{M}$ into an interval mechanism we now use a standard revelation argument: replace every message $m \in M$ with the corresponding interval $\mu(m)$, and adjust the function p (resp. q) such that whenever the agent reports the interval $\mu(m)$ in the "transformed" mechanism he would get the price (resp. quantity) that he would have got if he reported the message m in the "original" mechanism. The elements in the transformed message set are pairwise disjoint intervals whose union is Θ, and therefore they constitute a partition of Θ. IC, IR, privacy loss and profit are maintained under this transformation, which proves the lemma. ■

8.3 Proof of Proposition 2

We focus on the case where $c''' \geq 0$ and $v(\theta)$ is strictly less convex than $F(\theta)$. The proof strategy is to show that whenever an interval has more mass than its adjacent interval on the right, these two intervals can be "switched" to increase profit. That is, we considering changing the two intervals $[\theta_{t-1}, \theta_t]$ and $[\theta_t, \theta_{t+1}]$ into two new intervals $[\theta_{t-1}, \tilde{\theta}_t]$ and $[\tilde{\theta}_t, \theta_{t+1}]$, where $\tilde{\theta}_t$ is defined by $F(\tilde{\theta}_t) - F(\theta_{t-1}) = F(\theta_{t+1}) - F(\theta_t)$. By the assumption that the (original) left interval has greater mass, we have $\tilde{\theta}_t < \theta_t$.

Let $u, w, \tilde{u}, \tilde{w}$ denote the expected virtual valuation on the four intervals $[\theta_{t-1}, \theta_t]$, $[\theta_t, \theta_{t+1}]$, $[\theta_{t-1}, \tilde{\theta}_t]$, $[\tilde{\theta}_t, \theta_{t+1}]$ respectively. Then, as in the proof of Lemma 4, the profit

---

$q(m)\theta'' - p(m) \geq q(m')\theta'' - p(m')$. Since any $\theta \in (\theta', \theta'')$ is a convex combination of θ' and θ'', the above two inequalities lead to $q(m)\theta - p(m) \geq q(m')\theta - p(m')$. Thus m is a best-response of type θ. It is in fact a *strict* best-response because the last inequality is strict whenever $m' \neq m$; otherwise $q(m)\theta' - p(m) = q(m')\theta' - p(m')$ and $q(m)\theta'' - p(m) = q(m')\theta'' - p(m')$ hold simultaneously, showing that m' is a redundant copy of m. Hence for any θ strictly in between θ' and θ'', $\sigma(\theta)$ puts probability 1 on sending the message m.

[31] When $\mu(m)$ is a singleton, the message m is sent by exactly one type, and therefore m reveals this type in equilibrium.

increase due to switching the two intervals is given by

$$\Delta = (h(\tilde{w}) - h(u)) \cdot [F(\theta_t) - F(\theta_{t-1})] - (h(w) - h(\tilde{u})) \cdot [F(\theta_{t+1}) - F(\theta_t)] .$$

Observe that $(\tilde{w} - u) \cdot [F(\theta_t) - F(\theta_{t-1})] = (w - \tilde{u}) \cdot [F(\theta_{t+1}) - F(\theta_t)]$.[32] So to show Δ is positive we just need to show

$$\frac{h(\tilde{w}) - h(u)}{\tilde{w} - u} > \frac{h(w) - h(\tilde{u})}{w - \tilde{u}}.$$

We claim that the above inequality follows from $\tilde{w} + u > w + \tilde{u}$, which we will prove later. Indeed, as h is strictly convex, the RHS of this inequality increases in w. So it suffices to prove the weak version of the inequality assuming $\tilde{w} + u = w + \tilde{u}$. For that we rewrite it as $(w - \tilde{u}) \cdot \int_u^{\tilde{w}} h'(y)dy \geq (\tilde{w} - u) \cdot \int_{\tilde{u}}^{w} h'(y)dy$, which is in turn equivalent to

$$(w - \tilde{w} + u - \tilde{u}) \cdot \int_u^{\tilde{w}} h'(y)dy \geq (\tilde{w} - u) \cdot \left[\int_{\tilde{u}}^{u} h'(y)dy + \int_{\tilde{w}}^{w} h'(y)dy \right].$$

Since we assume $u - \tilde{u} = w - \tilde{w}$ (and $\tilde{w} > u$), this last inequality holds whenever h' is a concave function. Recall that $h'(y) = \phi(y)$ and $\phi'(y) = \frac{1}{c''(\phi(y))}$. So we have $h''(y) = \frac{1}{c''(\phi(y))}$, which is indeed decreasing in y because $c''' > 0$.

To complete the proof, it remains to verify that $\tilde{w} + u > w + \tilde{u}$. We introduce the function $G(x) = F^{-1}(x) \cdot (x - 1)$ for all $x \in [0, 1]$. Observe that $G(F(\theta)) = \theta(F(\theta) - 1)$, whose derivative is $v(\theta)f(\theta)$. Thus the expected virtual valuation $\tilde{w} = \mathbb{E}[v(\theta) \mid \tilde{\theta}_t \leq \theta \leq \theta_{t+1}]$ can be written as

$$\tilde{w} = \frac{\int_{\tilde{\theta}_t}^{\theta_{t+1}} v(\theta)f(\theta)d\theta}{F(\theta_{t+1}) - F(\tilde{\theta}_t)} = \frac{G(F(\theta_{t+1})) - G(F(\tilde{\theta}_t))}{F(\theta_{t+1}) - F(\tilde{\theta}_t)}.$$

For notational convenience, we let $a = F(\theta_{t-1}), b = F(\theta_t), c = F(\theta_{t+1})$ with $a < b < c$ and $b > \frac{a+c}{2}$. Then $F(\tilde{\theta}_t) = a + c - b < b$ and we have $\tilde{w} = \frac{G(c) - G(a+c-b)}{b-a}$. With similar computations for $u, w, \tilde{u}$, we only need to prove

$$\frac{G(c) - G(a + c - b) + G(b) - G(a)}{b - a} > \frac{G(c) - G(b) + G(a + c - b) - G(a)}{c - b}.$$

[32] Both are equal to $\int_{\theta_t}^{\theta_{t+1}} v(\theta)f(\theta)d\theta - \int_{\theta_{t-1}}^{\tilde{\theta}_t} v(\theta)f(\theta)d\theta$.

This is equivalent to

$$2(c-b)\cdot\int_{a+c-b}^{b}G'(x)dx>(a+c-2b)\cdot\left[\int_{a}^{a+c-b}G'(x)dx+\int_{b}^{c}G'(x)dx\right],$$

which holds so long as G' is a strictly concave function. This is indeed the case because $G'(F(\theta))=v(\theta)$, which is less convex than $F(\theta)$. The proposition follows. ■

8.4 Proof of Proposition 3

The following lemma gives a simple formula for the profit in the uniform-quadratic case.

Lemma 5 *In the uniform-quadratic case, the profit from any interval partition with cutoffs* $\underline{\theta}=\theta_0<\theta_1,\ldots,\theta_{n-1}<\theta_n=\overline{\theta}$ *is*

$$\Pi=\left(\frac{1}{6}\left(\overline{\theta}-\underline{\theta}\right)^2+\frac{1}{2}\underline{\theta}^2-\frac{1}{6\left(\overline{\theta}-\underline{\theta}\right)}\sum_{i=1}^{n}\left(\theta_i-\theta_{i-1}\right)^3\right),$$

which depends only on the lengths $\{\theta_i-\theta_{i-1}\}_{i=1}^{n}$.

