Theoretical Economics 16 (2021), 1557–1603 1555-7561/20211557

Bayesian privacy

Ran Eilat
Department of Economics, Ben-Gurion University of the Negev

Kfir Eliaz
Department of Economics, Tel-Aviv University and Department of Finance, University of Utah

Xiaosheng Mu
Department of Economics, Princeton University

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 their participants. We contribute to this discussion by proposing a Bayesian approach to measure loss of privacy in a mechanism. Specifically, we define the loss of privacy associated with a mechanism as the difference between the designer's prior and posterior beliefs about an agent's type, where this difference is calculated using Kullback–Leibler divergence, and where the change in beliefs is triggered by actions taken by the agent in the mechanism. We consider both ex post (for every realized type, the maximal difference in beliefs cannot exceed some threshold κ) and ex ante (the expected difference in beliefs over all type realizations cannot exceed some threshold κ) measures of privacy loss. Applying these notions to the monopolistic screening environment of Mussa and Rosen (1978), we study the properties of optimal privacy-constrained mechanisms and the relation between welfare/profits and privacy levels.

Keywords. Privacy, mechanism-design, relative entropy.

JEL classification. D47, D82.

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

Ran Eilat: eilatr@bgu.ac.il
Kfir Eliaz: kfire@tauex.tau.ac.il
Xiaosheng Mu: xmu@princeton.edu

We thank Benny Moldovanu and Doron Ravid for helpful conversations. Xiaosheng Mu acknowledges the hospitality of Columbia University and the Cowles Foundation at Yale University, which hosted him during parts of this research.

 https://doi.org/10.3982/TE4390

amounts of data on individuals' characteristics such as their tastes and willingness-to-pay for products and services. These data are valuable to a variety of entities including commercial firms as well as political institutions. If leaked to these entities, this information may be used against the users' interests. In light of this, there has been a growing sentiment that governments should enact laws that regulate the ability of private entities to collect and use personal information. If the growing concerns for maintaining privacy were to lead to regulations that impose privacy constraints on pricing mechanisms, how would that affect the design of these mechanisms, and what is the trade-off between profits, welfare and privacy? This paper takes a step towards addressing these questions by proposing a Bayesian approach to the measurement of loss of privacy and applying this approach to the design of optimal mechanisms that are restricted in the amount/precision of private information that they can elicit from participants.

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

Taking the idea of relative privacy to mechanism design, we suppose that the designer already possesses some information about the participants in the form of a prior belief over their "types." He updates this belief as a result of the participants' interactions with the mechanism, which reveal more information about them. For example, if the mechanism consists of a menu, then an individual's choice of a particular menu item enables the designer to learn additional information about this person's preferences. Consequently, the designer's posterior belief about the participant's type may be quite different from her prior belief. This suggests that the *difference* between the designer's prior and posterior beliefs should serve as the basis for measuring the loss of privacy associated with a particular mechanism.

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

their private types. Hence, the optimal solution entails complete loss of privacy: The designer has a degenerate posterior belief on the type of each participant.

To study the design of mechanisms that preserve some level of privacy, we follow the information theory literature and propose to measure a mechanism's inherent loss of privacy as the *maximal relative entropy* (i.e., the maximal Kullback–Leibler divergence) between the seller's posterior and prior beliefs. The posterior belief is derived from the "message" received in the mechanism, and the maximum KL divergence is computed across all messages (i.e., across all consumer types). We then augment the standard mechanism design problem by restricting attention to mechanisms with privacy loss at most κ. The parameter κ, which takes values between zero (full privacy) and infinity (no privacy), captures the strength of the privacy requirement. For any finite κ, our approach imposes an upper bound on the amount of new information that a designer may learn about any participant. While in our analysis we use the functional form of KL-divergence, our results extend to the more general class of *f-divergences* (see footnote 14 for details).

Before describing our results, we offer some discussion of this modeling framework. 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 (see, e.g., Acquisti and Grossklags (2005), Barth and de Jong (2017), and Kokolakis (2017)).[1] Our paternalistic approach is analogous to the one that is frequently taken in the design of commitment mechanisms, which try to help agents with self-control problems by limiting their choice sets even when there is no demand for such commitment devices on part of the agents.[2]

This perspective that a regulator imposes the exogenous privacy constraint motivates our choice of an information-theoretic measure of privacy loss. Our measure requires making no specific assumptions about the exact form of future interactions between the data holder and the agent. This approach reflects our view that the continuation game is often unknown (especially to the regulator), making it impossible to predict how data collected today will be used in the future.[3]

[1] Even if consumers were to fully endogenize the consequences of disclosing their private information, it is often the case that individuals neglect the effects of sharing their information on others. For a discussion of the negative externalities and the "public bad" properties of loss of privacy see, e.g., Fairfield and Engel (2015) and Choi et al. (2019).

[2] Laibson (2018) called it the commitment paradox and gave the following example: "College instructors adopt course policies that force students to focus on their course work: pop quizzes, classroom attendance requirements, cold calling, graded problem sets, deadlines, classroom wifi blocking, and classroom laptop bans. In my experience, most students do not welcome these paternalistic restrictions."

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

Thus, a regulator may prefer the relative entropy measure, which is portable and applicable to any mechanism design setting regardless of the particular type space. As our analysis demonstrates, relative entropy (or any f-divergence) also makes the problem especially tractable. In particular, our results suggest that the (ex post) relative entropy privacy measure admits an equivalent "reduced-form" implementation, where the regulator simply requires each outcome in the designer's menu to cater to a sufficiently big proportion of the population. This is discussed in further detail below.

Our main results highlight the following key properties of optimal privacy-preserving mechanisms. The optimal κ-constrained mechanism *partitions* the type space into finitely many intervals (whose number depends on κ), such that consumers truthfully announce to which interval their type belongs. Thus, even though there is a continuum of types, and the privacy constraint allows for a continuum of noisy messages,[4] maximal profits are attained with only finitely many messages. When the production cost function and the type distribution satisfy some additional properties that we provide, any optimal κ-constrained mechanism consists of intervals that are *monotonically ordered* according to their mass from left to right, so that the gain from differentiating higher types either diminishes or increases.

We show that an interval partition of the type space satisfies the κ-privacy constraint if and only if the probability mass of each interval is at least $e^{-\kappa}$. From a practical viewpoint, this latter constraint on interval size may be more easily implemented by a regulator. Specifically, to verify that a firm did not violate the privacy constraint, the regulator only needs to get the data on the proportion of consumers who bought each menu item. He then needs to check that each item is bought by at least $e^{-\kappa}$ fraction of the consumers. This equivalent implementation does not require the regulator to know the prior distribution of types, nor the equilibrium strategies. In fact, similar restrictions are already used in practice in related contexts. For instance, the US government often requires publicly released statistics (say, for research based on the US Census) to be based on cells containing a minimum number of people or firms. Our contribution is to show that such practices are optimal under a general notion of relative privacy.

We then study the welfare consequences of varying the privacy constraint κ. A priori, it is not clear whether stricter privacy requirements help or harm consumers since such requirements distort allocative efficiency. We show that when the marginal production cost is convex, buyer surplus is *maximized* at $\kappa = 0$, where every type receives the same quality, and it is minimized at $\kappa = \infty$, where types are fully separated. However, we find that when costs are quadratic and the prior density function is decreasing in the type, total welfare is *minimized* at $\kappa = 0$ (full privacy) and maximized at $\kappa = \infty$ (no privacy). This result suggests that regulators may face a trade-off between consumer protection and trade efficiency.

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

[4] For example, when each type θ reports $\theta + \varepsilon$, where ε is a continuous random variable.

in the interior of this interval, the privacy constraint is slack and profit is constant as κ increases within this interval.

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

We show that many of our findings continue to hold under the ex ante notion of privacy. However, there are a number of major differences. First, even though it is again without loss to consider mechanisms that involve interval partitions, existence of an optimal mechanism is no longer straightforward. The difficulty arises since the ex ante constraint, unlike our original ex post version, allows for partitions with countably many intervals. Without a finite upper bound on the number of intervals, it is more challenging to use a compactness argument to prove existence. To show that an optimal mechanism indeed exists, we show that there can be at most one interval of arbitrarily small mass, i.e., at most one set of types about which the seller attains very precise information. This property restores compactness and settles the existence issue; it also implies that for κ small (near full privacy), the optimal ex ante κ-constrained mechanism has exactly two intervals. Another distinction of the ex ante constraint is that the resulting optimal mechanism always exhausts the privacy constraint. Consequently, the optimal mechanism in the uniform-quadratic case is different from the one derived for ex post privacy: For $\kappa \in (\log(n-1), \log(n)]$, there is exactly one "short" interval and $n-1$ "long" intervals of equal length; the lengths are uniquely determined by the binding privacy constraint, and the position of the short interval does not matter.

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

We proceed as follows. In Section 2, we review the related literature. In Section 3, we introduce our ex post Bayesian privacy notion, and in Section 4 we use this notion to analyze optimal privacy-constrained mechanisms in a monopolistic screening setting. Section 5 presents an alternative, ex ante, Bayesian privacy measure and solves

for the corresponding optimal mechanisms under this constraint. Section 6 considers an extension where the seller possibly excludes a subset of consumers from trade, and Section 7 concludes.

2. Related literature

Our work is related to rational inattention (Sims (2003), Caplin and Dean (2015), Matějka and McKay (2015), Matějka (2016), Steiner et al. (2017), Caplin et al. (2019)). In that literature, an uninformed decision maker chooses the structure of a signal she wants to observe, subject to the information constraint that the signal can only contain a limited amount of information about the state (as measured by mutual information). Note that mutual information is equal to the expected KL divergence between the decision maker's posterior and prior beliefs. Thus, the rational inattention approach connects to our mechanism designer's problem under the ex ante privacy constraint. The key difference is that in our setting, the "signal" observed is the agent's message, which is bound by an additional incentive constraint. Thus, the seller in our model cannot directly choose what information to acquire, but rather needs to incentivize another party to disclose that information to her. Studying the interaction between the information constraint and the incentive constraint is the main objective of our work.

