Optimal Privacy-Constrained Mechanisms

Ran Eilat*, Kfir Eliaz†and Xiaosheng Mu‡

February 14, 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-ante (the expected difference in beliefs over all type realizations cannot exceed some threshold κ) and ex-post (for every realized type, the maximal difference in beliefs 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.

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

†School of Economics, Tel-Aviv University and Economics Dept., Columbia University. kfire@post.tau.ac.il.

‡Cowles Foundation and Columbia University. xm2230@columbia.edu

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. This data is 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 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 that the designer of a mechanism already possesses some information about participants in his mechanism in the form of a prior belief over their "types". He updates these beliefs as a result of the participants' interaction with the mechanism, which releases some information about them. For example, when a consumer who faces a menu of choices, say quantity-price pairs, selects a particular item in the menu, the seller learns additional information about this buyer. In particular, the seller knows the consumer is willing to pay the price he chose, and that no other quantity-price pair is preferred. Consequently, the seller's posterior belief about the buyer's type may be quite different than his prior belief. This suggests that the difference between the seller'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 privacy loss for mechanisms and apply it to screening mechanisms in which there is no strategic interaction between the participants. Specifically, we consider the classic Mussa-Rosen set-up in which a monopolist faces increasing costs for producing a higher quality (or quantity) of a product, and wishes to offer the optimal menu of quality-

price pairs to consumers with private marginal rates of substitution between quality and money. The standard solution implies that all the types that opt in effectively 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-theoretic literature and propose to measure a mechanism's inherent loss of privacy as the *expected relative entropy* (or Kullback-Leibler Divergence) between the designer's posterior and prior beliefs, where the expectation is taken with respect to the prior distribution over consumer types. We then augment the standard mechanism design problem by requiring that the privacy loss of the optimal solution is at most κ. The parameter κ, which takes values between 0 (full privacy) and infinity (no privacy), captures the strength of the privacy requirement.

We view this ex-ante notion of privacy as a conservative departure from the standard privacy-unconstrained approach in mechanism design in the following sense: It acknowledges 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"), and hence, allows the designer to preserve privacy in a differential manner across consumer types (so some types may release more information than other types) as long as on average, a given level of privacy is maintained. The privacy constraint may also be interpreted as a budget for "securing" sensitive data, such that more precise and detailed data is more costly to secure. The ex-ante constraint takes into account that the designer may find it profitable to allocate these costs in a differential manner across consumer types. This interpretation of the privacy constraint, and our ex-ante approach, also create an interesting link between privacy constraints and rational inattention, and we discuss this relation in the next section. Finally, the ex-ante approach also has the merit of making the analysis relatively tractable. Hence, adopting a well-studied measure of difference between distributions, and taking an ex-ante approach serves as a useful benchmark with which we can compare other measures of privacy.

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 (2017) and Kokolakis (2017)). Alternatively, our approach can be interpreted as assuming homogenous preferences over privacy that take a threshold form: A consumer transacts with a platform that has κ or less loss of privacy.

Our main results highlight key properties of the optimal privacy-preserving mechanisms. First, the optimal κ-constrained mechanism partitions the set of types into *finitely* many intervals (whose number depends on κ), such that consumers truthfully announce to which interval their type belongs, and the total loss of privacy is exactly κ. Thus, even though there is a continuum of types, and the privacy constraint allows for a continuum of noisy messages (e.g., when each type θ reports $\theta + \varepsilon$, where ε is a continuous random variable), maximal profits are attained with only finitely many messages. The second property relates to the structure of the intervals: There can be at most one interval with an arbitrarily small mass. In other words, there is at most one set of types with positive measure about which the monopolist attains very precise information. This property also implies that there exists a threshold $\underline{\kappa}$ such that for any $\kappa \leq \underline{\kappa}$, the optimal κ-constrained mechanism has exactly *two* intervals. 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, the optimal κ-constrained mechanism is unique up to reordering of the intervals and has the following properties. The number of intervals is equal to the smallest integer n^* whose natural logarithm is at least κ. There is exactly one "short" interval and $n^* - 1$ "long" intervals of equal length, such that privacy loss is precisely κ. In addition, the optimal mechanism exhibits an interesting trade-off between privacy and profits: As κ increases, there are *diminishing* returns to loss of privacy when the optimal number of intervals increase, but there are *increasing* returns over ranges of κ where the optimal number of intervals remains fixed (but their length changes). These properties of the optimal mechanism remain true for distributions that are close to the uniform.

Our ex-ante notion of privacy allows the designer to meet the privacy constraint even if he can learn almost perfectly about some small set of types. A more stringent notion of privacy would restrict the designer not to learn too much about *any* consumer type. To explore the implications of such a notion of privacy, we require that

the *largest* change in the designer's beliefs (as measured by relative entropy) must be at most κ. Under this ex-post privacy constraint, there exists an optimal mechanism that partitions the type space into finitely many intervals, each with a mass of at least $e^{-\kappa}$. In particular, in the uniform-quadratic case, for any $\kappa \in [\log(n), \log(n+1))$, the optimal κ-constrained mechanism partitions the types into n equal intervals.

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 measure 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, thereby allowing us to better understand the trade-offs between welfare/profits and privacy demands. 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).

There are many interesting questions left to explore in the study of privacy-preserving mechanisms. In particular, studying mechanisms with strategic interaction between participants raises some novel challenges. First, one-shot mechanisms may not be optimal in these environments. Second, one needs to take a stand on how privacy loss is aggregated across different individuals. This is particularly important since optimal mechanisms may exhibit a differential treatment of ex-ante identical agents. We discuss these issues in Section 6.

2 Related literature

On the one hand, our notion of privacy differs from the popular measure of "differential privacy" that is often used in the computer science literature. On the other hand, it coincides with how the rational inattention literature models the cost of information. Hence, our proposed framework creates an interesting link between these two distinct strands of literature. In this subsection we briefly summarize the main insights of these literatures and their relation to our research.[1]

The majority of theoretical work on privacy in computer science uses the notion of "differential privacy", which was introduced by Dwork et al. (2006). Roughly speaking, this notion means that changing the data of only a single individual, or alternatively, of only a single attribute of an individual, has a negligible effect on

[1]There are many works in these literatures, 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).

computations that are done on this data. In the context of mechanism design, Pai and Roth (2013) show that this notion can be defined 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 $u : X \to \mathbb{R}$,

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

This definition implies that the action 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 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 truthtelling is either "almost" or exactly weakly dominant (see e.g., McSherry and Talwar (2007), Kearns et al. (2012) and Nissim et al. (2012)).

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). Closely related, Gradwohl (2018) studies the problem of full implementation when agents prefer to protect their privacy.

Yet another literature in computer science deals with distortion and anonymization of databases and communication channels due to privacy concerns. Within this literature, several papers used information-theoretic measures to quantify privacy, like the notion of relative entropy that we use in our work. Noteworthy examples are Agrawal and Aggarwal (2001) who study privacy-preserving data-mining algorithms; Rebollo-Monedero et al. (2010) and Sankar et al. (2013) who study the privacy-distortion trade-off; Wang et al. (2016) who link between three different notions of privacy in the privacy-distortion context; and Díaz et al. (2003) who study the degree of anonymity provided by schemes for anonymous connections. The key distinction of the current paper is that we are interested in the strategic interaction between privacy, mechanism and agent behavior, while in this literature strategic behavior does not play any role.