Proof: When the agent's type is uniformly distributed over $[\underline{\theta},\overline{\theta}]$, the virtual value of type θ is given by $v(\theta)=2\theta-\overline{\theta}$, and the optimal quantity for any interval $[\theta_{i-1},\theta_i]$, as determined by Equation (4), is $\theta_i=\theta_i+\theta_{i-1}-\overline{\theta}$.

The profit as given by Equation (3) is:

$$\begin{aligned}\Pi(\omega)&=\sum_{i=1}^{n}\left(\theta_i+\theta_{i-1}-\overline{\theta}\right)\cdot\int_{\theta_{i-1}}^{\theta_i}\left(2x-\overline{\theta}\right)\frac{1}{\overline{\theta}-\underline{\theta}}dx-\frac{\left(\theta_i+\theta_{i-1}-\overline{\theta}\right)^2}{2}\frac{\theta_i-\theta_{i-1}}{\overline{\theta}-\underline{\theta}}\\&=\frac{1}{2\left(\overline{\theta}-\underline{\theta}\right)}\sum_{i=1}^{n}\left(\theta_i-\theta_{i-1}\right)\left(\theta_i+\theta_{i-1}-\overline{\theta}\right)^2\\&=\frac{1}{2\left(\overline{\theta}-\underline{\theta}\right)}\left(\sum_{i=1}^{n}\left(\theta_i-\theta_{i-1}\right)\left(\theta_i+\theta_{i-1}\right)^2-2\sum_{i=1}^{n}\left(\theta_i^2-\theta_{i-1}^2\right)\cdot\overline{\theta}+\sum_{i=1}^{n}\left(\theta_i-\theta_{i-1}\right)\cdot\overline{\theta}^2\right).\end{aligned}$$

The three terms in the parentheses above can be simplified as follows: $\sum_{i=1}^{n}\left(\theta_i-\theta_{i-1}\right)=\left(\overline{\theta}-\underline{\theta}\right)$, $\sum_{i=1}^{n}\left(\theta_i^2-\theta_{i-1}^2\right)=\left(\overline{\theta}^2-\underline{\theta}^2\right)$, and $\sum_{i=1}^{n}\left(\theta_i-\theta_{i-1}\right)\left(\theta_i+\theta_{i-1}\right)^2=\frac{4}{3}\left(\overline{\theta}^3-\underline{\theta}^3\right)-\frac{1}{3}\sum_{i=1}^{n}\left(\theta_i-\theta_{i-1}\right)^3$.[33] Plugging the three expressions back, we obtain the lemma. ■

[33] To simplify the third term we used the identity $(x-y)(x+y)^2=\frac{4}{3}\left(x^3-y^3\right)-\frac{1}{3}(x-y)^3$.

Back to the proposition, when $\kappa \in [\log(N), \log(N+1))$, we have $e^\kappa < N+1$. So by Lemma 2, any feasible interval mechanism has at most N intervals. Now by Lemma 5, among interval partitions with at most N intervals, maximizing profit is equivalent to minimizing

$$\sum_{i=1}^{N} x_i^3,$$

where $x_i = \theta_i - \theta_{i-1}$ is the length of the interval m_i (which can be zero if less than N intervals are used). Subject to $x_i \geq 0$ and the total length $\sum_{i=1}^N x_i = \overline{\theta} - \underline{\theta}$, the cubic sum $\sum_{i=1}^N x_i^3$ is clearly minimized when the intervals have equal lengths. Hence, the equal partition maximizes profit among all partitions with at most N intervals, even ignoring the feasibility constraint. Since it is κ-feasible, it must then be the ex-post κ-optimal mechanism. ■

8.5 Proof of Proposition 4

By the envelope theorem, the interim expected utility of a buyer with type $\hat{\theta}$ is given by $\int_{\theta \leq \hat{\theta}} q(\theta) d\theta$. Thus ex-ante buyer surplus can be computed as

$$\int \int_{\theta \leq \hat{\theta}} q(\theta) d\theta \; dF(\hat{\theta}) = \int q(\theta)(1 - F(\theta)) d\theta. \tag{9}$$

In what follows, we consider the effect of combining two adjacent intervals in a mechanism into a single interval. Specifically, let $\theta_{j-1}, \theta_j, \theta_{j+1}$ be three adjacent cutoffs in a constrained-optimal mechanism (for any κ). Write $v_j = \mathbb{E}[v(x) \mid x \in [\theta_{j-1}, \theta_j]]$, $v_{j+1} = \mathbb{E}[v(x) \mid x \in [\theta_j, \theta_{j+1}]]$, and $v = \mathbb{E}[v(x) \mid x \in [\theta_{j-1}, \theta_{j+1}]]$. Then the corresponding optimal quantities for these intervals are $q_j = \phi(v_j)$, $q_{j+1} = \phi(v_{j+1})$ and $q = \phi(v)$. Thus, the change in buyer surplus when "eliminating" the cutoff θ_j is

$$\begin{aligned}
\Delta :=& q \cdot \int_{\theta_{j-1}}^{\theta_{j+1}} (1 - F(\theta)) d\theta \quad - \quad q_j \cdot \int_{\theta_{j-1}}^{\theta_j} (1 - F(\theta)) d\theta \quad - \quad q_{j+1} \cdot \int_{\theta_j}^{\theta_{j+1}} (1 - F(\theta)) d\theta \\
=& (q - q_j) \cdot \int_{\theta_{j-1}}^{\theta_j} (1 - F(\theta)) d\theta \quad - \quad (q_{j+1} - q) \cdot \int_{\theta_j}^{\theta_{j+1}} (1 - F(\theta)) d\theta.
\end{aligned}$$

We will show $\Delta \geq 0$. Indeed, observe that

$$v(F(\theta_{j+1}) - F(\theta_{j-1})) = \int_{\theta_{j-1}}^{\theta_{j+1}} v(\theta) d\theta = v_j(F(\theta_j) - F(\theta_{j-1})) + v_{j+1}(F(\theta_{j+1}) - F(\theta_j)).$$

So $(v - v_j)(F(\theta_j) - F(\theta_{j-1})) = (v_{j+1} - v)(F(\theta_{j+1}) - F(\theta_j))$. Hence,

$$\frac{q - q_j}{q_{j+1} - q} = \frac{\phi(v) - \phi(v_j)}{\phi(v_{j+1}) - \phi(v)} \geq \frac{v - v_j}{v_{j+1} - v} = \frac{F(\theta_{j+1}) - F(\theta_j)}{F(\theta_j) - F(\theta_{j-1})} \geq \frac{\int_{\theta_j}^{\theta_{j+1}} (1 - F(\theta))d\theta}{\int_{\theta_{j-1}}^{\theta_j} (1 - F(\theta))d\theta}, \quad (10)$$

which precisely means that $\Delta \geq 0$. In the above, the first inequality holds because ϕ is concave (as its inverse function c' is assumed to be convex). The last inequality holds because it can be rewritten as

$$\frac{\int_{\theta_{j-1}}^{\theta_j} (1 - F(\theta))d\theta}{F(\theta_j) - F(\theta_{j-1})} \geq \frac{\int_{\theta_j}^{\theta_{j+1}} (1 - F(\theta))d\theta}{F(\theta_{j+1}) - F(\theta_j)},$$

where the LHS is $\frac{\int_{\theta_{j-1}}^{\theta_j} (1-F(\theta))d\theta}{\int_{\theta_{j-1}}^{\theta_j} f(\theta)d\theta} \geq \frac{1-F(\theta_j)}{f(\theta_j)}$ by the assumption that $\frac{1-F(\theta)}{f(\theta)}$ is decreasing, and similarly the RHS is at most $\frac{1-F(\theta_j)}{f(\theta_j)}$. This completes the proof that $\Delta \geq 0$.