The privacy constraint in our model entails that, *in equilibrium*, agents cannot communicate all their private information to the designer. Several papers have investigated a related question of optimal mechanism design with limited communication, by imposing exogenous restrictions on the cardinality of the action space available to agents. Notable examples are Green and Laffont (1986, 1987), Melumad et al. (1992), Blumrosen et al. (2007), Bergemann et al. (2012), Kos (2012), and Blumrosen and Feldman (2013). In a different setting, Mookherjee and Tsumagari (2014) studied 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) studied the interaction between communication capacity and incentive feasibility by quantifying the "cost of selfishness"—the amount of excess information (bits) that needs to be exchanged to implement a given social choice function, relative to the case in which agents honestly report their types. More generally, optimization problems over partitions have been considered in Alonso and Matouschek (2008) and Frankel (2014), who study delegation, as well as in Crémer et al. (2007), who study language design within an organization.

A number of papers have studied privacy in dynamic models, where the agent's preferences for privacy is derived from how the designer can use the information against him in the future. Such models appear in Taylor (2004), Calzolari and Pavan (2006), Conitzer et al. (2012) among others. In contrast, our approach based on the divergence between prior and posterior beliefs is reduced-form and especially applicable to the question of regulation, where it is not a priori clear what the seller would/could do with the collected consumer data. Static models about privacy have been studied in Gradwohl (2018), who analyzes the problem of full implementation when agents prefer to protect their privacy, in Gilboa-Freedman and Smorodinsky (2018), who axiomatize preferences over privacy

and in Bird and Neeman (2020), who study how to regulate what information a firm can use when interacting with a consumer so as to maximize the consumer's welfare.[5] Finally, in the two-sided market setting, Hidir and Vellodi (2018) and Ichihashi (2019) have considered letting the consumer directly communicate to the seller via cheap talk or Bayesian persuasion. In our model, however, the consumer reveals information about his private type by responding to the seller's mechanism.

Turning to the computer science literature, our notion of privacy differs from the popular measure of "differential privacy."[6] Introduced by Dwork et al. (2006), differential privacy roughly means that changing the data of a single individual (or of a single attribute of an individual) should have a negligible effect on computations done on the entire data. In the context of mechanism design, Pai and Roth (2013) showed that this notion can be formalized as follows. Suppose there are n individuals, who each draws a private type from some set T. Define a mechanism M as a mapping from profiles of types $t \in T^n$ to distributions over some set of outcomes X. Then M is ϵ-*differentially private* if for all pairs of type profiles (t, t') that differ only in t_i, and for any payoff function $u : X \to \mathbb{R}$, it holds that

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

This definition implies that the message of a single player has a negligible effect on the outcome, such that any action is almost weakly dominant in the sense that it cannot affect a player's payoff by a multiplicative factor of more than 2ϵ, regardless of the other players' actions. In light of this, several studies in computer science have used the above notion to design mechanisms where truth-telling is either almost or exactly weakly dominant; see, e.g., McSherry and Talwar (2007), Kearns et al. (2014), and Nissim et al. (2012). Other studies attempt to incorporate agents with privacy concerns into mechanism design, assuming an additive cost for loss of privacy that increases with the level of differential privacy (i.e., with the ϵ above). Examples are Ghosh and Roth (2011), Ligett and Roth (2012), Fleischer and Lyu (2012).

What distinguishes our approach is that we measure privacy loss relative to a prior, whereas differential privacy is independent of the prior. Additionally, note that the differential privacy constraint is stated solely in terms of the (random) outcome distribution, sidestepping any action taken in the mechanism. This captures privacy loss from the perspective of an *outside observer*, but does not correspond to what the *mechanism designer* may learn through observing participants' choices. In this paper, we focus on constraining the amount of information learned by the mechanism designer, and study a benchmark in which the designer cannot commit not to access the data received.

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

[6]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).

Thus, we measure privacy loss using posterior beliefs that are derived from the messages sent, rather than the outcomes induced. In particular, the designer in our model *does not gain* from randomizing over allocations.

A notion related to differential privacy is that of "informational smallness" (see, e.g., Gul and Postlewaite (1992) and McLean and Postlewaite (2002)). In an environment with asymmetric information, an agent is informationally small if, given the prior distribution over states, the probability that revealing his information would have a non-negligible effect on the posterior distribution is small (assuming that the other agents reveal their information truthfully). Like our privacy measure, informational smallness is also prior-dependent. The main difference is that we measure the designer's reduction in uncertainty relative to the prior, whereas informational smallness measures the sensitivity of the posterior belief to each agent's report.

Another line of research in computer science deals with distortion and anonymization of databases and communication channels due to privacy concerns. This literature uses KL divergence to measure loss of privacy in a setting where a sender wants to send part of a dataset to a receiver in such a way that the receiver learns as little as possible about some other sensitive part of the data. The problem is formulated as finding the message that minimizes the KL divergence between the common prior on the sensitive data and the posterior, given the sender's message, subject to the constraint that the receiver attains some level of utility. Examples of works in this literature include Agrawal and Aggarwal (2001), Díaz et al. (2002), Rebollo-Monedero et al. (2009), Sankar et al. (2013), and Wang et al. (2016). However, these papers do not consider the strategic interaction between privacy, mechanism, and agent behavior, as we do here.

3. Bayesian privacy

A single agent has a privately known type $\theta \in \Theta$ and a type-dependent vNM utility function $u : Z \times \Theta \to \mathbb{R}$, where Z is the set of outcomes to be determined by a mechanism. The mechanism designer does not know the agent's type θ, but has a prior belief about θ represented by a probability measure F on Θ, with strictly positive density f (with respect to a base Borel measure). The designer's payoff function is $\pi : Z \times \Theta \to \mathbb{R}$.

To maximize expected payoff, the designer devises a mechanism $\mathbb{M} = \langle M, g \rangle$, where M is an arbitrary set of messages, and $g : M \to Z$ is a function that maps each message in M to an outcome $z \in Z$.[7] A strategy for the agent is a measurable function $\sigma : \Theta \to \Delta(M)$. A strategy σ^* is said to be an equilibrium in the mechanism $\mathbb{M}$ if each type $\theta \in \Theta$ best responds to $\mathbb{M}$.

At the outset, the designer already has some information about the agent in the form of his prior belief F. Given a mechanism $\mathbb{M}$ and an equilibrium σ^*, when the agent sends message $m \in M$, the designer updates her information to the posterior belief $F(\cdot|m, \sigma^*)$. This change of beliefs entails loss of privacy for the agent. For convenience, we define $F(\cdot|m, \sigma^*) = F$ for any message m sent with zero probability by every type in σ^*.

[7] We restrict the message set to be a Polish space, and endow it with its Borel σ-algebra. We let ΔM denote the set of all Borel probability measures over M endowed with the weak* topology.

We measure the loss of privacy entailed by a message $m \in M$ in an equilibrium σ^* by the *relative entropy* between the posterior belief triggered by m and the prior belief: If the posterior measure $F(\cdot|m, \sigma^*)$ has density $f(\cdot|m, \sigma^*)$, the relative entropy (or Kullback–Leibler divergence) from $F(\cdot|m, \sigma^*)$ to F is defined by[8]

$$D_{\mathrm{KL}}\big(F(\cdot|m, \sigma^*)\|F\big) = \int_\Theta f\big(\theta|m, \sigma^*\big) \cdot \log \frac{f\big(\theta|m, \sigma^*\big)}{f(\theta)}\, d\theta. \tag{1}$$

If instead $F(\cdot|m, \sigma^*)$ does not admit a density, we define $D_{\mathrm{KL}}(F(\cdot|m, \sigma^*)\|F) = +\infty$ (when $F(\cdot|m, \sigma^*)$ has an atom this definition preserves continuity). Throughout the paper, "log" represents the natural logarithm.

We define the *ex post loss of privacy* associated with an equilibrium σ^* in a mechanism $\mathbb{M}$ to be the *maximum* divergence between the possible posteriors and the prior.

Definition 1. *The ex post loss of privacy in an equilibrium σ^* of a mechanism $\mathbb{M}$ is given by*

$$I\big(\mathbb{M}, \sigma^*\big) = \sup_m \big[D_{\mathrm{KL}}\big(F\big(\cdot|m, \sigma^*\big)\|F\big)\big]$$

Thus, in equilibrium, the change in the designer's beliefs is no more than $I(\mathbb{M}, \sigma^*)$ regardless of the agent's type. In the next section, we study the problem of a designer who is restricted to use only mechanisms for which the above measure does not exceed some threshold.

Note that privacy loss, according to Definition 1, is not a property only of the mechanism $\mathbb{M}$—it also depends on which equilibrium is played. Thus, in case of multiple equilibria, to apply our notion of Bayesian privacy one needs to specify the relevant equilibrium. As we explain in the next section, this will not be an issue in the application that we analyze in this paper.

Note also that using relative entropy to quantify changes in beliefs does not bind our privacy measure to a particular metric over the set of types. To be specific, one could alternatively measure belief changes in a way that takes into account not only the relative probabilities between prior and posterior beliefs, but also where these probabilities lie according to a certain distance function between types (such as the Euclidean metric). However, sticking to any particular metric might be misleading because how the collected data will be used in future interactions is often unknown at present. Furthermore, a metric-based measure is highly sensitive to the context. In some contexts, there is greater loss of privacy in knowing that the agent is either of two "far-away" types (according to the adopted metric) than knowing that he is one of two "close-by" types, while the opposite is true in other contexts. For example, in the context of body weight where distance between types is measured in pounds, there is greater loss of privacy in knowing that an individual is either extremely obese or extremely underweight (i.e., far-away

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

types) than knowing that his weight is either 150 lbs or 160 lbs (close-by types). In contrast, in the context of geographic location where distance between types is measured in miles, there is greater loss of privacy in knowing that the agent's residential address is one of two adjacent streets in Provo, Utah, than in knowing that his address is either a street in Los Angeles or some street in New York.[9]

The above examples suggest that when information reveals the agent type to belong to a subset of types, privacy loss should be inversely related to the prior probability of this subset: finer information leads to greater privacy loss. Moreover, a metric-less measure of privacy loss should only depend on this prior probability. Our relative entropy measure exactly captures these properties; see equation (2) below.

Next, we apply our definition of privacy loss to the well-known problem of monopolistic screening.