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 different restrictions on the cardinality of the action space available to the agents. Notable examples are Kos (2012), Blumrosen et al. (2007), Blumrosen and Feldman (2013), Bergemann et al. (2012), Melumad et al. (1992) and Green and Laffont (1987). 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 and bounding 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.[2]

Finally, our work is closely related to the growing literature on rational inattention with information costs (see, e.g., Sims (2003), Matějka (2016), Matějka and McKay (2015) and Maćkowiak and Wiederholt (2015)). In this literature, an uninformed decision maker (DM) chooses the structure of a signal he wants to observe, subject to the constraint that the signal can only contain a limited amount of information.[3] In fact, the choice of the DM is tantamount to choosing a distribution of posterior beliefs, subject to the information capacity constraint (and the martingale condition of beliefs). Note that when a privacy-constrained designer chooses a mechanism (with a corresponding equilibrium), he also implicitly chooses a distribution of posterior beliefs, subject to the same information constraint.[4] However, while the rationally inattentive DM is bound only by the information constraint, the mechanism designer is bound also by an incentive constraint – the participating agent(s) must be willing to share the information in equilibrium. Studying the interaction between the information constraint and the incentive constraint is the main objective of our work.

[2]Green and Laffont (1986) study a model in which a principal can restrict the capacity of a communication channel between an agent and his obedient subordinate. Like in our model, the capacity of the channel is quantified using an information-theoretic measure (mutual information). Unlike our work, there is no conflict of interests between the agent and the subordinate. Therefore, when the (informed) agent designs the optimal communication protocol there are no incentive constraints involved.

[3]The amount of information is measured as the expected reduction in entropy between the prior and posterior beliefs (that the signal induces) regarding the state of the world.

[4]Formally, the designer chooses a set of messages for the agent(s) and a function that maps between the (profile of) messages and consequences. However, in equilibrium messages can be identified with the posterior beliefs they induce regarding the agent type.

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 $q > 0$. 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 $v(\underline{\theta}) > 0$.[5] To facilitate some technical arguments, we make the slightly stronger assumption that v is continuously differentiable and $v' > 0$.

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:

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

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 $p(\theta) = q(\theta)\theta - \int_{\underline{\theta}}^{\theta} q(x)\, dx$.

---

[5] 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. The strict inequality $v(\underline{\theta}) > 0$ is used in the proof that an optimal mechanism exists. But we note that a slightly modified argument applies if $v(\underline{\theta}) = 0$ and additionally $c''(0) > 0$.

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 for the buyer.

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:[6]

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

If $F(\cdot|m)$ contains atoms we define $D_{KL}(F(\cdot|m)\,||F) = +\infty$.[7] Throughout the paper, "log" represents the natural logarithm.

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

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

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

where $\mathbb{E}_m$ is evaluated according to the probability that each message $m \in M$ is sent in an equilibrium of $\mathbb{M}$.[8,9]

---

[6] 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.

[7] The integral on the RHS of (1) can be evaluated whenever $F(\cdot|m)$ is absolutely continuous with respect to F. Since we have assumed that F admits a density w.r.t. Lebesgue measure, absolute continuity is guaranteed when $F(\cdot|m)$ also admits a density. And when $G := F(\cdot|m)$ contains atoms, our definition that $D_{KL}(G||F) = +\infty$ preserves continuity of the relative entropy function in G.

[8] In calculating $I(\mathbb{M})$ we adopt the convention that $0 \cdot \infty = 0$, and therefore $I(\mathbb{M})$ 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 $I(\mathbb{M}) = +\infty$ according to our definition.

[9] 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/buyer indifference only arise when there are messages that lead to the same quantity-price

4 Optimal privacy-constrained mechanisms

4.1 Interval 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:

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

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

   for all $\theta \in \Theta$, all $m \in supp(\sigma(\theta))$ 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 $\mathbb{M}$:

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

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

3. *Privacy constraint* -

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

We refer to any mechanism that satisfies the above constraints as a *κ-feasible mechanism*. 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? Do some buyer types disclose more information than others? What is the maximal amount of information that is revealed by any buyer type? Is the privacy constraint even binding?

Note that in standard mechanism design the monopolist maximizes his expected profit subject only to the incentive-compatibility and individual-rationality constraints. 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

---

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.

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.

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. Mechanisms that rely on mixed strategies are "wasteful" in the sense that the seller could relax the privacy constraint by inducing pure strategies without affecting the outcome. This means that we can without loss assume the supports of the seller's posterior beliefs constitute a partition of $[\underline{\theta}, \overline{\theta}]$. The single-crossing property of the buyer's preferences further implies that the sets of types that "pool" together are convex. Hence, the aforementioned partition consists of intervals, leading to the lemma.

Next, we use the interval characterization to derive the quantity and price that a κ-optimal mechanism assigns to each message. Given a feasible mechanism $\mathbb{M} = \langle M, p, q \rangle$, in which all the messages $m \in M$ are intervals and each type $\theta \in \Theta$ reports the interval to which it belongs, the expected profit for the seller from employing $\mathbb{M}$ is given by:[10]

$$\Pi(\mathbb{M}) = \sum_{m \in M} \left[q(m) \int_{\underline{m}}^{\overline{m}} v(\theta) f(\theta) \, d\theta - c(q(m)) \cdot [F(\overline{m}) - F(\underline{m})] \right] \tag{2}$$

where $\underline{m}$ and $\overline{m}$ are the lower and upper bounds, respectively, for any interval $m \in M$. Therefore, the quantity that maximizes the expected profit while maintaining IC and

[10] 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 (2) is obtained from this formula using integration by parts.

IR is uniquely determined by:[11]

$$c'\left(q\left(m\right)\right)=\mathbb{E}_{F}\left[v\left(x\right)\ \mid x\in[\underline{m},\overline{m}]\right]\qquad\text{for any } m\in M \tag{3}$$

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

$$p\left(m\right)=q\left(m\right)\cdot\underline{m}-\sum_{m'\in M\text{ s.t }\overline{m}'\leq\underline{m}}\left(\overline{m}'-\underline{m}'\right)\cdot q\left(m'\right)\qquad\text{for any } m\in M \tag{4}$$

where the summation is over all the intervals m' that are "lower" than m. It follows that the assignment of types to quantity-price pairs in any κ-optimal mechanism is completely determined by the interval partition. Therefore, in the rest of the analysis we focus on characterizing the set of intervals M in the κ-optimal mechanism.