Now, starting from any mechanism, repeatedly combining adjacent intervals eventually leads to the fully pooling mechanism, which yields weakly higher buyer surplus. Thus $\kappa = 0$ maximizes buyer surplus. Similarly $\kappa = \infty$ minimizes buyer surplus. ■

8.6 Proof of Proposition 5

With quadratic costs, total welfare contributed by a buyer of type θ is $\theta q(\theta) - (q(\theta))^2/2$. Thus ex-ante total welfare is

$$\int q(\theta) \cdot \left(\theta - \frac{q(\theta)}{2}\right) f(\theta)d\theta.$$

Note that on each interval $[\theta_{i-1}, \theta_i]$, $q(\theta)$ is constant and equal to the expected virtual valuation on this interval. Thus the above can be equivalently written as

$$\int q(\theta) \cdot \left(\theta - \frac{v(\theta)}{2}\right) f(\theta)d\theta. \quad (11)$$

Compared with the above Equation (9) for buyer surplus, the difference here is that the function $\left(\theta - \frac{v(\theta)}{2}\right) f(\theta)$ takes the place of $1 - F(\theta)$.

Note that $f(\theta)$ is increasing implies that $\theta - \frac{v(\theta)}{2}$ is decreasing, since its derivative is $-\frac{(1-F(\theta))f'(\theta)}{2f(\theta)^2}$. Thus we can repeat the above argument to show that combining two intervals increases total welfare, which must be maximized at $\kappa = 0$ and minimized at $\kappa = \infty$. This proves the first half of the proposition.

As for the second half, we can apply a symmetric argument: If $f(\theta)$ is decreasing, then $\theta - \frac{v(\theta)}{2}$ is increasing. This implies that combining two intervals would decrease total welfare, since the last inequality in Equation (10) would be reversed (and the first inequality would hold equal thanks to quadratic costs). Therefore total welfare would be minimized with a single interval, and maximized with full screening. ■

8.7 Proof of Proposition 10

Consider a symmetric auction that asks each agent which of n intervals his type belongs to. Suppose the partition has cutoffs $\underline{\theta} = \theta_0 < \theta_1 < \cdots < \theta_{n-1} < \theta_n = \overline{\theta}$. Then, for each buyer, the probability of winning upon reporting the interval $[\theta_{i-1}, \theta_i]$ is computed as

$$q_i = \sum_{k=0}^{m-1} \frac{1}{k+1} \binom{m-1}{k} \left(\frac{\theta_{i-1} - \theta_0}{\theta_n - \theta_0}\right)^{m-k-1} \left(\frac{\theta_i - \theta_{i-1}}{\theta_n - \theta_0}\right)^{k},$$

where each element in the sum corresponds to the event that $m-k-1$ opponents report an interval lower than $[\theta_{i-1}, \theta_i]$ and k opponents report the interval $[\theta_{i-1}, \theta_i]$, in which case the buyer wins the object with probability $\frac{1}{k+1}$. Simplifying the right-hand side yields[34]

$$q_i = \frac{1}{(\theta_n - \theta_0)^{m-1}} \cdot \frac{1}{m} \cdot \frac{(\theta_i - \theta_0)^m - (\theta_{i-1} - \theta_0)^m}{\theta_i - \theta_{i-1}}.$$

By the envelope theorem, type θ_i's interim expected utility is given by the integral of quantities assigned to lower types, which is

$$u_i = \sum_{j=1}^{i} (\theta_j - \theta_{j-1})\, q_j = \frac{1}{m} \frac{(\theta_i - \theta_0)^m}{(\theta_n - \theta_0)^{m-1}}.$$

---

[34] To see this note that $\frac{1}{k+1}\binom{m-1}{k} = \frac{1}{m}\binom{m}{k+1}$, and thus:

$$\sum_{k=0}^{m-1} \frac{1}{k+1} \binom{m-1}{k} (\theta_{i-1} - \theta_0)^{m-k-1} (\theta_i - \theta_{i-1})^k = \frac{1}{m} \sum_{k=0}^{m-1} \binom{m}{k+1} (\theta_{i-1} - \theta_0)^{m-(k+1)} (\theta_i - \theta_{i-1})^k$$
$$= \frac{1}{m} \frac{1}{\theta_i - \theta_{i-1}} \sum_{l=1}^{m} \binom{m}{l} (\theta_{i-1} - \theta_0)^{m-l} (\theta_i - \theta_{i-1})^l$$
$$= \frac{1}{m} \frac{1}{\theta_i - \theta_{i-1}} \left(\sum_{l=0}^{m} \binom{m}{k} (\theta_{i-1} - \theta_0)^{m-l} (\theta_i - \theta_{i-1})^l - (\theta_{i-1} - \theta_0)^m \right)$$
$$= \frac{1}{m} \frac{(\theta_i - \theta_0)^m - (\theta_{i-1} - \theta_0)^m}{\theta_i - \theta_{i-1}}.$$

The expected payment when reporting the interval $[\theta_{i-1}, \theta_i]$ is

$$p_i = \theta_i \cdot q_i - u_i = \frac{1}{(\theta_n - \theta_0)^{m-1}} \cdot \frac{1}{m} \cdot \frac{\theta_{i-1}(\theta_i - \theta_0)^m - \theta_i(\theta_{i-1} - \theta_0)^m}{\theta_i - \theta_{i-1}}$$

The total profit from all buyers is therefore:

$$\begin{aligned}
\Pi(\mathbb{M}) &= m \cdot \sum_{i=1}^{n} \left(\frac{\theta_i - \theta_{i-1}}{\theta_n - \theta_0} \right) \cdot p_i \\
&= \frac{1}{(\theta_n - \theta_0)^m} \cdot \sum_{i=1}^{n} \left(\theta_{i-1}(\theta_i - \theta_0)^m - \theta_i(\theta_{i-1} - \theta_0)^m \right) \\
&= \theta_0 + \frac{1}{(\theta_n - \theta_0)^m} \sum_{i=1}^{n} \left(\theta_i - \theta_0 \right) \left(\theta_{i-1} - \theta_0 \right) \left((\theta_i - \theta_0)^{m-1} - (\theta_{i-1} - \theta_0)^{m-1} \right).
\end{aligned}$$

We denote $z_i \equiv \frac{\theta_i - \theta_0}{\theta_n - \theta_0}$. Then the seller seeks to maximize the expression

$$\hat{\Pi} \equiv \sum_{i=1}^{n} z_i \cdot z_{i-1} \left(z_i^{m-1} - z_{i-1}^{m-1} \right), \tag{12}$$

subject to $0 = z_0 < z_1 < \cdots < z_n = 1$ and the ex-post privacy constraint that requires $z_i - z_{i-1} \geq e^{-\kappa}$ for all $1 \leq i \leq n$. In what follows we show that when $m \geq 5$, the solution to this problem is unique, with all but one interval having the same mass $z_2 - z_1 = z_3 - z_2 = \cdots = z_n - z_{n-1} = e^{-\kappa}$, and the lowest interval having greater mass $z_1 \in [e^{-\kappa}, 2e^{-\kappa})$. As a corollary, the optimal number of intervals n is also uniquely determined, as stated in the Proposition.

