On FTC's "Do Not Track"

Raymond G. Sin
Hong Kong University of Science and Technology
rsin@ust.hk

and

Jia Jia
Hong Kong University of Science and Technology
jiajia@ust.hk

Abstract

The Federal Trade Commission (FTC) has recently proposed a "Do Not Track" mechanism in response to the fervent call for protecting consumer's privacy online. We argue that restricting information collection is a misplaced focus in addressing Internet privacy, and develop a mechanism that helps alleviate consumer's privacy concerns without sacrificing online firms' business benefits from using customer information. Building on the proposed mechanism, we derive alternative regulatory tools that can be readily available to policy makers, and investigate their respective effectiveness in improving social welfare.

We demonstrate that by leaving consumers partial control on how their information is used, the firm can devise a contract that serves the entire market while effectively catering to the privacy needs of different consumers. Further, results from our policy analysis suggest that imposing a requirement on preserving a portion of customer information purely for generating personalization is a superior strategy to restricting the firm's ability to collect personal information. Our modeling approach offers an alternative to the reliance on external instruments in traditional contract design, and extends the principal-agent framework to a three-tier interaction in a non-price context. Further, our work is one of the first that respond to the FTC's initiatives to pursue legislative options in protecting consumer's online privacy, and offers important guidelines to regulators for governing the information practice of online companies.

Keywords: "Do Not Track", mechanism design, policy analysis, privacy

1. Introduction

The Federal Trade Commission (FTC) has recently proposed a "Do Not Track" mechanism in response to the fervent call for protecting consumer's online privacy. This initiative focuses on restricting Internet companies' ability to gather personal information by granting consumers full authority to negate the tracking of their browsing and purchasing behaviors. While receiving support from the general public, the proposed approach neglects the sustainability of online businesses, such as Google, Facebook, and Last.fm, that are offering free personalizable services/contents to consumers under the premise that the cost of doing so can be justified by commercially exploiting the information gathered from consumers through their usage of these services. The predicament, in sum, is that without the ability to use customer information to generate revenues, firms would not have incentives to offer free services; without a mechanism in place to govern the usage and protection of customer data, consumers would not share their personal information (Giles 2010).

Given that privacy concerns threaten not only the profitability of personalization providers but also the very existence of the market itself, it is of particular interests to both the industry and regulators to identify an effective strategy to not only alleviate consumer's privacy concerns, but also preserve the firm's business benefits from using customer information. The goal of this paper is to devise such a mechanism, with a particular focus on the welfare implications of regulating online firm's information practice from a policy maker's perspective. In particular, our research seeks to address the following questions: 1. Should a regulator focus on restricting the firm's ability to collect customer information, or its subsequent usage of the gathered information? 2. What is the firm's corresponding optimal responses given the restrictions and how would they affect consumer surplus?

We portray a market where consumer's privacy concerns are sufficiently high to threaten the viability of the personalization market, and propose that by reserving a predetermined amount of information from secondary/commercial use (a concept referred to as "privacy preservation" in this paper), both the firm and the consumers are able to benefit from engaging in the personalization exchange. Our proposed mechanism leads to two alternative regulatory interventions, known as the "maximum information acquisition" and "minimum privacy preservation", respectively. The former policy corresponds to restricting the amount of information a firm can collect from consumers (for example, a regulator may allow the firm to collect demographic information from the user, but not her real-time location), while the later corresponds to imposing a minimum requirement on the portion of gathered information from which the firm can use only to deliver personalized services but not for commercial purposes (for example, if the preservation ratio is set at 1/2, then the firm is only allowed to share half of the gathered information with its restaurant partners to generate targeted advertisements). Our results suggest that, by leaving consumers partial control on how their information is used, the firm can devise a contract that serves the entire market while effectively catering to the privacy needs of different consumers. Further, results from our policy analysis suggest that imposing a requirement on preserving a portion of customer information purely for generating personalization is a superior strategy to restricting the firm's ability to collect personal information.

Our study is of theoretical and practical importance: Theoretically, we introduce information partition as a novel solution to the collapse in the number of available instruments that arises from the personalization-privacy tradeoff. By partitioning the set of information gathered from the consumer into two subsets that are used only for generating personalization and for generating revenues for the firm respectively, our solution approach offers an alternative to the reliance on external instruments (such as monetary transfer) in traditional mechanism designs. Further, our model is the first that extends the principal-agent framework to incorporate the effects of a third-party decision in the contract design in a non-price context, thus capturing the realistic dynamics of a three-tier interaction among the regulator, the firm, and consumers in the market for personalization.