To do this, we first rewrite the seller's optimization problem in terms of the interval partition. In particular, we will compute the privacy measure of any mechanism that uses intervals as messages. Note that when the seller sees a message m in equilibrium, his posterior density updates to $f(\theta\mid m)=\frac{f(\theta)}{F(\overline{m})-F(\underline{m})}$ for $\theta\in[\underline{m},\overline{m}]$, and $f(\theta\mid m)=0$ otherwise. The relative entropy between this posterior belief and the prior is computed as $\int_{\underline{m}}^{\overline{m}}f(\theta\mid m)\log\frac{f(\theta|m)}{f(\theta)}d\theta=-\log\left[F(\overline{m})-F(\underline{m})\right]$. Since the message m is sent in equilibrium with probability $F(\overline{m})-F(\underline{m})$, we deduce that for interval mechanisms $\mathbb{M}$, the ex-ante loss of privacy is given by:

$$I(\mathbb{M})=\sum_{m\in M}-\left[F(\overline{m})-F(\underline{m})\right]\cdot\log\left[F(\overline{m})-F(\underline{m})\right]. \tag{5}$$

Consider the discrete distribution g_M over the elements of M induced by the prior. That is, $g_M(m)=F(\overline{m})-F(\underline{m})$ is the ex-ante probability that the buyer's type θ belongs to the interval m. Then the above equation (5) can be compactly written as $I(\mathbb{M})=H(g_M)$, which is the Shannon entropy of the discrete distribution g_M.

This discussion yields the following result:

Lemma 2 *The profit maximization problem is equivalent to finding a set of intervals M that partition $[\underline{\theta},\overline{\theta}]$ and satisfy $H(g_M)\leq\kappa$, such that (2) is maximized subject to these constraints and with quantities given by (3).*

---

[11] Since c is strictly convex and $c'(0)=0$, the first order condition (3) uniquely determines the value of the optimal $q\left(m\right)$. The fact that $v\left(\cdot\right)$ is increasing ensures that $q\left(\cdot\right)$ is "increasing in m." Thus higher types receive higher quantity in equilibrium and local IC implies global IC.

4.2 Existence and further properties

So far we have set aside an important technical issue of whether a κ-optimal mechanism exists. To see why existence is not straightforward in our setting, recall Lemma 2 from above. Although that lemma provides a simple constrained optimization program in terms of the intervals, the space over which the seller optimizes is *not compact*. Indeed, compactness is guaranteed with any finite upper bound on the number of intervals used in the mechanism, but a priori the seller could even partition buyer types into *countably* many intervals.

We will however show that an optimal mechanism exists and consists of finitely many intervals.

Proposition 1 *There exists a 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.*[12]

The proof goes as follows: Consider a sequence of 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 κ-feasible mechanisms. We will replace each mechanism $\mathbb{M}_j$ by another κ-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 $\tilde{M}_\infty$. By Lemma 2 and 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 $\Pi(\tilde{M}_\infty) \geq \Pi^*$ and $\tilde{M}_\infty$ would be κ-optimal.

It remains to find the appropriate replacements $\tilde{\mathbb{M}}_j$. We first observe that starting from any mechanism $\mathbb{M}_j$, merging two adjacent intervals in M_j into a single interval (and adjusting the quantities/prices accordingly) always strictly decreases the profit. However, by doing so the seller is able to save on the 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

[12] 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 rather elementary and only makes use of the tradeoff between privacy and profit when merging/dividing intervals.

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.[13]

Below we collect a few other results that emerge from this proof:

Proposition 2 *Under the Bayesian privacy measure, the privacy constraint is exhausted in any κ-optimal mechanism $\mathbb{M}$. That is, $I(\mathbb{M}) = \kappa$.*

As discussed above, the intuition is that the seller always benefits from refining the information he elicits about the buyer's type (i.e., dividing an interval into two subintervals). By choosing one of the subintervals to be "small," the *average* privacy constraint is still satisfied. Note however that this argument and conclusion does not extend to an alternative ex-post notion of privacy loss, which we discuss in Section 5.

Now that we know $I(\mathbb{M}) = \kappa$ in the optimal mechanism, we can use Equation (5), along with a well-known result from the literature, to put a lower bound on the number of messages in any κ-optimal mechanism:

(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.*

Corollary 1 *In any κ-optimal mechanism $\mathbb{M} = \langle M, p, q \rangle$, the message set M consists of at least e^{κ} elements.*

---

[13] 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.

On the other hand, we show that when the privacy constraint is stringent, 2 messages are sufficient to implement the optimal mechanism.[14]

Proposition 3 *There exists $\underline{\kappa} > 0$ such that in any κ-optimal interval mechanism $\mathbb{M} = \langle M, p, q \rangle$ with $0 < \kappa \leq \underline{\kappa}$, the message set M consists of exactly two intervals.*

This result is proved via a lemma stating that there can be at most one interval with arbitrarily small mass (according to F). Compared to the above proof sketch for Proposition 1, the next lemma additionally rules out the existence of two "small" intervals that are not adjacent.

Lemma 3 *For every $k > 0$, there exists $\epsilon > 0$ such that in any κ-optimal interval mechanism $\mathbb{M} = \langle M, p, q \rangle$ with $\kappa \leq k$, at most one interval in M has mass $< \epsilon$.*

Knowing that an optimal mechanism exists also allows us to derive the first order conditions for an interval partition to be optimal. To be concrete, let the intervals in M be $m_1 = [\theta_0, \theta_1], m_2 = [\theta_1, \theta_2], \ldots, m_n = [\theta_{n-1}, \theta_n]$, with $\underline{\theta} = \theta_0 < \theta_1 < \cdots < \theta_n = \overline{\theta}$. For brevity we denote $q_i \equiv q(m_i)$, so that q_i is the quantity offered to buyers with type in $[\theta_{i-1}, \theta_i]$. The following is a necessary condition for the "cutoffs" $\theta_1, \ldots, \theta_{n-1}$ to be optimal:

Lemma 4 *Given $n > 1$, if an interval mechanism $\mathbb{M} = \langle M, p, q \rangle$ maximizes profit among κ-feasible mechanisms with n intervals, then there exists a constant $\lambda \geq 0$ such that for all $i \in \{1, ..., n-1\}$:*

$$[(q_{i+1} - q_i) \cdot v(\theta_i) - (c(q_{i+1}) - c(q_i))] = \lambda \cdot \left[\log \frac{F(\theta_i) - F(\theta_{i-1})}{F(\theta_{i+1}) - F(\theta_i)}\right]$$

where, by Equation (3), q_i is determined by $c'(q_i) = \mathbb{E}_F[v(x) \mid x \in [\theta_{i-1}, \theta_i]]$.

Note that this lemma provides necessary conditions for optimality *given* the number of intervals n, but it does not characterize the optimal n for general distribution F and cost function $c(\cdot)$. Without imposing additional structure on these primitives, it is difficult to provide a complete characterization of κ-optimal mechanisms (which need not be unique) that describes the number of intervals and their properties. In light of this difficulty, we illustrate next the structure of a κ-optimal mechanism in the uniform-quadratic case.

[14] There are of course other optimal mechanisms that involve more (redundant) messages. This is why the following proposition is stated with the restriction to interval mechanisms.

4.3 Uniform-quadratic case

Suppose that $\theta \sim U[\underline{\theta}, \overline{\theta}]$ and $c(q) = \frac{q^2}{2}$. To find the κ-optimal mechanism for any given $\kappa > 0$ we proceed in two steps. First, given any $n \geq 1$, we find the profit-maximizing κ-feasible mechanism with exactly n intervals. We call this mechanism the (n, κ)-optimal mechanism. Then, we find the number of intervals n such that the (n, κ)-optimal mechanism yields the highest profit.