Toward this goal, we first argue that in *any* optimal solution, the intervals are ordered in *decreasing mass* from left to right. That is, $z_i - z_{i-1} \geq z_{i+1} - z_i$ for all $i \in \{1, ..., n-1\}$. To prove this, it suffices to consider the effect of "switching" two adjacent intervals on the profit $\hat{\Pi}$. Since this switch essentially replaces z_i with the number $z_{i-1} + z_{i+1} - z_i$, the result reduces to showing the following inequality: For any positive numbers $a < b < c < d$ with $d - c = b - a$, it holds that

$$ac(c^{m-1} - a^{m-1}) + cd(d^{m-1} - c^{m-1}) > ab(b^{m-1} - a^{m-1}) + bd(d^{m-1} - b^{m-1}).$$

Simplifying, we need to show that $(c - b)(d^m - a^m) > (d - a)(c^m - b^m)$, which can also be rewritten as

$$(c - b) \int_a^d x^{m-1} \, dx > (d - a) \int_b^c x^{m-1} \, dx.$$

That is, we need to show

$$(c-b)\left(\int_a^b x^{m-1}\,dx+\int_c^d x^{m-1}\,dx\right)>(b-a+d-c)\int_b^c x^{m-1}\,dx.$$

Since $b-a=d-c$, we can further rewrite it as

$$(c-b)\left(\int_a^b x^{m-1}+(a+d-x)^{m-1}\,dx\right)>(b-a)\left(\int_b^c x^{m-1}+(a+d-x)^{m-1}\,dx\right).$$

Since the function x^{m-1} is strictly convex when $m>2$, the integrand on the LHS is in fact uniformly larger than the integrand on the RHS, which proves the result.

Next we argue that $z_1<2(z_2-z_1)$, meaning that the longest interval is less than twice the length of the second longest. Indeed, if this were not the case, we could break the longest interval into two equal subintervals and still satisfy the ex-post privacy constraint. As can be easily seen from Equation (12), this modification strictly increases the profit, contradicting optimality.

Finally, we argue that in the optimal solution, the second-longest interval already exhausts the privacy constraint (and so must every "higher" interval). Indeed, if $z_2-z_1>e^{-\kappa}$, we could increase z_1 slightly without violating the privacy constraint. The effect on profit of this change is

$$\frac{\partial\hat{\Pi}}{\partial z_1}=z_2^{m-1}-mz_1^{m-1}$$

We have already shown that $z_1<2(z_2-z_1)$, so $z_2>1.5z_1$. Further note that for $m\geq 5$, $1.5^{m-1}>m$. Thus the above display implies that increasing z_1 would strictly increase profit, again leading to a contradiction.

Summarizing the above, we must have $z_2-z_1=z_3-z_2=\cdots=z_n-z_{n-1}=e^{-\kappa}$ and $z_1\in[e^{-\kappa},2e^{-\kappa})$. This proves the Proposition.

We mention that a similar (but more involved) analysis yields the optimal partitions for the cases $m=3$ and $m=4$ as well. As for $m=2$, the profit as given by Equation (12) can be simplified as

$$\sum_{i=1}^n z_i^2z_{i-1}-z_{i-1}^2z_i=\frac{1}{3}\sum_{i=1}^n\left(z_i^3-z_{i-1}^3-(z_i-z_{i-1})^3\right)=\frac{1}{3}-\frac{1}{3}\sum_{i=1}^n(z_i-z_{i-1})^3.$$

So maximizing profit is this auction problem is equivalent to minimizing the cubic sum of the interval lengths, as in the single-agent uniform-quadratic case. Hence the solution is an even partition with as many intervals as allowed by the privacy constraint. ■

9 Appendix 2 - Proofs for the ex-ante measure

9.1 Proof of Proposition 6

We expand on the proof sketch outlined in the main text. As discussed, the key is to find a replacement $\tilde{\mathbb{M}}$ for any mechanism $\mathbb{M}$ such that profit is not decreased, and the number of intervals in $\tilde{M}$ is bounded.

Step 1. Find a "big" interval. Set $l = e^{-\kappa}$. We first show that any ex-ante κ-feasible interval mechanism contains a "big" interval with mass $\geq l$ (according to F). Indeed, from Equation (6) we have

$$I(\mathbb{M}) = \sum_i -\left[F(\theta_i) - F(\theta_{i-1})\right] \cdot \log\left[F(\theta_i) - F(\theta_{i-1})\right] \leq \kappa.$$

Since $\sum_i \left[F(\theta_i) - F(\theta_{i-1})\right] = 1$, there exists some i s.t. $-\log\left[F(\theta_i) - F(\theta_{i-1})\right] \leq \kappa$. In other words, the interval m_i has mass at least $e^{-\kappa}$.

Fixing this choice of l, we define ϵ to be a small positive constant as given by Lemma 6 below. Starting from $\mathbb{M}$, we will now look for the replacement $\tilde{\mathbb{M}}$.

Step 2. From countable to finite. We first find a replacement $\hat{\mathbb{M}}$ with at least as much profit and only *finitely* many intervals. Suppose p is an accumulation point of the cutoffs in $\mathbb{M}$. Then on the left of p we can order the intervals in M from left to right as $m_1, m_2, \ldots$, with m_i converging to p. In particular, the mass of m_i converges to zero, and we can find some m_s and m_{s+1} both with mass $< \epsilon$. Applying Lemma 6 below, we can merge the intervals m_s and m_{s+1} and divide the "big" interval into two subintervals, in such a way that the (ex-ante) privacy measure is unchanged and profit is strictly increased. The achieved profit gain is sufficient to cover the loss from additionally combining all the (countably many) intervals $m_t, m_{t+1}, \ldots$, so long as we choose t to be sufficiently large. As this last step also relaxes the privacy constraint, we obtain a replacement mechanism in which p is no longer an accumulation point of intervals on its left. Doing the same exercise for intervals on the right of p yields a mechanism in which p is not an accumulation point.

In fact, we can achieve this replacement with some extra properties. Note that whenever an accumulation point p exists, the "big" interval must have mass strictly greater than $l = e^{-\kappa}$; otherwise the privacy constraint requires every interval in M to have mass exactly l, a contradiction. Thus by choosing m_s and m_{s+1} to have sufficiently small mass, we can ensure that when they are merged and the "big" interval is divided into two subintervals, *the bigger subinterval still has mass* $> l$. In other words, we can perform the

replacement in such a way that the *same big interval* is sequentially divided (each time creating a small subinterval on the left and a big one on the right). The benefit is that as we get rid of the accumulation points in $\mathbb{M}$ one by one (which may be countably many), we obtain a sequence of replacement mechanisms that become finer in the original "big" interval in $\mathbb{M}$ and more coarse everywhere else. This sequence converges, and the limit mechanism has at most one accumulation point in the "big" interval.[35] By merging and dividing once more, we arrive at $\hat{\mathbb{M}}$ with finitely many intervals and weakly higher profit than $\mathbb{M}$.

Step 3. From finite to bounded. We now demonstrate how to replace the finite mechanism $\hat{\mathbb{M}}$ with yet another mechanism $\tilde{\mathbb{M}}$ with higher profit and at most $N := \frac{2}{\epsilon} + 4$ intervals. Starting from $\hat{\mathbb{M}}$, if there are *two pairs* of adjacent intervals (i.e., 4 distinct ones) all with mass $< \epsilon$, then we combine both pairs at the same time and used the privacy measure saved from one of the mergers to divide the "big" interval into two subintervals. The privacy constraint is relaxed, and by Lemma 6 below, total profit is increased if we choose the merger that induces greater profit loss.