Practically, the contract proposed in this study can be implemented through new industry standards such as the Accountable Hyper Text Transfer Protocol (HTTPA) (Seneviratne and Kagal 2011), which automates both the evaluation of the extent to which a certain (commercial) usage matches the previously agreed restrictions and the maintenance of a provenance log of data usage. These emerging technologies not only allow consumers control over how their shared information is used, but they also enable identification of potential violators, thus making it feasible for consumers and regulators to hold firms accountable for foul information practices.

Given the limited understanding of how consumer privacy concerns impact the adoption of personalization and thus the profitability of online companies (Bélanger and Crossler forthcoming, Smith, et al. forthcoming), our work not only bridges a gap in existing literature but is also one of the first that respond to the FTC's initiatives to pursue legislative options in protecting consumer's online privacy (FTC 2009, News 2010). Specifically, results from our welfare analysis offer important guidelines to the regulator on not only *what* devices can be applied in governing the information practice of online firms, but more importantly, exactly *how* and *why* enhancements in social efficiency can be achieved through a minimum privacy preservation policy. Our research has far-reaching implications, especially given the astonishing growth in the delivery of personalized digital contents for mobile devices that marks the inception of ubiquitous personalization in the digital era.

2. Literature Review

Personalization is an integral part of e-commerce strategy today. A unique feature of personalization is that users can enjoy more personalized services only if they are willing to disclose more information about themselves. Chellappa and Sin (2005) are among the first to formally study consumers' tradeoffs between personalization and privacy in the online context. They empirically examine the respective roles of consumers' valuation of personalized services and the privacy costs associated with sharing personal information in consumers' likelihood of adopting online personalization. Their research establishes that values for personalization and concerns for privacy are indeed two independent constructs, yet they jointly determine a consumer's decision to use personalized services. Further, different individuals may experience varying degrees of privacy concerns even when disclosing the same amount of information or if the same information is being used for secondary purposes. Such systematic differences in the preference for privacy can be attributed to individuals defining the information spaces of their social lives differently, being exposed to different situational cues and primed differently with regards to the consequences, or simply associating different probabilities or values on a given outcome (Hann, et al. 2007). Adding to the complexity of heterogeneity in consumers' attitudes towards privacy is the fact that privacy concerns are typically unobservable to a third-party (Chellappa and Shivendu 2010, Kobsa 2007). In other words, firms need to take into account not only these individual differences but also the lack of knowledge of such when making technology investment decisions and formulating its personalization strategy. As a result, the most privacy sensitive consumers may opt out completely from participating while those who choose to participate may over-represent their privacy concerns, leading to an inefficient market and the "privacy paradox" identified in extant literature (Awad and Krishnan 2006, Smith, et al. forthcoming). In particular, consumers' tendency to misrepresent their privacy sensitivity is a form of adverse selection, to which a solution cannot be attained through privacy-enhancing technologies alone (Kobsa 2007).

In this regard, the theory of incentives proposes that contract design can be a valuable tool for the decision maker to address such an issue. In particular, through a menu of options that specify the output/production level and the corresponding compensation, a principal can offer the appropriate incentives that induces the agents to behave desirably (so that in equilibrium, agents who are less productive will select a contract with lower production requirement with less compensation, while those who are most productive will select a contract with highest production requirement and compensation) (Laffont and Martimort 2002). These two components, also known as the "instruments", are required to be ex ante separable and independent of each other to ensure validity of the contract and its incentive compatibility. Ever since the seminal works by Mussa and Rosen (1978) and Maskin and Riley (1984) on monopolistic nonlinear pricing, this contract-theoretic approach has been applied to examine a broad variety of issues, such as governmental regulation of the private monopolist (Baron and Myerson 1982) and procurement (Laffont and Tirole 1993). In addition to the rich literature in economics, there has recently been a few applications in Information Systems on topics ranging from versioning and pricing of information goods (Huang and Sundararajan forthcoming, Sundararajan 2004b), to optimal sampling strategy and digital rights management (Sundararajan 2004a). Mechanism design is also recommended as a useful framework for studying the market for personal information, where firms can derive a mechanism to "purchase" information from consumers (Murthi and Sarkar 2003, Rust, et al. 2002). However, since both the benefit (value from personalization) and the cost (privacy concerns) are driven by the same underlying factor (information sharing) in the context of personalization (Adomavicius and Tuzhilin 2002, Volokh 2000), compensation and production are intrinsically correlated, hence poses a serious challenge to the contract design. Chellappa and Shivendu (2010) propose couponing as a compensation instrument in designing the menu to address consumers' privacy concerns associated with sharing personal information. Despite the novelty of their modeling approach, the applicability of such side payments or monetary subsidies is limited in practice. The closest real-life examples are Amazon's "A9 Instant Reward" and Microsoft's "Bing cashback" programs[1]. Unfortunately, both programs suffered similar fates in being discontinued within two years of their introduction (Turner and Wolfson 2010); most online firms today still rely primarily on personalized services alone in exchange for customer information. These observations