In the uniform-quadratic case, the optimal mechanism $\mathbb{M} = \langle M, p, q \rangle$ with n intervals admits a simple structure: If $\log n \leq \kappa$ then all the intervals in M have the same length; If $\log n > \kappa$ then a necessary condition is that M consists of exactly one "short" interval and $n-1$ equally "long" ones. The lengths are uniquely determined by the binding privacy constraint $I(\mathbb{M}) = \kappa$, and the position of the "short" interval within M does not matter. Formally:

Lemma 5 *In the uniform-quadratic case, given any $n \geq 1$ and $\kappa > 0$, the (n, κ)-optimal mechanism $\mathbb{M} = \langle M, p, q \rangle$ is such that*

1. *If $\log n \leq \kappa$ then M consists of n intervals of equal length $(= \frac{1}{n}\left(\overline{\theta} - \underline{\theta}\right))$.*
2. *If $\log n > \kappa$ then exactly one of the intervals in M has length l_s and the remaining $n-1$ intervals in M have length l_l. These lengths $l_s < l_l$ are uniquely determined by the following two equations:*

$$\overline{\theta} - \underline{\theta} = l_s + (n-1)\, l_l \tag{6}$$

$$\kappa = -\frac{l_s}{\left(\overline{\theta} - \underline{\theta}\right)} \log \frac{l_s}{\left(\overline{\theta} - \underline{\theta}\right)} - (n-1) \cdot \frac{l_l}{\left(\overline{\theta} - \underline{\theta}\right)} \log \frac{l_l}{\left(\overline{\theta} - \underline{\theta}\right)} \tag{7}$$

The mechanism is unique, up to reordering of the intervals.

The proof consists of three steps. First, we show that the order of the intervals in M does not change the expected profit when prices and quantities are optimally adjusted. Clearly it also does not affect the entailed loss of privacy. Next, we show that if the privacy constraint is binding, the first order conditions can be satisfied only if the intervals in M have at most two lengths (equivalently, two possible mass). Finally, we use the second order conditions to show that in any optimal solution, $n-1$ intervals have the same length and the last interval has weakly shorter length.

We now proceed to characterize the optimal number of intervals in the κ-optimal mechanism. Let n^*_κ denote the *smallest integer* for which $\log n^*_\kappa \geq \kappa$ (that is, $n^*_\kappa \in \mathbb{N}$ is such that $\kappa \in (\log(n^*_\kappa - 1), \log(n^*_\kappa)]$). We then have

Proposition 4 *In the uniform-quadratic case, given any κ, the κ-optimal mechanism is the (n^*_κ, κ)-optimal mechanism as described in Lemma 5.*

Proof: By Corollary 1, the number of intervals in a κ-optimal mechanism is at least $e^\kappa > n^*_\kappa - 1$. By way of contradiction, assume this number is greater than n^*_κ. Then by Lemma 5, at least n^*_κ of the intervals in M have the same "big" mass. It follows that each of the intervals in the mechanism $\mathbb{M}$ has mass smaller than $\frac{1}{n^*_\kappa} < e^{-\kappa}$, so that $-\log\left[F(\overline{m}) - F(\underline{m})\right] > \kappa$ for each $m \in M$. By Equation (5), we then have $I(\mathbb{M}) > \kappa \sum_m \left[F(\overline{m}) - F(\underline{m})\right] = \kappa$, leading to a contradiction. ■

Note that by Corollary 1, the number of messages in a κ-optimal mechanism in the uniform-quadratic case is equal to the lower bound on the number of intervals among all mechanisms that satisfy $H\left(g_M\right) = \kappa$. In other words, there are no partitions with less than n^*_κ intervals that exhaust the privacy constraint, and even though there are partitions with more than n^*_κ intervals that meet the privacy constraint, they are not optimal. The above proof actually shows that for $n > n^*_\kappa$, the (n, κ)-optimal mechanism does not exist if we require the mechanism to have *exactly* n intervals. The optimum is only achieved when $n - n^*_\kappa$ of these intervals are degenerate.

The structure of the κ-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 κ-optimal mechanism as a function of κ.

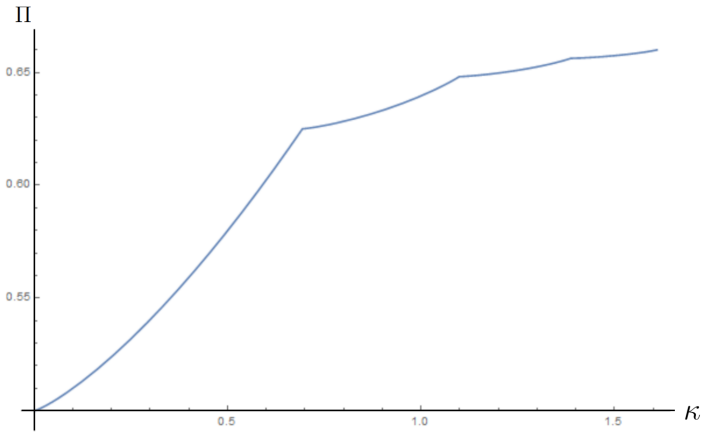

Figure 1. The privacy-profit frontier in the uniform-quadratic case

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. But as we continue to lower privacy, expected profit rises at an increasing rate until a new interval is added.

While the results in this section are derived for the uniform prior distribution, their qualitative properties (such as the shape of the privacy-profit frontier) are robust to small changes in this distribution. This is because the set of κ-optimal mechanisms, when viewed as a correspondence from the distribution F to the space of interval partitions, is *upper-hemicontinuous*.[15] To see this, recall that Lemma 2 expresses the seller's problem as a constrained optimization program. The objective function is clearly continuous, and the constraint $H(g_M) \leq \kappa$ is both upper- and lower-hemicontinuous.[16] So the set of optimizers is upper-hemicontinuous by the Theorem of the Maximum.

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 provide a thorough analysis of how κ changes these quantities. Throughout this section we assume quadratic costs, that is $c(q) = \frac{q^2}{2}$. We also 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.

It is immediate to notice that the expected profit Π is at least weakly increasing in κ. This is because higher κ only relaxes the privacy constraint (P) in the seller's problem. Moreover, by Proposition 2 we know that the privacy constraint is binding in κ-optimal mechanisms. So we have the following stronger result:

---

[15] One metric on the space of finite partitions is the following: If M consists of cutoffs $\{\theta_0, \ldots, \theta_n\}$ and M' consists of cutoffs $\{\theta'_0, \ldots, \theta'_m\}$, 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 verse for each θ'_j.

[16] To show it is lower-hemicontinuous, let M be a partition with cutoffs $\{\theta_0, \ldots, \theta_n\}$ such that $\sum_{i=1}^{n} -[F(\theta_i) - F(\theta_{i-1})] \cdot \log[F(\theta_i) - F(\theta_{i-1})] \leq \kappa$. Take any sequence of distributions F^j that converge (weakly) to F. We define M^j to be the partition with cutoffs θ_i^j given by $F^j(\theta_i^j) = F(\theta_i)$, for all $1 \leq i \leq n$. Then the privacy measure of M^j under the prior F^j is the same as the privacy measure of M under the prior F, so that M^j is κ-feasible under F^j. Weak convergence in the distribution implies $\theta_i^j \to \theta_i$ as $j \to \infty$. Hence we have lower-hemicontinuity.