Hence whenever $\hat{\mathbb{M}}$ contains two pairs of adjacent "small" intervals, it can be replaced with a mechanism $\hat{\mathbb{M}}^{(1)}$ with higher profit and *one less interval in total*. The latter property ensures that when iterating this process, we will eventually reach a mechanism $\tilde{\mathbb{M}}$ in which at most one pair of adjacent intervals both have mass $< \epsilon$. Excluding this pair and the two intervals next to them, at least half of the remaining intervals have mass $\geq \epsilon$. So the total number of intervals in $\tilde{M}$ is bounded by $N = \frac{2}{\epsilon} + 4$. ■

Lemma 6 *Given $l > 0$, there exists $\epsilon \in (0, l)$ with the following property. If any interval mechanism $\mathbb{M}$ has two adjacent small intervals both of mass $< \epsilon$ as well as a big interval of mass $\geq l$, then when merging the two small intervals and using the saved ex-ante privacy measure to divide the big interval into two subintervals, the profit gain in the latter step is at least twice as big as the profit loss in the former step.*

Proof: Suppose there are two adjacent intervals with mass $x, y < \epsilon$; assume without loss that $x \leq y$. If we combine them into a single interval, the profit loss is on the order of $xy(x + y)$ by Lemma 4. Meanwhile, Equation (6) implies that the amount of privacy measure saved is

$$\alpha = (x + y)\log(x + y) - x\log x - y\log y = x\log(1 + \frac{y}{x}) + y\log(1 + \frac{x}{y}). \qquad (13)$$

[35] If we do not divide the same big interval repeatedly, then it is possible that new accumulation points arise in the iterative process. That would complicate the argument.

By assumption, there exists another interval of mass $L \geq l$. We use the saved privacy measure to break this interval into two: That is, we look for a subinterval of mass $\delta \in (0, \frac{L}{2})$ such that the total privacy measure is restored. This requires

$$L \log L - (L - \delta) \log(L - \delta) - \delta \log \delta = \alpha.$$

From this we obtain[36]

$$\delta \cdot |\log \delta| \geq \frac{\alpha}{2}. \tag{14}$$

We claim that (13) and (14) together imply $\delta \geq x\sqrt{x+y}$ (whenever $x \leq y < \epsilon$). For this it suffices to show that

$$x\sqrt{x+y} \cdot \log(\frac{1}{x\sqrt{x+y}}) < \frac{x \log(1 + \frac{y}{x})}{2} < \frac{\alpha}{2}.$$

Rearranging, the above inequality is equivalent to

$$\frac{1}{x\sqrt{x+y}} < \left(1 + \frac{y}{x}\right)^{\frac{1}{2\sqrt{x+y}}}.$$

For small x, y, the exponent $\frac{1}{2\sqrt{x+y}}$ is at least 4. So by binomial expansion, the RHS above has size at least

$$\binom{\frac{1}{2\sqrt{x+y}}}{4} \cdot \left(\frac{y}{x}\right)^4 \geq \left(\frac{1}{8\sqrt{x+y}}\right)^4 \cdot \frac{y}{x} = \frac{y}{4096x(x+y)^2} \geq \frac{1}{8192x(x+y)}.$$

This is indeed greater than the LHS, which was $\frac{1}{x\sqrt{x+y}}$.

Hence we have shown that when using the saved capacity to divide the big interval into two subintervals, the smaller subinterval has mass $\delta \geq x\sqrt{x+y}$. By Lemma 4, the resulting profit gain is on the order of $\delta(L - \delta)L \geq \frac{L^2\delta}{2}$. Since $L \geq l$ which is given, this profit gain is at least on the order of $\delta \geq x\sqrt{x+y}$. This greatly exceeds the initial profit loss (which is about $xy(x+y)$) due to combining two small intervals, completing the proof.

---

[36] By the Mean Value Theorem, $L \log L - (L - \delta) \log(L - \delta) = \delta(1 + \log \zeta)$ for some $\zeta \in (L - \delta, L)$. So $\delta(1 + \log \frac{\zeta}{\delta}) = \alpha$. Since $\zeta \geq \frac{L}{2} \geq \delta$, this implies

$$\delta \leq \alpha = x \log(1 + \frac{y}{x}) + y \log(1 + \frac{x}{y}) \leq x \cdot \frac{y}{x} + y \cdot \frac{x}{y} = x + y \leq \frac{1}{e}.$$

Thus we further have $1 + \log \zeta \leq 1 \leq -\log \delta$. From $\delta(1 + \log \frac{\zeta}{\delta}) = \alpha$ we then deduce $\delta \cdot |\log \delta| \geq \frac{\alpha}{2}$.

9.2 Proof of Proposition 8

We argue that Proposition 8 follows from Lemma 3, which we prove below. Indeed, that lemma implies the existence of some $\epsilon > 0$ such that any ex-ante κ-optimal interval mechanism with $\kappa \leq 1$ contains at most one interval with mass $< \epsilon$. For this ϵ, define $\underline{\kappa} = -\epsilon \log \epsilon$. Then in any κ-optimal mechanism with $\kappa \leq \underline{\kappa} < 1$, Equation (6) and feasibility implies

$$\sum_i - \left[F(\theta_i) - F(\theta_{i-1})\right] \cdot \log \left[F(\theta_i) - F(\theta_{i-1})\right] \leq \kappa \leq -\epsilon \log \epsilon.$$

In particular, $\left[F(\theta_i) - F(\theta_{i-1})\right] \cdot \log \left[F(\theta_i) - F(\theta_{i-1})\right] > \epsilon \log \epsilon$ holds for every interval. Note that the function $x \log x$ is decreasing for $x \in [0, \frac{1}{e}]$ and increasing for $x \in [\frac{1}{e}, 1]$. Thus the preceding inequality implies either $F(\theta_i) - F(\theta_{i-1}) < \epsilon$, or $F(\theta_i) - F(\theta_{i-1}) > \frac{1}{2}$.

In words, each interval in M has mass either less than ϵ or greater than $\frac{1}{2}$. By definition of ϵ, there is at most on interval with mass $< \epsilon$. It is also clear that at most one interval can have mass $> \frac{1}{2}$. Hence any ex-ante κ-optimal interval mechanism with $\kappa \leq \underline{\kappa}$ consists of at most two intervals. Since the ex-ante privacy constraint is exhausted, exactly two intervals are employed. ■

9.3 Proof of Lemma 3

In the proof of Proposition 6, we showed that in any ex-ante κ-feasible mechanism there is a "big" interval of mass at least $e^{-\kappa} \geq e^{-k}$. So by Lemma 6, there cannot be two adjacent intervals both with mass $< \epsilon$ (for some small ϵ).

It remains to deal with the situation where two small intervals are not adjacent. The proof strategy is to move one of these intervals to be next to the other, and to show that the profit change is at most on the order of xy, where x, y are the mass of these small intervals. Once this is shown, we can repeat the argument in the proof of Lemma 6, merging the now adjacent small intervals and dividing the big interval. As computed in that proof, the profit gain in the last step is on the order of $x\sqrt{x+y}$, which exceeds any profit loss incurred earlier. This would complete the proof.