4. Optimal privacy-constrained screening mechanisms

We consider the classic Mussa–Rosen (1978) set-up of monopolistic screening. A seller (she) wishes to sell a product of quality $q \in \mathbb{R}^+$ to a buyer (he), in exchange for payment $p \in \mathbb{R}$. Thus, the set of outcomes Z is the set of price and quality pairs. The seller's profit is given by

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

where $c(\cdot)$ is a twice-continuously differentiable cost function that satisfies $c(0) = c'(0) = 0$ and $c''(q) > 0$ for all $q \geq 0$.[10] The buyer's type $\theta \in [\underline{\theta}, \overline{\theta}]$ represents his willingness to pay per unit of quality, which is unknown to the seller. If the buyer consumes q and pays p, his utility is

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

We assume that the buyer's virtual valuation, $v(\theta) \equiv \theta - (1 - F(\theta))/f(\theta)$, is increasing in θ and satisfies $v(\underline{\theta}) \geq 0$. To facilitate some technical arguments, we make the slightly stronger assumption that v is continuously differentiable and $v' > 0$.

Positive virtual valuation allows us to focus on the case in which the seller wants to include all buyer types, and the only question is what quality and price should be offered to each buyer type. Nonetheless, our results extend to the case where $v(\underline{\theta})$ is negative; see Section 6 for details.

To sell the good, the monopolist uses a mechanism $\mathbb{M} = \langle M, p, q \rangle$, where $p : M \to \mathbb{R}^+$ 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 quality $q(m)$ and charges the price $p(m)$.[11]

The seller's expected profit in an equilibrium σ^* of a mechanism $\mathbb{M}$ is given by

$$\Pi\big(\mathbb{M}, \sigma^*\big) = \mathbb{E}_m\big[p(m) - c\big(q(m)\big)\big]$$

[9]Another related issue is that it is often not the collected data per se that matters for privacy considerations, but rather other variables that are correlated with the data. It is thus impossible to a priori choose one "right" metric that fits all possible relevant variables in all future scenarios.

[10]In this application, the designer's payoff function (i.e., profit) is independent of the agent's type θ.

[11]The functions $p(\cdot)$ and $q(\cdot)$ correspond to the outcome function $g(\cdot)$ in the previous section.

where $\mathbb{E}_m$ is evaluated according to the probability that each message $m \in M$ is sent under the strategy σ^*.

Note that a priori we allow the buyer to randomize over messages that he finds indifferent. However, it will follow from our analysis (Lemma 1 below) that almost all buyer types *do not randomize* at the optimum.

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

Recall that, in general, the privacy loss associated with a mechanism $\mathbb{M}$ depends on the equilibrium σ^* that is played. However, in our monopolistic screening setting multiple equilibria only arise when a positive measure of buyer types are indifferent between two messages, which must then lead to the same quality-price pair. As we discuss below, such messages are "wasteful" and without loss excluded from the optimal mechanism. Hence, to simplify the exposition, from now on we omit the reference to the equilibrium σ^* from the notation of privacy-loss. Thus, $I(\mathbb{M})$ is the privacy loss entailed by the (essentially unique) equilibrium of $\mathbb{M}$.

4.1 *The design problem*

A regulator requires the seller to design a mechanism that does not exceed some privacy capacity $\kappa > 0$. The seller's 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 $\Pi(\mathbb{M}) = \mathbb{E}_m[p(m) - c(q(m))]$ subject to three constraints:

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

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

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

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

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

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*.[12] 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

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

(Proposition 1)? Do some buyer types disclose more information than others (Proposition 2)? Is the privacy constraint even binding (Proposition 3 and ensuing discussion)?

4.2 *Interval mechanisms*

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

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 (see the Appendix for the formal proof). First, messages that lead to the same quality-price pair can be combined without affecting the outcome of the mechanism. Due to convexity of the relative entropy function, this also relaxes the privacy constraint. We can thus without loss assume that different messages in the mechanism are strictly ranked by the quality levels they are mapped into. Next, we argue that the set of types selecting the same message in equilibrium must form an interval; this is because buyer preference is single crossing in the type and quality served. Finally, any two of these intervals do not intersect each other except at the boundary, since the indifference condition holds for at most one type. The lemma then follows.[13]

We highlight that our model *does allow* the buyer to mix over different messages, in case he is indifferent. But Lemma 1 shows that buyer randomization *does not happen* in the optimal mechanism. This result stands in contrast to the rational inattention literature, where entropy-based constraints often lead to stochastic, nonpartitional signals. The distinction arises because in our setting, stochastic "signals" (i.e., messages) are constrained by the buyer's indifference condition and turn out to be costly for the seller.

It is also worth noting that the above argument only uses the convexity of KL divergence, so that combining "redundant" messages decreases the maximum relative entropy between prior and posterior beliefs. Because of this, Lemma 1 would continue

[13]The same argument and result hold even if the designer's objective is to maximize a weighted sum of profit and consumer surplus.

to hold if the privacy measure were an arbitrary convex function of the distribution of posterior beliefs (as often assumed in the literature on costly information acquisition).

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

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

Since the ex post privacy constraint requires this to be at most κ, we derive the following result.

LEMMA 2. *An interval mechanism is κ-feasible if and only if each of the intervals has mass at least $e^{-\kappa}$ according to the prior distribution F.*[14]

Taken together, Lemmas 1 and 2 imply that the seller's problem reduces to selecting a finite menu of quality-price pairs such that the mass of consumers who take any option from the menu must exceed $e^{-\kappa}$. This allows for a natural reformulation of the problem by specifying the privacy constraint in terms of $\lambda \equiv e^{-\kappa}$, the lowest admissible probability mass of an interval of pooled types. The advantage of using this alternative parametrization is that, in contrast to the parameter κ, the parameter λ is expressed in probability units which are easily interpretable.

Assuming now that the intervals in M are $m_1 = [\theta_0, \theta_1]$, $m_2 = [\theta_1, \theta_2]$, $\ldots$, $m_n = [\theta_{n-1}, \theta_n]$, with $\underline{\theta} = \theta_0 < \theta_1 < \cdots < \theta_n = \overline{\theta}$. Given the number n and cutoffs $\theta_1, \ldots, \theta_{n-1}$, we can derive the optimal quality and price that a κ-optimal mechanism assigns to each message. Specifically, note that the expected profit for the seller from employing this interval mechanism is given by[15]

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

[14] Essentially the same result holds if KL divergence is replaced by a more general class of *f-divergences* which take the form of $\int_\theta \phi[f(\theta \mid m)/f(\theta)]\, dF(\theta)$ for some convex function ϕ, with KL being the special case of $\phi(x) = x \log x$ (Ali and Silvey (1966)). Indeed, Lemma 2 would state that in any κ-feasible interval mechanism, each interval has mass at least l, where l is defined such that $l \cdot \phi(\frac{1}{l}) = \kappa$. Since this lower bound on interval mass is the only feasibility constraint (for the ex post privacy model), our subsequent results also generalize to this class of divergences so long as κ is suitably scaled according to ϕ.

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

where $v(\theta)$ is the virtual valuation of type θ. Therefore, the quality that maximizes the expected profit while maintaining IC and IR is uniquely determined by[16]

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

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

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

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

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

4.3 *Characterization*

With the above preliminary analysis, we can show that a κ-optimal (interval) mechanism exists. Indeed, by Lemma 1, it is sufficient to show there exists a profit-maximizing interval partition. For this, we apply a compactness argument. Consider the following metric on the space of finite partitions: If M consists of cutoffs $\{\theta_1, \ldots, \theta_{n-1}\}$ and M' consists of cutoffs $\{\theta'_1, \ldots, \theta'_{m-1}\}$, then define $d(M, M')$ to be the smallest $\delta \geq 0$ such that for each θ_i there exists θ'_j within δ distance from it, and vice versa. It is straightforward to check this is indeed a metric, and that the resulting qualities q_i and profit $\Pi(\mathbb{M})$ are continuous with respect to this metric.

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

Proposition 1. *There exists a κ-optimal mechanism $\mathbb{M} = \langle M, p, q\rangle$, such that M consists of intervals (each with mass at least $\lambda = e^{-\kappa}$) that partition $[\underline{\theta}, \overline{\theta}]$, and each type $\theta \in \Theta$ reports the interval to which it belongs. The number of intervals in the optimal partition is bounded above by $1/\lambda$ and bounded below by $1/2\lambda$.*

The bounds on the optimal number of intervals provide a direct sense of how the privacy requirement constrains the screening ability. The upper bound of $1/\lambda$ follows

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

from the fact that each feasible interval has mass at least λ. The lower bound of $1/(2\lambda)$ holds because each interval in the *optimal* partition has mass *at most* 2λ; otherwise, it could be split into two subintervals both with mass exceeding λ. By Lemma 2, the resulting partition would be feasible, and Lemma 4 in the Appendix shows that this split would increase the seller's profit.

An immediate corollary of Proposition 1 is the following.

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

The first part is immediate when the privacy constraint is expressed in terms of λ. Indeed, when $\kappa < \log(2)$ then $\lambda > 0.5$, implying that at least 50% of the consumers must choose each menu item. Thus, an optimal mechanism can offer no more than one quality-price pair. As for the second part, $\kappa \in [\log(2), \log(3))$ means that $\lambda \in (1/3, 1/2)$, implying that the optimal mechanism can offer at most two quality-price pairs. Moreover, by Lemma 4 in the Appendix, the seller always benefits from having more information in the form of dividing an interval into two subintervals (so long as the resulting partition is still feasible). In particular, any feasible two-item menu leads to higher profit compared to a singleton menu, leading to the above result.

In general, it is difficult to fully characterize the optimal number n of intervals and their cutoffs for an arbitrary distribution F and cost function c, without imposing additional structure on these primitives. Our next result provides a step toward that characterization by showing that under certain conditions on F and c, the optimal mechanism has the property that the precision of information that the designer collects is *monotonic* in the agent's type: i.e., she either knows more about lower types or knows more about higher types.

To present this result, we introduce the function $G(x) = F^{-1}(x) \cdot (x-1)$ for all $x \in [0,1]$. Observe that $G(F(\theta)) = \theta(F(\theta)-1)$, whose derivative with respect to θ is $v(\theta)f(\theta)$. Hence, by the chain rule, $G'(F(\theta)) = v(\theta)$. We say that $v(\theta)$ is *strictly more concave* (*convex*) than $F(\theta)$ if $G'(x)$ is strictly concave (convex) in x.