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

Therefore, profit is minimized when the monopolist is required to provide full privacy ($\kappa = 0$), and it is maximized when he is allowed to fully separate the buyer types ($\kappa = \infty$).

For the buyer, the opposite is true:

Proposition 5 *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.*

The intuition for this result is that whenever the seller obtains finer information about the buyer in the form of dividing an interval into two subintervals, the buyer is worse-off in terms of ex-ante expected utility.

Finally, a regulator might be interested in finding the level of κ that maximizes total welfare. When the density function $f(\cdot)$ is monotone, the following proposition provides a characterization:

Proposition 6 *Suppose 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 $\kappa = \infty$.*

It is interesting to note that in the uniform-quadratic case, total welfare of any κ-optimal mechanism is independent of the privacy capacity κ.

5 Ex-post privacy-constrained mechanisms

So far we have analyzed an ex-ante notion of privacy loss. Under this criterion, the monopolist can satisfy the privacy constraint even when he learns almost perfectly about some small sets of buyer types (although by Lemma 3, there is at most one such set *at the optimum*). In this section we explore a more stringent notion of privacy, requiring that the designer not to learn too much about *any* buyer type. Formally, we strengthen the *average* privacy constraint $I(\mathbb{M}) \leq \kappa$ to its *ex-post* version:

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

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

where the supremum is taken over messages $m \in M$ that are sent with positive probability in equilibrium.

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

It turns out that ex-post κ-optimal mechanisms also take the interval partition form:

Proposition 7 *There exists an ex-post 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.*

Note that the ex-post privacy constraint directly implies that each interval in M has mass *at least* $e^{-\kappa}$. So in contrast to Corollary 1, here the total number of intervals in M is *bounded above* by e^{κ}. This upper bound also makes it easier to establish the existence of an optimal mechanism, since compactness is now guaranteed.

As another contrast with the ex-ante privacy notion, we observe that the ex-post privacy constraint is in general not exhausted in the optimal mechanism. This is a simple corollary of the following characterization in the uniform-quadratic case:

Proposition 8 *In the uniform-quadratic case, given any $\kappa \in [\log(n), \log(n+1))$, the ex-post κ-optimal mechanism divides the type space into n equal intervals.*

When $\kappa = \log(n)$, the ex-ante and ex-post constrained-optimal mechanisms coincide and exhaust both privacy constraints. But when $\log(n) < \kappa < \log(n+1)$, the ex-ante κ-optimal mechanism consists of $n+1$ intervals (not all equal), whereas the ex-post κ-optimal mechanism contains n equal intervals. In this case the ex-post privacy constraint is slack.

6 Discussion

6.1 Revelation principle

The revelation principle typically 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 the transitions from one stage to another are deterministic. However, sequential mechanisms with *random transitions* may do better than any one-shot mechanism. To illustrate this, suppose that the agent's type θ is uniformly distributed on $[0,1]$. Consider first the following one-shot mechanism. If the agent reports $\theta > 0.5$, the good is offered at price 0.375; if he reports $0.25 < \theta \leq 0.5$, then with probability $\frac{1}{2}$ the good is offered at price 0.25 and otherwise no interaction occurs; finally if he reports $\theta < 0.25$, no interaction occurs;. It is easy to check that every agent type is willing to truthfully report the interval that contains his type. Through this mechanism, the designer learns whether θ belongs to $[0, 0.25], [0.25, 0.5]$ or $[0.5, 1]$.

Now we construct a sequential mechanism that gives the agent the same incentives but preserves more privacy. In the first stage, the agent is asked whether his type is above or below 0.5. If he says "above," the good is offered at price 0.375. If he says "below," then with probability $\frac{1}{2}$ the interaction ends. With remaining $\frac{1}{2}$ probability the mechanism enters the second stage, in which the agent is asked whether θ is above or below 0.25. If yes, the good is offered at price 0.25; otherwise no trade occurs. Although the agent's incentives (and profit) are the same as in the previous one-shot mechanism, the ex-ante privacy constraint is relaxed since when $\theta \leq 0.5$, the designer only learns the exact interval $[0, 0.25]$ or $[0.25, 0.5]$ with probability $\frac{1}{2}$. Note however that it would be without loss to restrict to one-shot mechanisms under the *ex-post* privacy notion, which considers the most informative realization.

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.[17] One approach is to measure the average loss of privacy across all agents. An alternative approach is to require that the maximal loss of privacy for any agent is at most κ. 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 a sequential mechanism may preserve more privacy by collecting information from a small number of agents. For example, a

[17] This is particularly important since the literature on optimal mechanisms with restricted message spaces has highlighted the usefulness of asymmetric mechanisms.

descending price auction is strategically equivalent to a sealed-bid first price auction, but collects information only about the winning bidder.[18]

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 two 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 the virtual valuation is non-negative. We restrict attention to symmetric static mechanisms and the ex-ante privacy measure (each agent's privacy loss $\leq \kappa$).

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, consider the class of mechanisms where the types are partitioned into intervals, each buyer reports the interval to which his type belongs, and the higher bidder is awarded the good (with ties broken evenly). The optimal mechanism within this class turns out to be very similar to the optimal mechanism we derived for a single buyer in the uniform-quadratic case:

Proposition 9 *The optimal symmetric static mechanism in the two buyer problem with uniform distribution and no production costs partitions the types in the same way as the κ-optimal mechanism in the one buyer problem with quadratic costs. That is, for any $\kappa \in (\log(n-1), \log(n)]$, M consists of $n-1$ intervals of length l_l and one interval of length l_s. These lengths $l_s \leq l_l$ are uniquely determined as in Lemma 5.*

A pricing rule that assures incentive compatibility in this mechanism is the following: the winner pays $\frac{\theta_i \cdot \theta_{i-1} - \underline{\theta}^2}{\theta_i + \theta_{i-1} - 2\underline{\theta}}$, where $[\theta_{i-1}, \theta_i]$ is the interval that he reported, and the loser pays 0. Extending the analysis to any number of bidders is more involved and is left for future research.

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.

[18] The revelation principle would be maintained under the most stringent privacy measure, which considers the biggest loss of privacy across all agents and all of their type realizations.

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-ante stage (i.e., averaging over the possible realizations of the consumer type, the loss of privacy must be below some bound) and at the ex-post stage (i.e., for every realized type, the loss of privacy must be below some bound). We also demonstate 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? We hope that future research will provide answers to these and related questions.

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., 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, NY, USA.

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

[5] **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.

[6] **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, NY, USA.

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

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

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

[10] **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.

[11] **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.

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

[13] **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, NY, USA.

[14] **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, NY, USA.

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

[16] **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.

[17] **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.

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

[19] **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.

[20] **Kokolakis, S.** 2017. "Privacy Attitudes and Privacy Behavior: A Review of Current Research on the Privacy Paradox Phenomenon." *Computers and Security*, 64: 122-134.

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

[22] **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.

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

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

[25] **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.

[26] **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, USA.

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

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

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

[30] **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, NY, USA.

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

[32] **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).

[33] **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.

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

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

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

[37] **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.

8 Appendix 1 - Proofs

8.1 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.[19]

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

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

By the single-crossing property of the agent's preferences, the set $\mu(m)$ is either an interval or a singleton.[20] However, since κ is finite, there can be only a zero-measure

---

[19] 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.