To be more specific, suppose the two small intervals are $[\theta_{i-1}, \theta_i]$ and $[\theta_j, \theta_{j+1}]$, for some $i < j$. Set $x = F(\theta_i) - F(\theta_{i-1})$ and $y = F(\theta_{j+1}) - F(\theta_j)$. Consider *moving the small interval on the left toward the right while maintaining its mass*: We can do this sequentially by replacing θ_i with $\tilde{\theta}_i = F^{-1}(F(\theta_{i+1}) - x)$, then replacing θ_{i+1} with $\tilde{\theta}_{i+1} = F^{-1}(F(\theta_{i+2}) - x)$, so on and so forth until $\tilde{\theta}_{j-1} = F^{-1}(F(\theta_j) - x)$ and the two small intervals become adjacent. This process preserves the ex-ante privacy measure, and

it remains to estimate the profit change.

Note that in each step, the two intervals $[\tilde{\theta}_{t-1}, \theta_t]$ and $[\theta_t, \theta_{t+1}]$ are changed into two new intervals $[\tilde{\theta}_{t-1}, \tilde{\theta}_t]$ and $[\tilde{\theta}_t, \theta_{t+1}]$. Thus, as in the proof of Proposition 2, the profit increase is given by

$$\begin{aligned}\Delta_t = \quad & h(\tilde{u}) \cdot \left[F(\tilde{\theta}_t) - F(\tilde{\theta}_{t-1})\right] + h(\tilde{w}) \cdot \left[F(\theta_{t+1}) - F(\tilde{\theta}_t)\right] \\ & - h(u) \cdot \left[F(\theta_t) - F(\tilde{\theta}_{t-1})\right] - h(w) \cdot [F(\theta_{t+1}) - F(\theta_t)]\end{aligned} \tag{15}$$

where $u, w, \tilde{u}, \tilde{w}$ represent the expected virtual valuation on the intervals $[\tilde{\theta}_{t-1}, \theta_t]$, $[\theta_t, \theta_{t+1}]$, $[\tilde{\theta}_{t-1}, \tilde{\theta}_t]$, $[\tilde{\theta}_t, \theta_{t+1}]$ respectively.

We first consider the difference $h(\tilde{w}) \cdot \left[F(\theta_{t+1}) - F(\tilde{\theta}_t)\right] - h(u) \cdot \left[F(\theta_t) - F(\tilde{\theta}_{t-1})\right]$. By construction, $F(\theta_{t+1}) - F(\tilde{\theta}_t) = F(\theta_t) - F(\tilde{\theta}_{t-1}) = x$, so this difference simplifies to $(h(\tilde{w}) - h(u)) \cdot x$. Moreover, as we showed in the proof of Lemma 4,

$$u = \mathbb{E}[v(\theta) \mid \tilde{\theta}_{t-1} \le \theta \le \theta_t] = v(\theta_t) + O(\theta_t - \tilde{\theta}_{t-1}) = v(\theta_t) + O(F(\theta_t) - F(\tilde{\theta}_{t-1})) = v(\theta_t) + O(x)$$

where "$O(\cdot)$" is the standard big O notation with implied constants depending on the distribution and cost function. Thus $h(u) = h(v(\theta_t)) + O(x)$ and similarly $h(\tilde{w}) = h(v(\theta_{t+1})) + O(x)$. It follows that

$$h(\tilde{w}) \cdot \left[F(\theta_{t+1}) - F(\tilde{\theta}_t)\right] - h(u) \cdot \left[F(\theta_t) - F(\tilde{\theta}_{t-1})\right] = [h(v(\theta_{t+1})) - h(v(\theta_t))] \cdot x + O(x^2).$$

Next we consider the other difference $h(\tilde{u}) \cdot \left[F(\tilde{\theta}_t) - F(\tilde{\theta}_{t-1})\right] - h(w) \cdot [F(\theta_{t+1}) - F(\theta_t)]$ in Equation (15). It simplifies to $(h(\tilde{u}) - h(w)) \cdot [F(\theta_{t+1}) - F(\theta_t)]$. Moreover,

$$\begin{aligned}\tilde{u} &= \frac{\int_{\tilde{\theta}_{t-1}}^{\tilde{\theta}_t} v(\theta) f(\theta)\, d\theta}{F(\tilde{\theta}_t) - F(\tilde{\theta}_{t-1})} = \frac{\int_{\tilde{\theta}_{t-1}}^{\tilde{\theta}_t} v(\theta) f(\theta)\, d\theta}{F(\theta_{t+1}) - F(\theta_t)} = w + \frac{\int_{\tilde{\theta}_{t-1}}^{\theta_t} v(\theta) f(\theta)\, d\theta - \int_{\tilde{\theta}_t}^{\theta_{t+1}} v(\theta) f(\theta)\, d\theta}{F(\theta_{t+1}) - F(\theta_t)} \\ &= w + \frac{[v(\theta_t) - v(\theta_{t+1})] \cdot x}{F(\theta_{t+1}) - F(\theta_t)} + \frac{\int_{\tilde{\theta}_{t-1}}^{\theta_t} [(v(\theta) - v(\theta_t)] \cdot f(\theta)\, d\theta - \int_{\tilde{\theta}_t}^{\theta_{t+1}} [v(\theta) - v(\theta_{t+1})] \cdot f(\theta)\, d\theta}{F(\theta_{t+1}) - F(\theta_t)} \\ &= w + \frac{[v(\theta_t) - v(\theta_{t+1})] \cdot x}{F(\theta_{t+1}) - F(\theta_t)} + O(x^2),\end{aligned}$$

where the last step holds because for each $\theta \in [\tilde{\theta}_{t-1}, \theta_t]$, the difference between $[(v(\theta) - v(\theta_t)] \cdot f(\theta)$ and $[(v(\theta + \theta_{t+1} - \theta_t) - v(\theta_{t+1})] \cdot f(\theta + \theta_{t+1} - \theta_t)$ is at most on the order of $(\theta_t - \theta) \cdot (\theta_{t+1} - \theta_t) = O(x) \cdot [F(\theta_{t+1}) - F(\theta_t)]$. Thus $h(\tilde{u}) = h(w) + h'(w) \cdot \frac{[v(\theta_t) - v(\theta_{t+1})] \cdot x}{F(\theta_{t+1}) - F(\theta_t)} + O(x^2)$.

It follows that

$$h(\tilde{u}) \cdot \left[F(\tilde{\theta}_t) - F(\tilde{\theta}_{t-1})\right] - h(w) \cdot [F(\theta_{t+1}) - F(\theta_t)] = h'(w) \cdot [v(\theta_t) - v(\theta_{t+1})] \cdot x + O(x^2)$$

Taken together, we have estimated the RHS of Equation (15), so that

$$\Delta_t = \{h(v(\theta_{t+1})) - h(v(\theta_t)) - [v(\theta_{t+1}) - v(\theta_t)] \cdot h'(\mathbb{E}[v(\theta) \mid \theta \in [\theta_t, \theta_{t+1}]])\} \cdot x + O(x^2).$$

Summing across $t \in \{i, \ldots, j-1\}$, we obtain that when moving the small interval on the left to be adjacent to the one on the right, the total profit change is[37]

$$\Delta_{LR} = O(x^2) + \sum_{t=i}^{j-1} \{h(v(\theta_{t+1})) - h(v(\theta_t)) - [v(\theta_{t+1}) - v(\theta_t)] \cdot h'(\mathbb{E}[v(\theta) \mid \theta \in [\theta_t, \theta_{t+1}]])\} \cdot x.$$

If we instead move the small interval on the right to be adjacent to the one on the left, then total profit change is similarly computed as

$$\Delta_{RL} = O(y^2) - \sum_{t=i}^{j-1} \{h(v(\theta_{t+1})) - h(v(\theta_t)) - [v(\theta_{t+1}) - v(\theta_t)] \cdot h'(\mathbb{E}[v(\theta) \mid \theta \in [\theta_t, \theta_{t+1}]])\} \cdot y.$$

Note the minus sign in front of the second term; this is because when moving from the right to the left, the ordering of the subscripts need to be reversed.