PROPOSITION 2. *Suppose that the cost function $c(\cdot)$ has a nonnegative third derivative, and that the virtual valuation $v(\theta)$ is strictly more concave than $F(\theta)$. Then any κ-optimal mechanism consists of intervals that are ordered in increasing mass from left to right; that is, $F(\theta_{i+1}) - F(\theta_i) \geq F(\theta_i) - F(\theta_{i-1})$ for all $i \in \{1, \ldots, n-1\}$. Symmetrically, the intervals in the optimal mechanism would be ordered in decreasing mass if $c''' \leq 0$ and $v(\theta)$ is strictly more convex than $F(\theta)$.*

We mention that a sufficient condition for $v(\theta)$ to be more concave than $F(\theta)$ is that the prior density function $f(\theta)$ is increasing and log-concave; in fact, $F(\theta)$ would be convex and $v(\theta)$ concave.[17] For example, this is satisfied when $F(\theta) = \theta^r - s$ with $\underline{\theta} =$

[17]We have $v''(\theta) = (f'(\theta)/f(\theta))' \cdot (1-F(\theta))/f(\theta) - (f'(\theta)/f(\theta)) \cdot (1 + f'(\theta)(1-F(\theta))/f(\theta)^2) < 0$.

$s^{(1/r)}$ and $\overline{\theta} = (s+1)^{(1/r)}$, where r and s are parameters satisfying $r > 1$ and $s > 1/r$ (the latter ensuring that virtual valuations are positive).

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

4.4 *Uniform-quadratic case*

To give a complete characterization of an optimal κ-constrained mechanism, in this subsection we consider a specific distribution of types and cost function. We will provide detailed results for this particular example to illustrate properties of the optimal mechanisms that we derived in the previous subsection. One specification that admits an elegant analytical characterization is the "uniform-quadratic" case, where F is uniform and c is quadratic. Since in this case $c''' = 0$ and $v(\theta)$ is *as convex as* $F(\theta)$ (in fact both are linear in θ), the results in Proposition 2 suggest that the ordering of intervals does not matter for profit. As we show below, this property enables us to focus on the lengths of the intervals and obtain a full characterization of the optimal partition.

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

In terms of the lower bound λ on the mass of each interval, the condition $\kappa \in [\log(N), \log(N+1))$ translates to $\lambda \in (1/(N+1), 1/N]$. Phrased in this way, the result says that in the uniform-quadratic case it is optimal to divide the type space into as many equal intervals as feasible.

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

4.5 *Welfare analysis*

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

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

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

As the uniform-quadratic case shows, this monotonicity is not necessarily strict everywhere. However, we do know that profit is uniquely maximized at $\kappa = \infty$ (when there is no privacy concern), and uniquely minimized at $\kappa = 0$ (with full privacy).[18]

On the other hand, one may expect that buyer surplus is higher when the privacy constraint is tighter. However, this is not obvious because privacy may hurt allocative efficiency so much that it actually damages the buyer's payoff. As far as we are aware, there is no general finding in the literature regarding how buyer surplus varies with the interval partition.

We are able to answer a simpler question of which privacy level maximizes/minimizes buyer surplus. Specifically, we have the following.

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

For an intuition of this result, we write buyer surplus as $\mathbb{E}[q(\theta) \cdot (1 - F(\theta))/f(\theta)]$, since any additional quality provided to type θ benefits all higher types by the same amount. Note that the quality $q(\theta)$ increases in θ while the inverse hazard rate $(1 - F(\theta))/f(\theta)$ decreases in θ. Thus for any given expected quality $\mathbb{E}[q(\theta)]$, the expected buyer surplus $\mathbb{E}[q(\theta) \cdot (1 - F(\theta))/f(\theta)]$ is maximized when $q(\theta)$ is constant in θ.[19]

Next, recall that for any interval mechanism, when $\theta \in [\theta_{i-1}, \theta_i]$ the optimal quality $q(\theta)$ is given by $c'(q(\theta)) = \mathbb{E}[v(\hat{\theta}) \mid \hat{\theta} \in [\theta_{i-1}, \theta_i]]$, which equates the marginal cost to the expected virtual valuation on the interval. Thus we have $\mathbb{E}[c'(q(\theta))] = \mathbb{E}[v(\theta)]$, meaning that the expected marginal cost does not depend on the partition. Since c' is by assumption a convex function, we conclude that the expected quality $\mathbb{E}[q(\theta)]$ is also maximized when $q(\theta)$ is constant.

Combining both effects described above, we see that the coarsest partition that pools all types maximizes expected buyer surplus (in fact, across *all* interval partitions). A similar argument shows that the finest partition (i.e., $\kappa = \infty$) minimizes buyer surplus.

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

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

[18]This is because profit strictly increases when an interval is divided into two subintervals (and qualities/prices are adjusted accordingly). See Lemma 4 in the Appendix.

[19]This is because if the quality $q(\theta)$ is not constant across types, then decreasing it for higher types and increasing it for lower types, in a way that preserves the expected quality $\mathbb{E}[q(\theta)]$, increases the expected buyer surplus $\mathbb{E}[q(\theta) \cdot (1 - F(\theta))/f(\theta)]$.

Our proof shows that with increasing density and quadratic cost, total welfare necessarily increases when the seller obtains less information in the form of combining two intervals into one. Thus, in this case the coarsest partition maximizes welfare while the finest partition minimizes welfare. One difficulty in obtaining a more general result lies in the fact that total welfare takes into account both buyer and seller surplus, which depend on the value distribution, the cost function, as well as their interaction.[20] Separate assumptions about each of these primitives typically do not lead to clear predictions.

5. Ex ante Bayesian privacy

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

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

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

where $\mathbb{E}_m$ is evaluated according to the probability that each message $m \in M$ is sent in an equilibrium of $\mathbb{M}$.[21]

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

5.1 *Existence*

To begin the analysis, we can use essentially the same argument to show that Lemma 1 also holds under the ex ante criterion. Thus, it is without loss to consider interval mechanisms. Using equation (2), we can compute the ex ante loss of privacy entailed by a (finite) interval mechanism to be

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

[20] Specifically, the marginal cost function determines optimal qualities, and the cost function itself maps these qualities into expected production cost (averaged across types using the prior distribution). This is a more complicated situation than our previous result about buyer surplus, which is fully determined by optimal qualities and the prior distribution. In other words, the effect of $c(\cdot)$ on buyer surplus can be analyzed separately from $F(\cdot)$, but such separation is lost in the study of total welfare.

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

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

This discussion suggests that under the ex ante privacy constraint, the profit maximization problem reduces to finding an interval partition M subject to $H(g_M) \leq \kappa$ that maximizes the profit in equation (3) with qualities given by (4). Note that the privacy constraint $H(g_M) \leq \kappa$ can be interpreted as requiring that the weighted geometric mean of the masses of the intervals must exceed $e^{-\kappa}$ (where the weight applied to each interval's mass is just this mass itself).

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

PROPOSITION 6. *There exists an ex ante κ-optimal mechanism $\mathbb{M} = \langle M, p, q\rangle$, such that M consists of finitely many intervals that partition $[\underline{\theta}, \overline{\theta}]$, and each type $\theta \in \Theta$ reports the interval to which it belongs.*[22]

We provide a sketch of the proof here, leaving further details to the Appendix. Consider a sequence of ex ante κ-feasible interval mechanisms $\mathbb{M}_j = \langle M_j, p_j, q_j\rangle$ such that $\Pi(\mathbb{M}_j)$ converges to the supremum profit Π^* across feasible mechanisms. We will replace each mechanism $\mathbb{M}_j$ by another feasible interval mechanism $\tilde{\mathbb{M}}_j = \langle \tilde{M}_j, \tilde{p}_j, \tilde{q}_j\rangle$, such that the new message set $\tilde{M}_j$ consists of at most N intervals, where N *is a constant that depends only on F and κ*. This upper bound N restores compactness and allows us to find a subsequence of the partitions $\{\tilde{M}_j\}$ that converges to some limit partition $\tilde{M}_\infty$, under the metric defined in Section 4.3. By continuity, $\tilde{M}_\infty$ is also a feasible mechanism, and it achieves the limit profit along the convergent subsequence. Therefore, if we could carry out the replacement in such a way that $\Pi(\tilde{\mathbb{M}}_j) \geq \Pi(\mathbb{M}_j)$, then $\Pi(\tilde{M}_\infty) \geq \limsup_j \Pi(\tilde{M}_j) = \Pi^*$ and $\tilde{M}_\infty$ would be ex ante κ-optimal.

It remains to find the appropriate replacements $\tilde{\mathbb{M}}_j$. As discussed, starting from any mechanism $\mathbb{M}_j$, merging two adjacent intervals in M_j into a single interval (and adjusting the qualities/prices accordingly) strictly decreases the profit. However, by doing so the seller is able to save on the ex ante privacy measure, which enables him to divide any other interval in M_j into two subintervals, increasing the profit. The key argument then is to *compare the profit gain in the latter step to the profit loss in the former*. We show that whenever two adjacent intervals are both of mass smaller than some constant ϵ, they can be combined to create enough slackness in the privacy constraint; and if the slackness is used to break another (big) interval into two, the seller achieves a net profit gain.

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

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 $N := 2/\epsilon + 1$ intervals, completing the proof.[23]

5.2 *Properties of optimal ex ante constrained mechanisms*

In contrast to the ex post problem, the ex ante privacy constraint always binds at the optimal mechanism. Moreover, unlike Lemma 2, the number of intervals now admits a *lower bound* of $1/\lambda$ (where $\lambda \equiv e^{-\kappa}$).

PROPOSITION 7. *The privacy constraint is exhausted in any ex ante κ-optimal mechanism $\mathbb{M}$; that is, $I^{ea}(\mathbb{M}) = \kappa$. Moreover, any optimal mechanism involves at least $\frac{1}{\lambda}$ intervals.*

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

(COVER AND THOMAS (2006, THEOREM 2.6.4)). *If a discrete random variable X takes n values, then its Shannon entropy satisfies $H(X) \leq \log(n)$, with equality if and only if X has a uniform distribution.*

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

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

We prove the result via a lemma stating that there can be at most one interval with arbitrarily small mass. Thus, even though the average privacy constraint allows the seller

[23]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 she 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.

to learn almost perfectly about sets of small types, the optimal solution involves at most one such set. In the above proof sketch for Proposition 6, we have already discussed why having two adjacent "small" intervals is suboptimal. The next lemma additionally rules out the presence of two "small" intervals that are nonadjacent.

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

Finally, Proposition 2, which gives sufficient conditions for the optimality of intervals that are increasing/decreasing in mass, continues to hold under the ex ante privacy measure. This is because the proof of Proposition 2 is based on evaluating whether profit increases or decreases when two adjacent intervals are "switched." Since switching does not affect the amount of privacy loss under both the ex post and ex ante measures, this argument extends without change.