[20] 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 $q(m)\theta'' - p(m) \geq q(m')\theta'' - p(m')$. 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'

subset of messages $m \in M$ for which $\mu(m)$ is a singleton.[21] 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. ■

8.2 Proof of Proposition 1

8.2.1 Core argument

We follow the proof sketch outlined in the main text, leaving some technical details to later subsections. As described in the main text, 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 κ-feasible interval mechanism contains a "big" interval with mass $\geq l$ (according to F). Indeed, from Equation 5 we have

$$I(\mathbb{M}) = \sum_{m \in M} -\left[F(\overline{m}) - F(\underline{m})\right] \cdot \log\left[F(\overline{m}) - F(\underline{m})\right] \leq \kappa.$$

Note that $\sum_{m \in M} \left[F(\overline{m}) - F(\underline{m})\right] = 1$. So there exists some $m \in M$ such that $-\log\left[F(\overline{m}) - F(\underline{m})\right] \leq \kappa$. In other words, the interval m has mass at least $e^{-\kappa}$.

---

and θ'', $\sigma(\theta)$ puts probability 1 on sending the message m.

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

Fixing this choice of l, we define ϵ to be a small positive constant as given by Lemma 7 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 7 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 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.[22] 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 (4 distinct ones) all with mass $< \epsilon$, then we combine both pairs at the same time and

[22] 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.

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 7 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. ■

8.2.2 Estimate of profit gain/loss

Lemma 6 *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 (2) and (3) that Δ only depends on a, b, c and is independent of the remaining cutoffs.

Proof: We define two auxiliary functions. First, we implicitly define the function $\phi(x)$ as follows: $c'(\phi(x)) = x$ for all $x > 0$. By Equation (3) we have that $q(m) = \phi(\mathbb{E}v(m))$ for all $m \in M$.

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 $q > 0$, we deduce that $c''(\phi(x))$ is bounded above and away from zero whenever $\phi(x)$ is, which in turn holds when x is bounded above and away from zero. Thus for all $x \in [v(\underline{\theta}), v(\overline{\theta})]$, $\phi'(x)$ is bounded above and away from zero.

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 $x \in [v(\underline{\theta}), v(\overline{\theta})]$.

We now estimate the profit reduction when merging two intervals into a single one. Let $\mathbb{E}v(m)$ denote $\mathbb{E}_F[v(\theta) \mid \theta \in [\underline{m}, \overline{m}]]$. Then the profit of mechanism $\mathbb{M} = \langle M, p, q \rangle$ as given by Equations (2) and (3) can be rewritten as

$$\Pi(\mathbb{M}) = \sum_{m \in M} h(\mathbb{E}v(m)) \cdot [F(\overline{m}) - F(\underline{m})]$$

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

$$\begin{aligned} \Delta = \quad & h(\mathbb{E}[v(\theta) \mid a \le \theta \le b]) \cdot [F(b) - F(a)] + h(\mathbb{E}[v(\theta) \mid b \le \theta \le c]) \cdot [F(c) - F(b)] \\ & - h(\mathbb{E}[v(\theta) \mid a \le \theta \le c]) \cdot [F(c) - F(a)]. \end{aligned} \tag{8}$$

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

$$v_1 \cdot [F(b) - F(a)] + v_2 \cdot [F(c) - F(b)] = \int_a^b v(\theta) f(\theta) d\theta + \int_b^c v(\theta) f(\theta) d\theta = v \cdot [F(c) - F(a)]. \tag{9}$$

Thus from Equation (8) 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 (8) and using (9), 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 (9) 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.

8.2.3 Comparison of two profit changes

Lemma 7 *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 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 6. Meanwhile, Equation (5) implies that the amount of privacy measure saved is

$$\alpha = (x + y) \log(x + y) - x \log x - y \log y = x \log(1 + \frac{y}{x}) + y \log(1 + \frac{x}{y}). \qquad (10)$$

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[23]

$$\delta \cdot |\log \delta| \geq \frac{\alpha}{2}. \tag{11}$$

We claim that (10) and (11) 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 6, 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.

8.3 Proof of Proposition 2

This result follows directly from Lemma 6 above: If the privacy constraint were slack, the seller could divide any interval in M into two subintervals and strictly increase the profit. By choosing one of the subintervals to be very small, he would still satisfy the privacy constraint. This contradicts optimality. ■

---

[23] 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 $\delta\cdot|\log\delta| \geq \frac{\alpha}{2}$.

8.4 Proof of Proposition 3

We argue that Proposition 3 follows from Lemma 3, which we prove below. Indeed, that lemma implies the existence of some $\epsilon > 0$ such that any 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 (5) and feasibility implies

$$\sum_{m \in M} -\left[F(\overline{m}) - F(\underline{m})\right] \cdot \log\left[F(\overline{m}) - F(\underline{m})\right] \leq \kappa \leq -\epsilon \log \epsilon.$$

In particular, $\left[F(\overline{m}) - F(\underline{m})\right] \cdot \log\left[F(\overline{m}) - F(\underline{m})\right] > \epsilon \log \epsilon$ holds for every interval $m \in M$. 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 $F(\overline{m}) - F(\underline{m}) < \epsilon$, or $F(\overline{m}) - F(\underline{m}) > \frac{1}{2}$ (which is a rough estimate).

In other 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 κ-optimal interval mechanism with $\kappa \leq \underline{\kappa}$ consists of at most two intervals. Since the privacy constraint is exhausted, exactly two intervals are employed. ■

8.5 Proof of Lemma 3

In the proof of Proposition 1, we showed that in any κ-feasible mechanism there is a "big" interval of mass at least $e^{-\kappa} \geq e^{-k}$. So by Lemma 7, 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 7, 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 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 $[\tilde{\theta}_t, \theta_{t+1}]$. Recall from the proof of Lemma 6 that

$$\Pi(\mathbb{M}) = \sum_{m \in M} h(\mathbb{E}v(m)) \cdot (F(\overline{m}) - F(\underline{m})).$$

Thus the profit increase in each step is given by

$$\begin{aligned} \Delta_t = \quad & h(\tilde{u}) \cdot \Big[F(\tilde{\theta}_t) - F(\tilde{\theta}_{t-1})\Big] + h(\tilde{w}) \cdot \Big[F(\theta_{t+1}) - F(\tilde{\theta}_t)\Big] \\ & - h(u) \cdot \Big[F(\theta_t) - F(\tilde{\theta}_{t-1})\Big] - h(w) \cdot [F(\theta_{t+1}) - F(\theta_t)] \end{aligned} \tag{12}$$

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 $h(\tilde{w}) \cdot \Big[F(\theta_{t+1}) - F(\tilde{\theta}_t)\Big] - h(u) \cdot \Big[F(\theta_t) - F(\tilde{\theta}_{t-1})\Big]$. 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 6,

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

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

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

Next we consider the other difference $h(\tilde{u}) \cdot \Big[F(\tilde{\theta}_t) - F(\tilde{\theta}_{t-1})\Big] - h(w) \cdot [F(\theta_{t+1}) - F(\theta_t)]$ in Equation (12). 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 (12), 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[24]

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

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 7. This again contradicts optimality, and hence there cannot even be two small intervals that are non-adjacent. ■

[24] 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 7, their sum is still $O(x^2)$.