Now observe that if we compute the weighted sum $y \cdot \Delta_{LR} + x \cdot \Delta_{RL}$, then the second term is cancelled out. This yields

$$y \cdot \Delta_{LR} + x \cdot \Delta_{RL} = O(x^2 y + y^2 x).$$

Therefore Δ_{LR} and Δ_{RL} cannot both be very negative. To be concrete we may without loss assume $\Delta_{LR} \geq -O(xy)$. Then in moving the small interval on the left to the right, the initial profit loss (if any) is small relative to the profit gain provided in Lemma 6. This again contradicts optimality, and hence there cannot even be two small intervals that are non-adjacent. ■

[37] There are $j - i$ terms of order at most x^2, and since $j - i$ is bounded by the total number of intervals which in turn is bounded by Lemma 6, their sum is still $O(x^2)$.

9.4 Proof of Proposition 9

We will show that in any profit-maximizing partition with exactly n intervals, all but one of the intervals have the same lengths and the last interval has weakly smaller length. To see how this claim implies the result, suppose $\kappa \in (\log(N-1), \log(N)]$. By Proposition 7, the number of intervals in the optimal solution satisfies $n \geq e^{\kappa} > N-1$. Thus $n \geq N$. Moreover, if $n > N$, then at least N intervals would have the same lengths. It would follow that *each* of the intervals in the optimal partition has *mass* smaller than $\frac{1}{N}$, which is at most $e^{-\kappa}$. This would contradict feasibility, i.e., Equation (6). Hence, with the preceding claim, we will know that the optimal partition involves exactly N intervals. The two lengths are then uniquely determined, as described in the Proposition. Finally, the ordering of the intervals does not matter for either the privacy measure or for profit, thanks to Lemma 5.

It remains to prove the above claim that characterizes the lengths of the optimal intervals. For an interval partition, let $x_i = \frac{\theta_i - \theta_{i-1}}{\bar{\theta} - \underline{\theta}}$ denote the probability mass of the i-th interval. In what follows we will work with the probability masses $\{x_i\}$ instead of the cutoffs $\{\theta_i\}$. By Lemma 5 and Equation (6), seller's profit maximization problem under the ex-ante privacy constraint can be rewritten as the following constrained minimization problem:

$$\min \sum_{i=1}^{n} x_i^3$$
$$s.t.\ x_i \geq 0, \quad \sum_{i=1}^{n} x_i = 1, \quad \sum_{i=1}^{n} x_i \log x_i \leq -\kappa.$$

For fixed n, the Lagrangian is given by:

$$\mathcal{L}\left(\alpha, \beta, \{x_i\}_{i=1}^{n}\right) = \sum_{i=1}^{n} x_i^3 + \alpha(1 - \sum_{i=1}^{n} x_i) - \beta(\sum_{i=1}^{n} x_i \log x_i + \kappa).$$

Thus, whenever each x_i is strictly positive and $\{x_i\}$ is a local constrained minimizer, the first order conditions imply

$$3x_i^2 - \beta \log x_i = \alpha + \beta \qquad \text{for all } 1 \leq i \leq n. \tag{16}$$

If $\beta \leq 0$, then the function $3x^2 - \beta \log x$ is monotonically increasing. Thus every x_i is the same, which completes the proof. Otherwise assume $\beta > 0$. In this case the derivative of the function $3x^2 - \beta \log x$ is $6x - \frac{\beta}{x}$, which is monotonically increasing and crosses 0

at $\hat{x} = \sqrt{\frac{\beta}{6}}$. Thus, the function $3x^2 - \beta \log x$ decreases on $[0, \hat{x}]$ and increases on $[\hat{x}, \infty)$. Equation (16) yields that x_i can take at most two values $\underline{x}$ and $\overline{x}$, with $\underline{x} < \hat{x} < \overline{x}$.

Below we analyze the second order conditions to show that at most one x_i can be equal to $\underline{x}$. Suppose for the sake of contradiction that in the optimal solution $\beta > 0$ and $x_1 = x_2 = \underline{x}$. Let $g(x) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i \log x_i)' \in \mathbb{R}^2$ denote the constraint values. Then its Jacobian $D_g(x)$ is the $2 \times n$ matrix whose first row is all 1s and whose second row is $(1 + \log x_1, \ldots, 1 + \log x_n)$. Consider $v = (1, -1, 0, \ldots, 0)' \in \mathbb{R}^n$. Then clearly v belongs to the null space of $D_g(x)$.

The second derivative of the Lagrangian $\mathcal{L}(\alpha, \beta, x)$ with respect to (the vector) x is the diagonal matrix $H = \text{diag}(6x_1 - \frac{\beta}{x_1}, \ldots, 6x_n - \frac{\beta}{x_n})$. It is easy to see that

$$v'Hv = 6x_1 - \frac{\beta}{x_1} + 6x_2 - \frac{\beta}{x_2} = 2\left(6\underline{x} - \frac{\beta}{\underline{x}}\right),$$

which is negative because $\underline{x} < \hat{x}$. But this fails the second derivative test for constrained local minimums; see, e.g., Simon and Blume (1994), page 468. Hence the proof is complete. ■

References

[1] **Acquisti, A. and Grossklags, J.** 2005. "Privacy and Rationality in Individual Decision Making." *IEEE Security and Privacy* 3.1: 26-33.

[2] **Acquisti, A., Taylor, C., and Wagman, L.**. 2016. "The Economics of Privacy." *Journal of Economic Literature*, 54(2): 442-492.

[3] **Agrawal, D. and Aggarwal, C.** 2001. "On the Design and Quantification of Privacy Preserving Data Mining Algorithms." In Proceedings of the 20th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '01), 247-255. ACM, New York.

[4] **Alonso, R. and Matouschek, N.** 2008. "Optimal Delegation." *Review of Economic Studies*, 75(1): 259-293.

[5] **Babaioff, M., Blumrosen, L., and Schapira, M.** 2013. "The Communication Burden of Payment Determination." *Games and Economic Behavior*, 77(1): 153-167.

[6] **Barth, S. and de Jong, M. D.** 2017. "The Privacy Paradox–Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior–A Systematic Literature Review." *Telematics and Informatics*. 34: 1038-1058.

[7] **Bergemann, D., Shen, J., Xu, Y., and Yeh, E.** 2012. "Multi-dimensional Mechanism Design with Limited Information", In Proceedings of the 13th ACM Conference on Electronic Commerce (EC'12), 162-178. ACM, New York.

[8] **Blumrosen, L. and Feldman, M.** 2013. "Mechanism Design with a Restricted Action Space." *Games and Economic.Behavior* 82: 424-443.

[9] **Blumrosen, L., Nisan, N., and Segal, I.** 2007. "Auctions with Severely Bounded Communication." *Journal of Artificial Intelligence Research.* 28: 233-266.

[10] **Calzolari, G. and Pavan, A.** 2006. "On the Optimality of Privacy in Sequential Contracting." *Journal of Economic theory*, 130: 168-204.