The welfare results Proposition 4 and Proposition 5 also generalize to the current setting, since their proofs only involve welfare comparison when combining two adjacent intervals. Whether we adopt the ex post or ex ante privacy measure, the coarsest interval partition corresponds to $\kappa = 0$, and the finest partition to $\kappa = \infty$. These levels of privacy requirement thus maximize/minimize buyer surplus and/or total welfare.

5.3 *Uniform-quadratic case revisited*

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

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

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

$$\overline{\theta} - \underline{\theta} = \ell_s + (N-1)\ell_b$$

$$\kappa = -\frac{\ell_s}{(\overline{\theta} - \underline{\theta})} \log \frac{\ell_s}{(\overline{\theta} - \underline{\theta})} - (N-1) \cdot \frac{\ell_b}{(\overline{\theta} - \underline{\theta})} \log \frac{\ell_b}{(\overline{\theta} - \underline{\theta})}$$

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

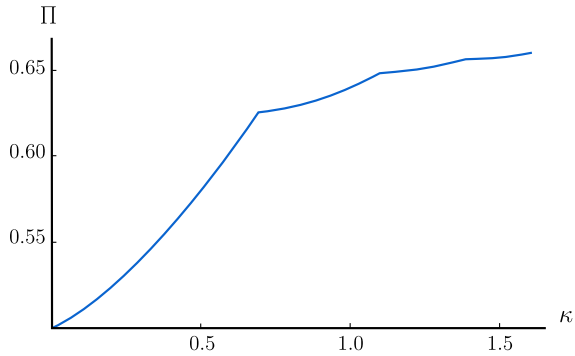

FIGURE 1. The privacy-profit frontier in the uniform-quadratic case.

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

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

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

We mention that the qualitative features of the privacy-profit frontier are robust to small changes in the prior distribution (away from uniform). This is because the set of ex ante κ-optimal mechanisms, when viewed as a correspondence from the distribution F to the space of interval partitions, is upper hemicontinuous.[24]

[24]This follows from the maximum theorem, since the seller solves a constrained optimization problem in which profit is continuous in the interval partition (with respect to the metric defined in Section 4.3), and the constraint $H(g_M) \leq \kappa$ is both upper and lower hemicontinuous.

6. Negative virtual values

So far we have focused on the case where all types have nonnegative virtual values. While our results are more easily stated under this simplifying assumption, they generalize to the case where $v(\underline{\theta})$ is negative. In this section, we discuss this extension and describe how some of the results need to be slightly modified. Throughout this section, we work with the ex post privacy constraint.

First of all, the optimality of interval mechanisms reported in Lemma 1 holds without change. This is because our proof of this result primarily uses the convexity of KL-divergences and the single-crossing property of buyer preferences, which continue to hold even with negative virtual values.

It is also still true that every interval in the partition has mass at least $\lambda = e^{-\kappa}$, although the lowest interval may now have a negative average virtual value and lead to an optimal quality of zero. Note that if there are more than one interval with negative average virtual value, then these intervals can be combined without affecting the privacy constraint or profit. So below we assume there is at most one such interval. A corollary is that apart from the lowest *two* intervals, all higher intervals consist of types with positive virtual values.[25] Because of this, Proposition 2 (regarding the intervals ordered in increasing/decreasing mass) holds except possibly for the lowest two intervals.

The next result generalizes Proposition 3 and completely characterizes the optimal interval partition in the uniform-quadratic case.

Proposition 10. *Suppose $F \sim U[\underline{\theta}, \overline{\theta}]$ with $\overline{\theta} > 2\underline{\theta}$ (so that $v(\underline{\theta}) < 0$), and $c(q) = q^2/2$. Then, given any positive integer N and any $\lambda = e^{-\kappa} \in (1/(N+1), 1/N]$, the κ-optimal mechanism takes one of three possible forms described below:*

1. *if $\underline{\theta} \geq ((N-1)/(2N-1)) \cdot \overline{\theta}$, then dividing the type space into N equal intervals is optimal;*
2. *if $\underline{\theta} < ((N-1)/(2N-1)) \cdot \overline{\theta}$ and $\overline{\theta}/(2N-1) > \lambda(\overline{\theta} - \underline{\theta})$, then it is optimal to have N intervals with cutoffs $\theta_j = ((j+N-1)/(2N-1)) \cdot \overline{\theta}$ for $1 \leq j \leq N-1$;*
3. *otherwise, there exists a unique positive integer $m \leq N$ such that*

$$\frac{\overline{\theta}}{2m-3} > \lambda(\overline{\theta} - \underline{\theta}) \geq \frac{\overline{\theta}}{2m-1},$$

in which case it is optimal to have m intervals. Except for the lowest interval, the remaining $m-1$ intervals all have mass exactly λ.

To interpret this solution, note that $\lambda \leq 1/N$ means it is feasible to divide the type space into N equal intervals, but $\lambda > 1/(N+1)$ means there cannot be more than N intervals. As we showed in Proposition 3, N equal intervals is optimal in an environment where virtual values are all positive. More generally, the condition $\underline{\theta} \geq ((N-1)/(2N-$

[25] Formally, the claim is that it is without loss to consider optimal partitions $\underline{\theta} = \theta_0 < \theta_1 < \cdots < \theta_n = \overline{\theta}$ with $v(\theta_2) > 0$. The uniform-quadratic case studied below shows that $v(\theta_1)$ need not be positive.

$1)) \cdot \overline{\theta}$ corresponds to the lowest interval having positive *average* virtual value under the equal partition.[26] Thus, the first part of the above proposition says that if the "first best" solution of N equal intervals does not require the seller to withhold sale from the lowest interval, then it is in fact the optimal partition.

However, when this candidate solution involves no sale to the lowest interval, the optimal solution changes. Intuitively, in this case the seller would like to increase the length of the lowest interval to charge higher prices and extract more profits from the other intervals. The second part of Proposition 10 provides a "second best" solution, which involves a lowest interval with negative average virtual value, and $N-1$ other intervals with equal length. Note that as long as the mass lower bound of λ is not binding (reflected in the condition $\overline{\theta}/(2N-1) > \lambda(\overline{\theta}-\underline{\theta})$), the optimal *length* for the $N-1$ intervals is $\overline{\theta}/(2N-1)$ and does not depend on λ in this case. This length turns out to maximize the seller's profit from $N-1$ equal intervals when the privacy constraint is ignored.

Lastly, the parameters may be such that the second best no longer satisfies the privacy constraint. When this happens, the optimal partition involves a lowest interval with negative average virtual value, and $m-1$ other intervals with mass exactly λ. The number m as determined by the condition $\overline{\theta}/(2m-3) > \lambda(\overline{\theta}-\underline{\theta}) \geq \overline{\theta}/(2m-1)$ is the biggest number such that fitting in $m-1$ intervals with mass λ still makes the lowest one of them have positive average virtual value. This, as we discussed, is a key necessary property for the optimal partition.

7. Concluding remarks

This paper proposed a Bayesian approach to incorporating privacy constraints into mechanism design. The underlying idea is that the designer *already* has some prior information about the participants, and the loss of privacy induced by a mechanism should be measured as the *difference* between this prior information and the updated information that can be inferred from the agents' interaction with the mechanism. This entails an additional constraint—on top of the standard incentive-compatibility and individual-rationality constraints—that needs to be satisfied by a mechanism: The difference between the prior and posterior information must be below some threshold.

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

[26] Since $v(\theta) = 2\theta - \overline{\theta}$, the average virtual value on $[\underline{\theta}, \theta_1]$ is positive if and only if $\theta_1 \geq \overline{\theta} - \underline{\theta}$. This holds under the equal partition if and only if $\underline{\theta} + (1/N)(\overline{\theta}-\underline{\theta}) \geq \overline{\theta} - \underline{\theta}$.

Our approach opens the door to many interesting questions about mechanism design and privacy. In particular, how does our analysis extend to an environment with multiple agents? In Appendix C below, we derive the optimal symmetric mechanism under ex post privacy for uniformly distributed types. This leaves open the question of designing optimal privacy-preserving auctions when asymmetric treatment of bidders is allowed. We hope that future research will provide answers to these and related questions.

Appendix A: Proofs for the main model

A.1 *Preliminaries*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

When an interval $[a, c]$ is divided into $[a, b]$ and $[b, c]$, the profit gain is therefore

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

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

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

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

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

$$h(v_1) = h(v) + (v_1 - v)h'(v) + \frac{(v_1 - v)^2}{2} h''(\xi)$$

$$h(v_2) = h(v) + (v_2 - v)h'(v) + \frac{(v_2 - v)^2}{2} h''(\zeta)$$

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

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

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

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

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

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

A.2 *Proof of Lemma 1*

We will show that any mechanism $\mathbb{M}=\langle M, p, q\rangle$ that satisfies $I(\mathbb{M}) \leq \kappa$, 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 $D_{\mathrm{KL}}(F(\cdot|m)\|F)$ in its first argument, the entailed loss of privacy $I(\mathbb{M})$ is decreased. *This is true under both the ex post and ex ante measures of privacy.*[27]

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

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

Since the agent's preference satisfies increasing differences in (θ, q), the set $\mu(m)$ is either an interval or a singleton.[28] However, since κ is finite, there can be only a zero-measure subset of messages $m \in M$ for which $\mu(m)$ is a singleton.[29] 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.

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

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

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

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

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

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

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., quality) 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.

A.3 *Proof of Proposition 1*

Given Lemma 1 and Lemma 2, it only remains to prove the bounds on the optimal number of intervals. This follows from the argument given in the main text, after Proposition 1.