8.6 Proof of Lemma 4

By Lemma 2, the seller's problem is to find a vector of cutoffs $\omega = \left\{\theta_0 = \underline{\theta}, \theta_1, \cdots, \theta_{n-1}, \theta_n = \overline{\theta}\right\}$ that maximizes the expected profit

$$\Pi\left(\omega, q_i^*\left(\omega\right)\right) := \sum_{i=1}^{n}\left[q_i^*\left(\omega\right)\int_{\theta_{i-1}}^{\theta_i} v\left(x\right) f\left(x\right) dx - c\left(q_i^*\left(\omega\right)\right)\cdot\left[F\left(\theta_i\right) - F\left(\theta_{i-1}\right)\right]\right]$$

subject to the constraint

$$I\left(\omega\right) := \sum_{i=1}^{n} -\left[F\left(\theta_i\right) - F\left(\theta_{i-1}\right)\right]\cdot\log\left[F\left(\theta_i\right) - F\left(\theta_{i-1}\right)\right] \leq \kappa,$$

where $q_i^*\left(\omega\right)$ is determined by $c'\left(q_i\right) = \mathbb{E}_F\left[v\left(x\right) \mid x \in [\theta_{i-1}, \theta_i]\right]$.

We form the Lagrangian

$$\mathcal{L}\left(\omega\right) = \Pi\left(\omega, q_i^*\left(\omega\right)\right) + \lambda(\kappa - I\left(\omega\right))$$

where $\lambda \in \mathbb{R}^+$ is the Lagrange multiplier. By the envelope theorem we have that $\frac{d\Pi}{d\theta_i} = \frac{\partial\Pi}{\partial\theta_i}$, and therefore the first order conditions, that is $\frac{\partial\mathcal{L}(\theta)}{\partial\theta_i} = 0$ for all $i \in \{1, \ldots, n-1\}$, are given by (after cancelling out $f(\theta_i)$):

$$\left[\left(q_{i+1} - q_i\right)\cdot v\left(\theta_i\right) - \left(c(q_{i+1}) - c(q_i)\right)\right] = \lambda\left[\log\frac{F\left(\theta_i\right) - F\left(\theta_{i-1}\right)}{F\left(\theta_{i+1}\right) - F\left(\theta_i\right)}\right]$$

for all $i \in \{1, \ldots, n-1\}$. ■

8.7 Proof of Lemma 5

Given $\kappa > 0$ and $n \geq 1$, we will characterize the partition of Θ into n intervals that maximizes the expected profit subject to the privacy constraint. We represent any partition as a vector of cutoffs $\omega = (\theta_0, ..., \theta_n)$, such that $\underline{\theta} = \theta_0 \leq \theta_1 \leq \cdots \leq \theta_{n-1} \leq \theta_n = \overline{\theta}$.

Our proof strategy is as follows. First, we will write the explicit expressions of the expected profit $\Pi\left(\omega\right)$ and the loss of privacy $I\left(\omega\right)$ that are induced by a vector of cutoffs ω. We verify that the expected profit depends only on the lengths of these intervals, so it is invariant to their ordering.[25] The same is true for the entailed loss

[25]For instance, the profit is the same for $\omega = (\theta_0, ..., \theta_{k-1}, \theta_k, \theta_{k+1}, ..., \theta_n)$ and for $\omega' = (\theta_0, ..., \theta_{k-1}, \theta_{k-1} + (\theta_{k+1} - \theta_k), \theta_{k+1}, ..., \theta_n)$

of privacy.

Next, we will write the first order conditions of the problem in terms of the interval mass (equivalently, lengths), and show that there can be at most two different mass in the optimal solution. Finally, we write the second order conditions and argue that exactly one interval has weakly smaller mass.

Invariance to ordering. 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 (3), is $\theta_i = \theta_i + \theta_{i-1} - \overline{\theta}$.

The profit as given by equation (2) 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: $\sum_{i=1}^{n} \left(\theta_i - \theta_{i-1}\right) = \left(\overline{\theta} - \underline{\theta}\right)$ and $\sum_{i=1}^{n} \left(\theta_i^2 - \theta_{i-1}^2\right) = \left(\overline{\theta}^2 - \underline{\theta}^2\right)$ and $\sum_{i=1}^{n} \left(\theta_i - \theta_{i-1}\right) \left(\theta_i + \theta_{i-1}\right)^2 = \frac{4}{3}\left(\overline{\theta}^3 - \underline{\theta}^3\right) - \frac{1}{3}\sum_{i=1}^{n} \left(\theta_i - \theta_{i-1}\right)^3$.[26] Plugging the three expressions back, we deduce

$$\Pi(\omega) = \left(\frac{1}{6}\left(\overline{\theta} - \underline{\theta}\right)^2 + \frac{1}{2}\underline{\theta}^2 - \frac{1}{6\left(\overline{\theta} - \underline{\theta}\right)} \sum_{i=1}^{n} \left(\theta_i - \theta_{i-1}\right)^3 \right) \tag{13}$$

This expression depends only on the lengths $\theta_i - \theta_{i-1}$.

The loss of privacy that is entailed by the partition ω is:

$$I(\omega) = -\sum_{i=1}^{n} \frac{\left(\theta_i - \theta_{i-1}\right)}{\overline{\theta} - \underline{\theta}} \log \frac{\left(\theta_i - \theta_{i-1}\right)}{\overline{\theta} - \underline{\theta}} \tag{14}$$

This is also invariant to the ordering of the intervals.

First order conditions. Let $x_i = \frac{\theta_i - \theta_{i-1}}{\overline{\theta} - \underline{\theta}}$ 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 $\{\theta_i\}$.

For given κ and n, Equations (13) and (14) suggest that the seller faces the

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

following constrained minimization problem:

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

The Lagrangian is given by:

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

Whenever $\{x_i\}$ is a local constrained minimizer, the first order conditions imply

$$3x_i^2 - \beta \log x_i = \alpha + \beta \qquad \text{for all } 1 \leq i \leq n. \tag{15}$$

If $\beta \leq 0$, then the function $3x^2 - \beta \log x$ is monotonically increasing. Thus every x_i is the same. This corresponds to the case where $\log n \leq \kappa$; it is clear that $\min \sum_{i=1}^n x_i^3$ is achieved when each $x_i = \frac{1}{n}$, and the privacy constraint is slack.

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 (15) yields that x_i can take at most two values $\underline{x}$ and $\overline{x}$, with $\underline{x} < \hat{x} < \overline{x}$.

Second order conditions. We next 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 derivative/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), p. 468.