[11] **Choi, J.P., Jeon, D., and Kim, B.** 2019. "Privacy and Personal Data Collection With Information Externalities." *Journal of Public Economics*, 173: 113-124.

[12] **Conitzer, V., Taylor, C. R., and Wagman, L.** 2012. "Hide and Seek: Costly Consumer Privacy in a Market with Repeat Purchases." *Marketing Science*, 31: 277-292.

[13] **Cover, M. T. and Thomas, J. A.**. 2006. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, New York.

[14] **Crémer, J., Garicano, L., and Prat, A.** 2007. "Language and the Theory of the Firm." *Quarterly Journal of Economics*, 122(1): 373-407.

[15] **Díaz, C., Seys, S., Claessens, J., and Preneel, B.** 2002. "Towards Measuring Anonymity." In Proceedings of the 2nd International Conference on Privacy Enhancing Technologies (PET'02), Dingledine R. and Syverson P. (eds.), 54-68. Springer-Verlag, Berlin, Heidelberg.

[16] **Dwork, C., McSherry, F., Nissim, K., and Smith, A.** 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." In Theory of Cryptography. TCC 2006, Halevi S., Rabin T. (eds.), Lecture Notes in Computer Science, vol 3876. Springer, Berlin, Heidelberg.

[17] **Fadel, R. and Segal, I.** 2009. "The Communication Cost of Selfishness." *Journal of Economic Theory*, 144(5): 1895-1920.

[18] **Fairfield, J.A.T. and Engel, C.** 2015. "Privacy as a Public Good." *Duke Law Journal*, 65(3): 385-457.

[19] **Fleischer, L. and Lyu, Y.** 2012. "Approximately Optimal Auctions for Selling Privacy when Costs are Correlated with Data." In Proceedings of the 13th ACM Conference on Electronic Commerce (EC '12), 568-585. ACM, New York.

[20] **Frankel, A.** 2014. "Aligned Delegation." *American Economic Review*, 104: 66-83.

[21] **Ghosh, A. and Roth, A.** 2011. "Selling Privacy at Auction." In Proceedings of the 12th ACM Conference on Electronic Commerce (EC '11), 199-208. ACM, New York.

[22] **Gilboa-Freedman, G. and Smorodinsky, R.** 2018. "On the Properties that Characterize Privacy." Working Paper.

[23] **Gradwohl, R.** 2018. "Privacy in Implementation", *Social Choice and Welfare*, 50(3): 547-580.

[24] **Green, J. R. and Laffont, J. J.** 1986. "Incentive Theory with Data Compression." In Uncertainty, Information and Communication: Essays in Honor of Kenneth Arrow, Heller, W. P., Starr, R.M. and Starrett D.A. (eds.), 239-253. Cambridge: Cambridge Univ. Press.

[25] **Green, J. R. and Laffont, J. J.** 1987. "Limited Communication and Incentive Compatibility." In *Information, Incentives, and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*, Groves, T., Radner, R., and Reiter, S. (eds.), 308-329. Minneapolis: Univ. Minnesota Press.

[26] **Heffetz, O. and Ligett, K.** 2014. "Privacy and Data-Based Research." *Journal of Economic Perspectives*, 28(2): 75-98.

[27] **Hidir, S. and Vellodi, N.** 2018. "Personalization, Discrimination and Information Revelation." Working paper.

[28] **Ichihashi, S.** 2019. "Online Privacy and Information Disclosure by Consumers." *American Economic Review*, forthcoming.

[29] **Kearns, M., Pai, M., Roth, A., and Ullman, J.** 2014. "Mechanism Design in Large Games: Incentives and Privacy." *American Economic Review*, 104(5): 431-435.

[30] **Kokolakis, S.** 2017. "Privacy Attitudes and Privacy Behavior: A Review of Current Research on the Privacy Paradox Phenomenon." *Computers and Security*, 122-134.

[31] **Kos, N.**. 2012. "Communication and Efficiency in Auctions." *Games and Economic Behavior*, 75: 233-249.

[32] **Lan, S.** 2019. "If You Design the Encryption Correctly, There is Not Enough Time in the Universe to Crack it", *Globes*, 6 September 2019. Accessed 13 September 2019. www.globes.co.il/news/article.aspx?did=1001299378. In Hebrew.

[33] **Ligett K. and Roth. A.** 2012. "Take It or Leave It: Running a Survey when Privacy Comes at a Cost." In Proceedings of the 8th International Conference on Internet and Network Economics (WINE '12), Goldberg P.W. (eds.), 378-391. Springer-Verlag, Berlin, Heidelberg.

[34] **Maćkowiak, B. and Wiederholt, M.** 2015. "Business Cycle Dynamics under Rational Inattention." *The Review of Economic Studies*, 82(4): 1502-1532.

[35] **Matějka, F.** 2016. "Rationally Inattentive Seller: Sales and Discrete Pricing." *The Review of Economic Studies*, 83(3): 1125-1155.

[36] **Matějka, F. and McKay, A.** 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *The American Economic Review*, 105(1): 272-298.

[37] **McSherry, F. and Talwar, K.** 2007. "Mechanism Design via Differential Privacy." In Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS '07), 94-103. IEEE Computer Society, Washington DC.

[38] **Melumad, N., Mookherjee, D., and Reichelstein, S.** 1992. "A Theory of Responsibility Centers." *Journal of Accounting and Economics*, 15(4): 445-484.

[39] **Mookherjee, D. and Tsumagari, M.** 2014. "Mechanism Design with Communication Constraints." *Journal of Political Economy*, 122(5): 1094-1129.

[40] **Mussa, M. and Rosen, S.** 1978. "Monopoly and Product Quality." *Journal of Economic Theory*, 18(2): 301-317.

[41] **Nissim, K., Smorodinsky, R., and Tennenholtz, M.** 2012. "Approximately Optimal Mechanism Design via Differential Privacy." In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS '12), 203-213. ACM, New York.

[42] **Ollár, M., Rostek, M., and Yoon, J.H.** 2016. "Privacy-Preserving Market Design." Working Paper.

[43] **Pai, M. and Roth, A.** 2013. "Mechanism Design and Privacy." *SIGecom Exchanges*, 12(1): 8-29.

[44] **Pomatto, L., Strack, P., and Tamuz, O.** 2019. "The Cost of Information." Working Paper.

[45] **Rebollo-Monedero, D., Forne, J., and Domingo-Ferrer, J.** 2009. "From t-closeness-like Privacy to Postrandomization via Information Theory." *IEEE Transactions on Knowledge and Data Engineering*, 99(1).

[46] **Sankar, L., Rajagopalan, S. R., and Poor, H. V.** 2013. "Utility-privacy Tradeoffs in Databases: An Information-theoretic Approach." *IEEE Transactions on Information Forensics and Security*, 8(6): 838-852.

[47] **Simon, C. P. and Blume, L. E.** 1994. *Mathematics for Economists.* Norton New York.

[48] **Sims, C. A.** 2003. "Implications of Rational Inattention." *Journal of Monetary Economics*, 50(3): 665-690.

[49] **Taylor, C. R.** 2004. "Consumer Privacy and the Market for Customer Information." *Rand Journal of Economics*, 631-650.

[50] **Van Zandt, T.** 2007. "Communication Complexity and Mechanism Design." *Journal of European Economic Association*, 5: 543-553.

[51] **Wang, W., Ying, L., and Zhang, J.** 2016. "On the Relation between Identifiability, Differential Privacy, and Mutual-information Privacy." *IEEE Transactions on Information Theory*, 62(9): 5018-5029.