[1] A9 Instant Reward was more popularly known as the "$\pi / 2$ discount", which offered monetary rewards to users of their browser-embedded toolbar that delivers personalized search results. This program was launched in 2004 and discontinued in 2006. Bing cashback was launched in 2008 and discontinued in 2010.

point towards the need to explore alternative contract instruments for the market of personalization. Given that the current state of technology makes it feasible to hold accountable a firm's secondary use of information, our work proposes that some form of protection along this dimension can be used to construct a privacy-preserving contract.

Extant literature has established that firms can moderate consumers' privacy concerns if they can offer consumers a sense control over information access/collection and subsequent use (Hann, et al. 2007, Malhotra, et al. 2004, Tsai, et al. 2011). These findings suggest that by delegating consumers a certain level of control over their own information, firms can manage consumer's privacy concerns while strategically leveraging privacy protection to achieve their business objectives. However, the current focus on control over information collection (e.g., P3P architecture which requires online firms communicates purpose of information collection to Internet users and give consumers the ability to make choices upon each and every data request) is perhaps misplaced (Seneviratne and Kagal 2011); because future uses, sharing, and retention described by firms at the moment of data collection are not accountable ex post. A new stream of research has emerged and advocated accountability on subsequent use as a fundamental means through which our society addresses Internet privacy (Weitzner, et al. 2008). Specifically, this literature proposes making firms' information use transparent to consumers and regulators so that they may judge whether a particular use is appropriate, and seeks to develop the required infrastructure that enables individuals and institutions to be held accountable in case of misuse (e.g., TAMI (Transparent Accountable Datamining Initiative) project).

In a similar spirit, our work considers the loss of control over subsequent use of one's personal information as the primary source of consumer's privacy costs, and proposes protecting a predetermined portion of information from any secondary use to be an effective alternative that allows both the firm and the consumers to benefit from engaging in the personalization exchange. More importantly, our proposed concept of privacy preservation constitutes a practical solution for policy makers seeking to restore the delicate balance between allowing personal information to be used for business purposes and protecting consumer's privacy.

3. Model

3.1 Basic Setup

We consider a market where consumers are heterogeneous in their privacy sensitivity, denoted by $\theta \in [\underline{\theta}, \overline{\theta}]$, such that consumers with higher values of θ are more privacy sensitive. θ is privately known only to the consumers; from the firm's perspective it is drawn from a cumulative density $F(\cdot)$ with density function $f(\cdot)$. We assume that $f(\theta)$ is continuously differentiable, everywhere positive, and log-concave.[2]

A principal – an online firm – offers free personalization services that are considered valuable by the consumers. Consumers provide personal information that is required to enjoy the corresponding personalized services. Once this information (denoted by I) is acquired and processed, a spectrum of personalized offerings is being generated in return. The value (also referred to as convenience hereafter) that a consumer derives from consuming these services (denoted by $S(I)$) is assumed to be positively correlated with both the amount of information provided and the extent to which the firm is able to use such information in generating personalization. For tractability, we use a linear function to capture this intrinsic relationship between convenience and information acquisition:

$$S(I) = aI \tag{3.1}$$

[2] If $f(\theta)$ is continuously differentiable and log-concave, then it is unimodal (An 1998), and the reciprocal of its hazard rate, i.e., $F(\theta)/f(\theta)$, is non-decreasing, satisfying the monotone hazard rate property. Most common distributions satisfy these standard assumptions (Bagnoli and Bergstrom 2005). Further, with log-concavity of $f(\theta)$, $F(\theta)$ is also log-concave on $[\underline{\theta}, \overline{\theta}]$ (a standard proof is provided in the online appendix).

where the parameter a captures the firm's technological efficiency in generating convenience for the consumers from their personal information. This relationship is assumed to be deterministic, and is common knowledge to all parties in the market. Similar treatments can be found in Chellappa and Shivendu (2010).