A.4 *Proof of Proposition 2*

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

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

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

Observe that $(\tilde{w} - u) \cdot [F(\theta_t) - F(\theta_{t-1})] = (w - \tilde{u}) \cdot [F(\theta_{t+1}) - F(\theta_t)]$.[30] So to show Δ is positive we just need to show

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

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

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

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

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

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

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

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

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

This is equivalent to

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

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

A.5 *Proof of Proposition 3*

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

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

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

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

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

The profit as given by equation (3) is

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

$$-2\sum_{i=1}^{n}(\theta_i^2-\theta_{i-1}^2)\cdot\overline{\theta}+\sum_{i=1}^{n}(\theta_i-\theta_{i-1})\cdot\overline{\theta}^2\Bigg).$$

The three terms in the parentheses above can be simplified as follows: $\sum_{i=1}^{n}(\theta_i-\theta_{i-1})=(\overline{\theta}-\underline{\theta})$, $\sum_{i=1}^{n}(\theta_i^2-\theta_{i-1}^2)=(\overline{\theta}^2-\underline{\theta}^2)$, and $\sum_{i=1}^{n}(\theta_i-\theta_{i-1})(\theta_i+\theta_{i-1})^2=\frac{4}{3}(\overline{\theta}^3-\underline{\theta}^3)-\frac{1}{3}\sum_{i=1}^{n}(\theta_i-\theta_{i-1})^3$.[31] Plugging the three expressions back, we obtain the lemma. □

Back to the proposition, when $\kappa\in[\log(N),\log(N+1))$, we have $e^\kappa<N+1$. So by Lemma 2, any feasible interval mechanism has at most N intervals. Now by Lemma 5, among interval partitions with at most N intervals, maximizing profit is equivalent to minimizing $\sum_{i=1}^{N}x_i^3$, where $x_i=\theta_i-\theta_{i-1}$ is the length of the interval m_i (which can be zero if less than N intervals are used). Subject to $x_i\geq 0$ and the total length $\sum_{i=1}^{N}x_i=\overline{\theta}-\underline{\theta}$, the cubic sum $\sum_{i=1}^{N}x_i^3$ is clearly minimized when the intervals have equal lengths. Hence, the equal partition maximizes profit among all partitions with at most N intervals, even ignoring the feasibility constraint. Since it is κ-feasible, it must then be the ex post κ-optimal mechanism.

A.6 *Proof of Proposition 4*

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

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

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

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

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

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

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

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

$$\frac{q-q_j}{q_{j+1}-q}=\frac{\phi(v)-\phi(v_j)}{\phi(v_{j+1})-\phi(v)}$$

$$\geq\frac{v-v_j}{v_{j+1}-v}=\frac{F(\theta_{j+1})-F(\theta_j)}{F(\theta_j)-F(\theta_{j-1})}\geq\frac{\int_{\theta_j}^{\theta_{j+1}}\big(1-F(\theta)\big)\,d\theta}{\int_{\theta_{j-1}}^{\theta_j}\big(1-F(\theta)\big)\,d\theta},\tag{10}$$

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

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

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

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

A.7 *Proof of Proposition 5*

With quadratic costs, total welfare contributed by a buyer of type θ is $\theta q(\theta)-(q(\theta))^2/2$. Thus ex ante total welfare is $\int q(\theta)\cdot(\theta-\frac{q(\theta)}{2})f(\theta)\,d\theta$. 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, ex ante total welfare can be equivalently written as

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

Compared with the above equation (9) for buyer surplus, the difference here is that the function $(\theta-v(\theta)/2)f(\theta)$ takes the place of $1-F(\theta)$.

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

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

A.8 *Proof of Proposition 10*

Throughout, we assume $\lambda = e^{\kappa} \in (1/(N+1), 1/N]$. The seller's problem is to find an interval partition such that each interval has mass at least λ, and profit is maximal subject to buyer incentive constraints. Note that on any interval $[\theta_{i-1}, \theta_i]$, the optimal quality is either the average virtual value $\theta_{i-1} + \theta_i - \overline{\theta}$ when this is positive, or zero when the average virtual value is negative. Correspondingly, the profit contribution of any interval is either $(\theta_{i-1} + \theta_i - \overline{\theta})^2/2$, or zero. It is no greater than $(\theta_{i-1} + \theta_i - \overline{\theta})^2/2$ in either case.

Thus, for any interval partition with cutoffs $\underline{\theta} = \theta_0 < \theta_1, \ldots, \theta_{n-1} < \theta_n = \overline{\theta}$, total profit is *bounded from above* by what is calculated in Lemma 5, with equality if and only if the lowest interval has nonnegative average virtual value. Now recall that each interval has mass $\geq \lambda > 1/(N+1)$, so the number of intervals satisfies $n \leq N$. It follows that the profit given by Lemma 5 is maximized when $n = N$ and all intervals have equal length. This upper bound is achieved in our current setting if and only if the lowest interval under the equal partition has nonnegative average virtual value. This requires $\underline{\theta} + (\underline{\theta} + (\overline{\theta} - \underline{\theta})/N) - \overline{\theta} \geq 0$, which is precisely the condition stated in the first part of the proposition.

Below we consider the case where this first best is not achievable, i.e., $\underline{\theta} < ((N-1)/(2N-1)) \cdot \overline{\theta}$. Suppose the optimal partition has exactly $n-1$ intervals with positive average virtual values, starting from some type $\hat{\theta}$. Then the optimal mechanism simply excludes the types lower than $\hat{\theta}$ and reduces to an optimal mechanism for the uniform distribution on $[\hat{\theta}, \overline{\theta}]$. It follows from Lemma 5 that *these* $n-1$ intervals above $\hat{\theta}$ must have equal length (in order to minimize the cubic sum of their lengths). Thus, $n \leq N$, for otherwise the lowest of these $n-1$ equal intervals would have negative average virtual value, by the assumption that $\underline{\theta} < ((N-1)/(2N-1)) \cdot \overline{\theta}$.

We can now formulate the seller's problem as choosing a cutoff type $\hat{\theta}$ and a positive integer $n \leq N$ in order to minimize the profit from an equal partition of $[\hat{\theta}, \overline{\theta}]$ into $n-1$ equal intervals, subject to the following two constraints:

(i) $\hat{\theta} > ((n-2)/(2n-3)) \cdot \overline{\theta}$, so the lowest of these $n-1$ intervals has positive average virtual value;

(ii) $\overline{\theta} - \hat{\theta} \geq (n-1)\lambda(\overline{\theta} - \underline{\theta})$, so these intervals have mass at least λ.

The optimal $\hat{\theta}$ will turn out to also satisfy $\hat{\theta} < \overline{\theta} - \underline{\theta}$ and $\hat{\theta} - \underline{\theta} \geq \lambda(\overline{\theta} - \underline{\theta})$, which means there is another interval below $\hat{\theta}$ in the overall partition of $[\underline{\theta}, \overline{\theta}]$. This nth interval has negative average virtual value, and has mass at least λ. So the solution to our relaxed problem will be feasible.

To characterize the optimal $\hat{\theta}$ and n, we rely on the following lemma.

LEMMA 6. *For each positive integer $n > 1$, denote by $\Pi_{n-1}(\hat{\theta})$ the optimal profit from excluding types below $\hat{\theta}$ and dividing types above $\hat{\theta}$ into $n-1$ equal intervals. Then $\Pi_{n-1}(\hat{\theta})$ as a function of $\hat{\theta}$ is strictly increasing when $\hat{\theta} \in [((n-2)/(2n-3)) \cdot \overline{\theta}, (n/2n-1) \cdot \overline{\theta}]$, and is strictly decreasing when $\hat{\theta} \geq (n/(2n-1)) \cdot \overline{\theta}$.*

Indeed, using Lemma 5 we can compute that

$$\Pi_{n-1}(\hat{\theta}) = \frac{\overline{\theta}-\hat{\theta}}{\overline{\theta}-\underline{\theta}} \cdot \left[\left(\frac{1}{6} - \frac{1}{6(n-1)^2}\right)(\overline{\theta}-\hat{\theta})^2 + \frac{1}{2}\hat{\theta}^2\right],$$

where the term $(\overline{\theta}-\hat{\theta})/(\overline{\theta}-\underline{\theta})$ represents the fact that each type in $[\hat{\theta}, \overline{\theta}]$ occurs with a scaled density compared to the uniform distribution on this interval. With a little algebra, the derivative of the above function with respect to $\hat{\theta}$ simplifies to

$$\frac{\big((2n-3)\hat{\theta} - (n-2)\overline{\theta}\big) \cdot \big(n\overline{\theta} - (2n-1)\hat{\theta}\big)}{2(n-1)^2(\overline{\theta}-\underline{\theta})}.$$

This yields Lemma 6.

We will apply this lemma to deduce the second and third parts of Proposition 10. Suppose $\overline{\theta}/(2N-1) > \lambda(\overline{\theta}-\underline{\theta})$, then the seller can choose $n = N$ and $\hat{\theta} = (N/(2N-1)) \cdot \overline{\theta}$, while satisfying the privacy constraint (ii) above. Note that $\hat{\theta} < \overline{\theta} - \underline{\theta}$ by our previous assumption about $\underline{\theta}$, and $\hat{\theta} - \underline{\theta} > (1/N)(\overline{\theta}-\underline{\theta}) \geq \lambda(\overline{\theta}-\underline{\theta})$. So as we promised, the excluded types below $\hat{\theta}$ have negative average virtual value, and do not violate the privacy constraint.

This feasible solution is optimal when $n = N$, by Lemma 6. Thus, it only remains to show that having a smaller n is not profitable. For this, we note the following simple fact:

$$\Pi_{n-1}\left(\frac{n-2}{2n-3}\overline{\theta}\right) = \Pi_{n-2}\left(\frac{n-1}{2n-3}\overline{\theta}\right). \tag{12}$$

The reason is that the $n-1$ intervals associated with the LHS simply adds the interval $[((n-2)/(2n-3)) \cdot \overline{\theta}, ((n-1)/(2n-3)) \cdot \overline{\theta}]$ to the $n-2$ intervals on the RHS. But since this extra interval has average virtual value equal to zero, its optimal quality is zero whether or not it is included. So excluding it from the LHS has no effect on total profit.

From equation (12) and Lemma 6, we have

$$\Pi_{n-1}\left(\frac{n}{2n-1}\overline{\theta}\right) > \Pi_{n-1}\left(\frac{n-2}{2n-3}\overline{\theta}\right) = \Pi_{n-2}\left(\frac{n-1}{2n-3}\overline{\theta}\right).$$

This means the optimal profit with $n-1$ intervals is higher than the optimal profit with $n-2$ intervals, so on and so forth. Hence, the choices $n = N$ and $\hat{\theta} = (N/(2N-1)) \cdot \overline{\theta}$ are optimal among any $n \leq N$ and $\hat{\theta} \geq ((n-2)/(2n-3)) \cdot \overline{\theta}$. This proves the second part of the proposition.