Summary. By the above analysis, if $\log n \leq \kappa$ then the optimal solution involves

equally long intervals and satisfies the privacy constraint with slackness. If $\log n > \kappa$ then having equally long intervals violates the privacy constraint. So it must hold that $\beta > 0$ and x_i takes two values. Moreover, exactly one x_i takes the smaller value, which then pins down the different x_i as described in the lemma. ■

8.8 Proof of Proposition 5

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

$$\int \int_{\theta \leq \hat{\theta}} q(\theta) d\theta \; dF(\hat{\theta}) = \int q(\theta)(1 - F(\theta)) d\theta. \tag{16}$$

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 $q_j = \mathbb{E}[v(x) \mid x \in [\theta_{j-1}, \theta_j]]$, $q_{j+1} = \mathbb{E}[v(x) \mid x \in [\theta_j, \theta_{j+1}]]$, and $q = \mathbb{E}[v(x) \mid x \in [\theta_{j-1}, \theta_{j+1}]]$. Then 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 $\Delta \geq 0$, which implies the proposition.[27] Indeed, observe that

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

So $(q - q_j)(F(\theta_j) - F(\theta_{j-1})) = (q_{j+1} - q)(F(\theta_{j+1}) - F(\theta_j))$. Thus, $\Delta \geq 0$ is equivalent to

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

This holds because the LHS is just $\frac{\int_{\theta_{j-1}}^{\theta_j} (1 - F(\theta)) d\theta}{\int_{\theta_{j-1}}^{\theta_j} f(\theta) d\theta}$, which is at least $\frac{1 - F(\theta_j)}{f(\theta_j)}$ by the assumption that $\frac{1 - F(\theta)}{f(\theta)}$ is decreasing. Similarly the RHS of the above equation is at

[27] 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.

most $\frac{1-F(\theta_j)}{f(\theta_j)}$. Hence $\Delta \geq 0$ and the proposition follows.[28] ■

8.9 Proof of Proposition 6

Total welfare from a buyer of type θ is given by $\theta q(\theta) - (q(\theta))^2/2$. 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

$$\int q(\theta) \cdot \left(\theta - \frac{v(\theta)}{2}\right) f(\theta) d\theta. \tag{17}$$

Compared with the above Equation (16) for buyer surplus, the difference here is that the function $\left(\theta - \frac{v(\theta)}{2}\right) f(\theta)$ takes the place of $1 - F(\theta)$. If the function $\theta - \frac{v(\theta)}{2}$ decreases in θ, then the same argument as before shows that combining two intervals increases total welfare, which must be maximized at $\kappa = 0$ and minimized at $\kappa = \infty$.

It remains to show $\theta - \frac{v(\theta)}{2}$ is decreasing whenever $f(\theta)$ is increasing. This is because $2\theta - v(\theta) = \theta + \frac{1-F(\theta)}{f(\theta)}$, whose derivative is $-\frac{(1-F(\theta))f'(\theta)}{(f(\theta))^2}$. Thus the first half of the proposition is proved. The second half is proved by a symmetric argument: If $f(\theta)$ is decreasing then $\theta - \frac{v(\theta)}{2}$ is increasing, and combining two intervals decreases total welfare. ■

8.10 Proof of Proposition 7

Directly following the proof of Lemma 1, we can transform any ex-post κ-feasible mechanism into an interval mechanism that achieves the same profit and still satisfies the ex-post privacy constraint. The only step that requires some care is in eliminating "duplicate" messages. If m' is a duplicate of m, then after removing m' from M and adjusting the equilibrium, the seller's posterior belief given the message m in the new mechanism is an average of his posterior beliefs given the messages m and m' in the original mechanism. Both of these posterior beliefs have relative entropy at most κ from the prior, and so does the average belief. Thus removing duplicates preserves the ex-post privacy constraint.

Suppose $m \in M$ is an interval, then the relative entropy between the posterior upon seeing m and the prior is simply $-\log\left[F(\overline{m}) - F(\underline{m})\right]$. So the privacy constraint

[28] This argument generalizes to any cost function with $c''' \geq 0$.

requires that each interval m has mass at least $e^{-\kappa}$ according to F. It follows that any ex-post κ-feasible mechanism contains *at most* e^{κ} intervals. Hence an optimal interval mechanism exists by compactness. ■

8.11 Proof of Proposition 8

When $\kappa \in [\log(n), \log(n+1))$, any feasible interval mechanism has at most $e^{\kappa} < n+1$ intervals. Now recall from the proof of Lemma 5 that in the uniform-quadratic case, having n equal intervals achieves the greatest profit among all partitions with at most n intervals (even when the privacy constraint is ignored).[29] Thus the mechanism with n equal intervals, which is ex-post κ-feasible, must be the ex-post κ-optimal mechanism. ■

8.12 Proof of Proposition 9

Consider an arbitrary partition with cutoffs $\{\underline{\theta} = \theta_0, \ldots, \theta_n = \overline{\theta}\}$. For each interval $[\theta_{i-1}, \theta_i]$, the probability of winning is computed as

$$q_i = \frac{\theta_{i-1} - \underline{\theta} + (\theta_i - \theta_{i-1})/2}{\overline{\theta} - \underline{\theta}} = \frac{(\theta_i + \theta_{i-1})/2 - \underline{\theta}}{\overline{\theta} - \underline{\theta}},$$

which is the probability that the opponent type belongs to a lower interval or it belongs to the same interval and the tie is broken favorably. By the envelope theorem, type θ_i's interim expected utility is thus

$$u_i = \sum_{j=1}^{i} (\theta_j - \theta_{j-1}) \cdot q_j = \sum_{j=1}^{i} (\theta_j - \theta_{j-1}) \cdot \frac{(\theta_j + \theta_{j-1})/2 - \underline{\theta}}{\overline{\theta} - \underline{\theta}} = \frac{(\theta_i - \underline{\theta})^2}{2(\overline{\theta} - \underline{\theta})}$$

after some simplification. It follows that the expected payment when reporting the interval $[\theta_{i-1}, \theta_i]$ is given by

$$p_i = \theta_i \cdot q_i - u_i = \frac{\theta_i \theta_{i-1} - \underline{\theta}^2}{2(\overline{\theta} - \underline{\theta})}.$$

---

[29] From that proof, we know the seller seeks to minimize $\sum_{i=1}^{n} x_i^3$ subject to $\sum_{i=1}^{n} x_i = 1$. The minimum is clearly achieved when each $x_i = \frac{1}{n}$.

Therefore total profit from both buyers equals

$$\Pi(\mathbb{M}) = \frac{1}{(\overline{\theta} - \underline{\theta})^2} \sum_{i=1}^{n} (\theta_i \theta_{i-1} - \underline{\theta}^2) \cdot (\theta_i - \theta_{i-1}).$$

Since $\sum_i \underline{\theta}^2 \cdot (\theta_i - \theta_{i-1}) = \underline{\theta}^2 \cdot (\overline{\theta} - \underline{\theta})$ is a constant, the seller seeks to maximize the expression $\sum_i \theta_i \theta_{i-1}(\theta_i - \theta_{i-1})$. Now observe that

$$3 \sum_{i=1}^{n} \theta_i \theta_{i-1}(\theta_i - \theta_{i-1}) = \sum_{i=1}^{n} \left[\theta_i^3 - \theta_{i-1}^3 - (\theta_i - \theta_{i-1})^3\right] = \overline{\theta}^3 - \underline{\theta}^3 - \sum_{i=1}^{n} (\theta_i - \theta_{i-1})^3.$$

Hence the seller equivalently minimizes $\sum_{i=1}^{n} (\theta_i - \theta_{i-1})^3$. But recall from the proof of Lemma 5 that this is also the objective in the single-buyer uniform-quadratic case. Since the privacy constraint is also the same, so must be the solution. ■