Finally, we suppose $\underline{\theta} < ((N-1)/(2N-1)) \cdot \overline{\theta}$ and for some $m \leq N$,

$$\frac{\overline{\theta}}{2m-3} > \lambda(\overline{\theta}-\underline{\theta}) \geq \frac{\overline{\theta}}{2m-1}.$$

From the constraints (i) and (ii) above on $\hat{\theta}$, we deduce $\lambda(\overline{\theta}-\underline{\theta}) < \overline{\theta}/(2n-3)$. Thus, the fact that $\lambda(\overline{\theta}-\underline{\theta}) \geq \overline{\theta}/(2m-1)$ implies $n \leq m$. Consider the choices $n = m$ and $\hat{\theta} = \overline{\theta} - (m-1)\lambda(\overline{\theta}-\underline{\theta})$. The privacy constraint (ii) is exactly satisfied, and the virtual value constraint (i) is satisfied because $\lambda(\overline{\theta}-\underline{\theta}) < \overline{\theta}/(2m-3)$. Moreover, from $\lambda(\overline{\theta}-\underline{\theta}) \geq$

$\overline{\theta}/(2m-1)$ we deduce $\underline{\theta} \leq \overline{\theta} - \overline{\theta}/(\lambda(2m-1)) \leq \overline{\theta} - N\overline{\theta}/(2m-1) < ((m-1)/(2m-1)) \cdot \overline{\theta}$, with the last inequality being strict because we already know $\underline{\theta} < ((N-1)/(2N-1)) \cdot \overline{\theta}$ when $m = N$. Together with $\hat{\theta} = \overline{\theta} - (m-1)\lambda(\overline{\theta} - \underline{\theta}) \leq (m/(2m-1)) \cdot \overline{\theta}$, we deduce that the types below $\hat{\theta}$ have negative average virtual value. The mass of these types is $1 - (m-1)\lambda \geq 1 - (N-1)\lambda \geq \lambda$, so this solution is feasible.

Now observe that given the privacy constraint (ii), $\hat{\theta} = \overline{\theta} - (m-1)\lambda(\overline{\theta} - \underline{\theta})$ is the maximal value $\hat{\theta}$ can take subject to $n = m$. As $\hat{\theta} \leq (m/(2m-1)) \cdot \overline{\theta}$, we know from Lemma 6 that this choice of $\hat{\theta}$ is optimal for $n = m$. From that lemma, we also have

$$\Pi_{m-1}(\hat{\theta}) > \Pi_{m-1}\left(\frac{m-2}{2m-3}\overline{\theta}\right) = \Pi_{m-2}\left(\frac{m-1}{2m-3}\overline{\theta}\right).$$

Thus, the choice of $m-1$ intervals starting from $\hat{\theta}$ is better than the optimum with $m-2$ intervals, which is in turn better than any optimum with even fewer intervals. This proves the third/last part of Proposition 10.

Appendix B: Proofs for the ex ante measure

B.1 *Proof of Proposition 6*

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

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

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

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

Fixing this choice of l, we define ϵ to be a small positive constant as given by Lemma 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 (ex ante) privacy measure is unchanged and profit is strictly increased. The achieved profit gain is sufficient to cover the loss from additionally combining all the (countably many) intervals $m_t, m_{t+1}, \ldots$, so long as we choose t to be sufficiently large. As this last step also relaxes the privacy constraint, we obtain a replacement mechanism in which p is no longer an accumulation point of intervals on its left. Doing the same exercise for intervals on the right of p yields a mechanism in which p is not an accumulation point.

In fact, we can achieve this replacement with some extra properties. Note that whenever an accumulation point p exists, the "big" interval must have mass strictly greater than $l = e^{-\kappa}$; otherwise, the privacy constraint requires every interval in M to have mass exactly l, a contradiction. Thus by choosing m_s and m_{s+1} to have sufficiently small mass, we can ensure that when they are merged and the "big" interval is divided into two subintervals, *the bigger subinterval still has mass* $> l$. In other words, we can perform the replacement in such a way that the *same big interval* is sequentially divided (each time creating a small subinterval on the left and a big one on the right). The benefit is that as we get rid of the accumulation points in $\mathbb{M}$ one by one (which may be countably many), we obtain a sequence of replacement mechanisms that become finer in the original "big" interval in $\mathbb{M}$ and more coarse everywhere else. This sequence converges, and the limit mechanism has at most one accumulation point in the "big" interval.[32] 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 $N := 2/\epsilon + 4$ intervals. Starting from $\hat{\mathbb{M}}$, if there are *two pairs* of adjacent intervals (i.e., 4 distinct ones) all with mass $< \epsilon$, then we combine both pairs at the same time and used the privacy measure saved from one of the mergers to divide the "big" interval into two subintervals. The privacy constraint is relaxed, and by Lemma 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 $N = 2/\epsilon + 4$.

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 ex ante privacy measure to divide the big interval into two subintervals, the profit gain in the latter step is at least twice as big as the profit loss in the former step.*

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

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

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

By assumption, there exists another interval of mass $L \geq l$. We use the saved privacy measure to break this interval into two: That is, we look for a subinterval of mass $\delta \in (0, L/2)$ 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[33]

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

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

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

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 $1/(2\sqrt{x+y})$ is at least 4. So by binomial expansion, the RHS above has size at least

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

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

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

B.2 *Proof of Proposition 7*

The argument is already sketched in the main text, after Proposition 7. In particular, Lemma 4 ensures that the seller strictly benefits from dividing any interval into two subintervals.

[33] 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 $\delta(1+\log(\zeta/\delta)) = \alpha$. Since $\zeta \geq L/2 \geq \delta$, this implies

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

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

B.3 *Proof of Proposition 8*

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

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

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

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

B.4 *Proof of Lemma 3*

In the proof of Proposition 6, we showed that in any ex ante κ-feasible mechanism there is a "big" interval of mass at least $e^{-\kappa} \geq e^{-k}$. So by Lemma 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 ex ante privacy measure, and it remains to estimate the profit change.

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

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

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

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

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

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

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

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

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

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, $h(\tilde{u}) = h(w) + h'(w) \cdot ([v(\theta_t) - v(\theta_{t+1})] \cdot x/(F(\theta_{t+1}) - F(\theta_t))) + O(x^2)$. It follows that

$$\begin{aligned} &h(\tilde{u}) \cdot \big[F(\tilde{\theta}_t) - F(\tilde{\theta}_{t-1})\big] - h(w) \cdot \big[F(\theta_{t+1}) - F(\theta_t)\big] \\ &\quad = h'(w) \cdot \big[v(\theta_t) - v(\theta_{t+1})\big] \cdot x + O\big(x^2\big) \end{aligned}$$

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

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

Summing across $t \in \{i, \dots, j-1\}$, 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[34]

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

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

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

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 $y \cdot \Delta_{\mathrm{LR}} + x \cdot \Delta_{\mathrm{RL}}$, then the second term is cancelled out. This yields

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

Therefore, Δ_{LR} and Δ_{RL} cannot both be very negative. To be concrete, we may without loss assume $\Delta_{\mathrm{LR}} \geq -O(xy)$. 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 nonadjacent.

B.5 *Proof of Proposition 9*

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

It remains to prove the above claim that characterizes the lengths of the optimal intervals. For an interval partition, let $x_i = (\theta_i - \theta_{i-1})/(\overline{\theta} - \underline{\theta})$ denote the probability mass

[34]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 $O(x^2)$.

of the ith interval. In what follows, we will work with the probability masses $\{x_i\}$ instead of the cutoffs $\{\theta_i\}$. By Lemma 5 and equation (6), seller's profit maximization problem under the ex ante privacy constraint can be rewritten as the following constrained minimization problem: $\min \sum_{i=1}^n x_i^3$ subject to $x_i \geq 0$, $\sum_{i=1}^n x_i = 1$, and $\sum_{i=1}^n x_i \log x_i \leq -\kappa$. For a fixed n, the Lagrangian is given by

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

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

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

If $\beta \leq 0$, then the function $3x^2 - \beta \log x$ is monotonically increasing. Thus, every x_i is the same, which completes the proof; otherwise, assume $\beta > 0$. In this case, the derivative of the function $3x^2 - \beta \log x$ is $6x - \beta/x$, which is monotonically increasing and crosses 0 at $\hat{x} = \sqrt{\beta/6}$. Thus, the function $3x^2 - \beta \log x$ decreases on $[0, \hat{x}]$ and increases on $[\hat{x}, \infty)$. Equation (16) yields that x_i can take at most two values $\underline{x}$ and $\overline{x}$, with $\underline{x} < \hat{x} < \overline{x}$.

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

The second derivative of the Lagrangian $\mathcal{L}(\alpha, \beta, x)$ with respect to (the vector) x is the diagonal matrix $H = \operatorname{diag}(6x_1 - \beta/x_1, \ldots, 6x_n - \beta/x_n)$. 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 minima; see, e.g., Simon and Blume (1994, page 468). Hence, the proof is complete.

Appendix C: Extension to multiagent mechanisms

Extending our analysis to mechanisms with more than one agent presents some challenges. In particular, the notion of privacy loss needs to be extended to accommodate the possibility that different participants are exposed to different losses of privacy.[35] One approach is to require that the maximal loss of privacy for any agent is at most κ; an alternative is to measure the average loss of privacy across all agents. As in our single-agent model, the privacy notion also has to address the fact that loss of privacy may differ across types of the same agent.

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

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

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

PROPOSITION 11. *Suppose there are $m \geq 5$ buyers with uniformly distributed values, and no production costs. Then for any $\kappa \in [\log(n), \log(n+1))$, the optimal ex post privacy-constrained symmetric auction partitions the type space into n intervals. Each of the upper $n-1$ intervals has equal mass of $e^{-\kappa}$ (which is privacy-binding), and the lowest interval has the remaining mass of $1-(n-1)e^{-\kappa}$ (which is weakly greater).*

Our proof below also shows that with 2 buyers, the optimal auction partitions the set of types into n equally long intervals, which coincides with the solution for a single buyer under quadratic costs. The solutions to 3 or 4 buyers are more complex.

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

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

where each element in the sum corresponds to the event that $m-k-1$ opponents report an interval lower than $[\theta_{i-1}, \theta_i]$ and k opponents report the interval $[\theta_{i-1}, \theta_i]$, in which case the buyer wins the object with probability $1/(k+1)$. Simplifying the right-hand side yields[36]

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

[36]To see this note that $(1/(k+1))\binom{m-1}{k} = (1/m)\binom{m}{k+1}$, and thus

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

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

$$= \frac{1}{m} \frac{1}{\theta_i-\theta_{i-1}} \sum_{l=1}^{m} \binom{m}{l} (\theta_{i-1}-\theta_0)^{m-l}(\theta_i-\theta_{i-1})^l$$

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

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

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

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

The total profit from all buyers is therefore

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

We denote $z_i \equiv (\theta_i - \theta_0)/(\theta_n - \theta_0)$. Then the seller seeks to maximize the expression

$$\hat{\Pi} \equiv \sum_{i=1}^{n} z_i \cdot z_{i-1} \big(z_i^{m-1} - z_{i-1}^{m-1} \big), \tag{17}$$

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

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

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

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

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

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

That is, we need to show

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

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

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

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

Next, we argue that $z_1 < 2(z_2-z_1)$, meaning that the longest interval is less than twice the length of the second longest. Indeed, if this were not the case, we could break the longest interval into two equal subintervals and still satisfy the ex post privacy constraint. As can be easily seen from equation (17), this modification strictly increases the profit, contradicting optimality.

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

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

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

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

We mention that a similar (but more involved) analysis yields the optimal partitions for the cases $m=3$ and $m=4$ as well. As for $m=2$, the profit as given by equation (17) can be simplified as

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

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

References

Acquisti, Alessandro and Jens Grossklags (2005), "Privacy and rationality in individual decision making." *IEEE Security and Privacy*, 3, 26–33. [1559]

Acquisti, Alessandro, Curtis R. Taylor, and Liad Wagman (2016), "The economics of privacy." *Journal of Economic Literature*, 54, 442–492. [1563]

Agrawal, Dakshi and Charu Aggarwal (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. [1564]

Ali, S. M. and S. D. Silvey (1966), "A general class of coefficients of divergence of one distribution from another." *Journal of the Royal Statistical Society*, 28, 131–142. [1569]

Alonso, Ricardo and Niko Matouschek (2008), "Optimal delegation." *Review of Economic Studies*, 75, 259–293. [1562]

Babaioff, Moshe, Liad Blumrosen, and Michael Schapira (2013), "The communication burden of payment determination." *Games and Economic Behavior*, 77, 153–167. [1562]

Barth, Susanne and Menno D. T. de Jong (2017), "The privacy paradox–investigating discrepancies between expressed privacy concerns and actual online behavior—a systematic literature review." *Telematics and Informatics.*, 34, 1038–1058. [1559]

Bergemann, Dirk, Ji Shen, Yun Xu, and Edmund Yeh (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. [1562]

Bird, Daniel and Zvika Neeman (2020), "What should a firm know? Protecting consumers' privacy rents." Working Paper. [1563]

Blumrosen, Liad and Michal Feldman (2013), "Mechanism design with a restricted action space." *Games and Economic. Behavior*, 82, 424–443. [1562]

Blumrosen, Liad, Noam Nisan, and Ilya Segal (2007), "Auctions with severely bounded communication." *Journal of Artificial Intelligence Research*, 28, 233–266. [1562]

Calzolari, Giacomo and Alessandro Pavan (2006), "On the optimality of privacy in sequential contracting." *Journal of Economic theory*, 130, 168–204. [1562]

Caplin, Andrew and Mark Dean (2015), "Revealed preference, rational inattention and costly information acquisition." *American Economic Review*, 105, 2183–2203. [1562]

Caplin, Andrew, Mark Dean, and John Leahy (2019), "Rational inattention, optimal consideration sets and stochastic choice." Review of Economic Studies. (forthcoming). [1562]

Choi, P. Jay, Doh-Shin Jeon, and Byung-Cheol Kim (2019), "Privacy and personal data collection with information externalities." *Journal of Public Economics*, 173, 113–124. [1559]

Conitzer, Vincent, Curtis R. Taylor, and Liad Wagman (2012), "Hide and seek: Costly consumer privacy in a market with repeat purchases." *Marketing Science*, 31, 277–292. [1562]

Cover, M. Thomas, and Joy A. Thomas (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley, New York. [1576]

Crémer, Jacques, Luis Garicano, and Andrea Prat (2007), "Language and the theory of the firm." *Quarterly Journal of Economics*, 122, 373–407. [1562]

Díaz, Claudia, Stefaan Seys, Joris Claessens, and Bart Preneel (2002), "Towards measuring anonymity." In *Proceedings of the 2nd International Conference on Privacy Enhancing Technologies (PET'02)* (R. Dingledine and P. Syverson, eds.), 54–68, Springer, Berlin, Heidelberg. [1564]

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith (2006), "Calibrating noise to sensitivity in private data analysis." In *Theory of Cryptography. TCC 2006* (S. Halevi and Rabin T., eds.), volume 3876 of Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. [1563]

Fadel, Ronald and Ilya Segal (2009), "The communication cost of selfishness." *Journal of Economic Theory*, 144, 1895–1920. [1562]

Fairfield, Joshua A. T. and Christoph Engel (2015), "Privacy as a public good." *Duke Law Journal*, 65, 385–457. [1559]

Fleischer, Lisa K. and Yu-Han Lyu (2012), "Approximately optimal auctions for selling privacy when costs are correlated with data." In *PProceedings of the 13th ACM Conference on Electronic Commerce (EC '12)*, 568–585, ACM, New York. [1563]

Frankel, Alexander (2014), "Aligned delegation." *American Economic Review*, 104, 66–83. [1562]

Ghosh, Arpita and Aaron Roth (2011), "Selling privacy at auction." In *Proceedings of the 12th ACM Conference on Electronic Commerce (EC '11)*, 199–208, ACM, New York. [1563]

Gilboa-Freedman, Gail and Rann Smorodinsky (2018), "On the properties that characterize privacy." Working Paper. [1562]

Gradwohl, Ronen (2018), "Privacy in implementation." *Social Choice and Welfare*, 50, 547–580. [1562]

Green, Jerry R. and Jean-Jacques Laffont (1986), "Incentive theory with data compression." In *Uncertainty, Information and Communication: Essays in Honor of Kenneth Arrow* (W. P. Heller, R. M. Starr, and D. A. Starrett, eds.), 239–253, Cambridge Univ. Press, Cambridge. [1562]

Green, Jerry R. and Jean-Jacques Laffont (1987), "Limited communication and incentive compatibility." In *Information, Incentives, and Economic Mechanisms: Essays in Honor of Leonid Hurwicz* (T. Groves, R. Radner, and S. Reiter, eds.), 308–329, Univ. Minnesota Press, Minneapolis. [1562]

Gul, Faruk and Andrew Postlewaite (1992), "Asymptotic efficiency in large exchange economies with asymmetric information." *Econometrica*, 60, 1273–1292. [1564]

Heffetz, Ori and Katrina Ligett (2014), "Privacy and data-based research." *Journal of Economic Perspectives*, 28, 75–98. [1563]

Hidir, Sinem and Nikhil Vellodi (2018), "Personalization, discrimination and information revelation." Working paper. [1563]

Ichihashi, Shota (2019), "Online privacy and information disclosure by consumers." American Economic Review. (forthcoming). [1563]

Kearns, Michael, Mallesh M. Pai, Aaron Roth, and Jonathan Ullman (2014), "Mechanism design in large games: Incentives and privacy." *American Economic Review*, 104, 431–435. [1563]

Kokolakis, Spyros (2017), "Privacy attitudes and privacy behavior: A review of current research on the privacy paradox phenomenon." *Computers and Security*, 64, 122–134. [1559]

Kos, Nenad (2012), "Communication and efficiency in auctions." *Games and Economic Behavior*, 75, 233–249. [1562, 1596]

Laibson, David (2018), "Private paternalism, the commitment puzzle, and model-free equilibrium (Richard T. Ely lecture)." *American Economic Review Papers and Proceedings*, 108, 1–21. [1559]

Ligett, Katrina and Aaron Roth (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)* (P. W. Goldberg, ed.), 378–391, Springer, Berlin, Heidelberg. [1563]

Matějka, Filip (2016), "Rationally inattentive seller: Sales and discrete pricing." *The Review of Economic Studies*, 83, 1125–1155. [1562, 1575]

Matějka, Filip and Alisdair McKay (2015), "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *The American Economic Review*, 105, 272–298. [1562]

McLean, Richard and Andrew Postlewaite (2002), "Informational size and incentive compatibility." *Econometrica*, 70, 2421–2453. [1564]

McSherry, Frank and Kunal Talwar (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. [1563]

Melumad, Nahum, Dilip Mookherjee, and Stephan Reichelstein (1992), "A theory of responsibility centers." *Journal of Accounting and Economics*, 15, 445–484. [1562]

Mookherjee, Dilip and Masatoshi Tsumagari (2014), "Mechanism design with communication constraints." *Journal of Political Economy*, 122, 1094–1129. [1562]

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18, 301–317. [1557, 1566]

Nissim, Kobbi, Rann Smorodinsky, and Moshe Tennenholtz (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. [1563]

Ollár, Mariann, Marzena Rostek, and Ji Hee Yoon (2016), "Privacy-preserving market design." Working Paper. [1563]

Pai, Mallesh and Aaron Roth (2013), "Mechanism design and privacy." *SIGecom Exchanges*, 12, 8–29. [1563]

Rebollo-Monedero, David, Jordi Forné, and Josef Domingo-Ferrer (2009), "From t-closeness-like privacy to postrandomization via information theory." IEEE Transactions on Knowledge and Data Engineering, 99(1). [1564]

Sankar, Lalitha, S. Raj Rajagopalan, and H. Vincent Poor (2013), "Utility-privacy tradeoffs in databases: An information-theoretic approach." *IEEE Transactions on Information Forensics and Security*, 8, 838–852. [1564]

Simon, Carl P. and Lawrence E. Blume (1994), *Mathematics for Economists*. Norton, New York. [1596]

Sims, Christopher. A. (2003), "Implications of rational inattention." *Journal of Monetary Economics*, 50, 665–690. [1562]

Steiner, Jakub, Colin Stewart, and Filip Matějka (2017), "Rational inattention dynamics: Inertia and delay in decision-making." *Econometrica*, 85, 521–553. [1562]

Taylor, Curtis R. (2004), "Consumer privacy and the market for customer information." *Rand Journal of Economics*, 35, 631–650. [1562]

Van Zandt, Timothy (2007), "Communication complexity and mechanism design." *Journal of European Economic Association*, 5, 543–553. [1562]

Wang, Weina, Lei Ying, and Junshan Zhang (2016), "On the Relation between Identifiability, Differential Privacy, and Mutual-information Privacy." *IEEE Transactions on Information Theory*, 62, 5018–5029. [1564]

Co-editor Marina Halac handled this manuscript.

Manuscript received 9 July, 2020; final version accepted 30 December, 2020; available online 21 January, 2021.