In the absence of any privacy protection arrangement, consumers face the personalization-privacy tradeoff documented in extant literature. The disutility that arises from a consumer's privacy concern is assumed to be proportional to the full set of information collected by the firm, scaled by her privacy sensitivity parameter θ. The following function captures this tradeoff:

$$U(I,\theta) = S(I) - \theta I \tag{3.2}$$

Given $S = aI$, equation (3.2) can be rewritten as $U(I,\theta) = (a-\theta)I$. It implies that a consumer's decision on whether to subscribe to the proposed personalization services (or to "participate" in the market) depends on the relative magnitude of the efficiency coefficient a against her type coefficient θ. We label those consumers with $\theta > a$ "privacy seekers", as the costs of their privacy concerns outweigh the benefits derived from personalization so that they abstain from participating in the market; and those with $\theta \leq a$ "convenience seekers", as these consumers value convenience so much so that they are willing disclose as much information as possible in exchange for personalization. It can be observed that when the market comprises of only privacy seekers (i.e. $\underline{\theta} > a$), no consumer is willing to participate given the negative utility associated with using any level of personalization. This corresponds to the current FTC's proposal that allows consumers to completely negate firm's information collection, which renders the personalization market unviable. Our goal is to provide an alternative solution under such circumstances by introducing the notion of privacy preservation.

Privacy preservation (η) is defined as the portion of I (the set of information acquired by the firm) that is kept private by the firm and used purely for the purpose of generating personalization. Technology architecture (e.g., HTTPA (Seneviratne and Kagal 2011)) that holds the firm accountable for any violation to its privacy policy ensures η to be a valid instrument that constitutes the contract. With the credible commitment that η is not subject to secondary use, consumers' privacy concerns arise only from the disclosure of its complement i. Formally,

$$i = I - \eta \tag{3.3}$$

Hence under the preservation scheme, the utility that a consumer of type θ derives upon adopting personalization is given by:

$$U(i,\eta,\theta) = a(i+\eta) - \theta i \tag{3.4}$$

Firm's objective

The firm generates revenue from the set of information over which it can explore commercial possibilities, and incurs a constant marginal cost[3] in converting the acquired information into personalized services to its consumers. Without loss of generality, the marginal cost is normalized to 1. The firm's profit from serving each consumer is defined as:

$$\Pi(i,\eta) = \Phi(i) - (i+\eta) \tag{3.5}$$

where $\Phi(i)$ is the revenue function of the firm, and is assumed to be increasing (i.e., $\Phi_1(i) \geq 0$) and strictly concave (i.e., $\Phi_{11}(i) < 0$).

Due to information asymmetry on consumers' privacy sensitivity (θ), the firm's problem is to design a direct mechanism of $\{i(\theta),\eta(\theta)\}_{\theta\in[\underline{\theta},\bar{\theta}]}$ so that consumers of different types self-select into the desirable contract pair, allowing the firm to maximize profit. Formally, the firm's program is:

[3] This can be interpreted as a "resource cost" – the cost associated with the necessary computing resources in providing personalized content to a request (Liu, et al. 2010).

$$
\begin{aligned}
&\max_{i(\theta),\eta(\theta)} && \int_{\underline{\theta}}^{\bar{\theta}} \left[\Phi\left(i(\theta)\right) - \left(i(\theta) + \eta(\theta)\right)\right] f(\theta)\, d\theta \\
&s.t. && \theta = \arg\max_{\tilde{\theta}} a\left(i(\tilde{\theta}) + \eta(\tilde{\theta})\right) - \theta i(\tilde{\theta}) && \text{(I.C.)} \\
& && a\left(i(\theta) + \eta(\theta)\right) - \theta i(\theta) \geq 0 && \text{(I.R.)} \\
& && \eta(\theta) \geq 0, \text{ and } i(\theta) \geq 0 \qquad \forall \theta, \tilde{\theta} \in \left[\underline{\theta}, \bar{\theta}\right]
\end{aligned}
\tag{3.6}
$$

Our model formulation departs from traditional designs in that: 1) the two contracting variables in our model are additive separable, since our instruments (i, η) are set-complement; 2) the compensatory transfer is perceived differently by the principal and the agents; i.e. the principal's cost of delivering services is not identical to the benefits received by the agent.

3.2 Optimality of the privacy-preserving contract

In this section, we characterize the optimal contract of the firm and use it as a benchmark against which we compare the outcomes of two policy interventions in the next section. We show that the proposed contract pair allows the firm to implement an incentive-compatible menu that fully separates the market.

PROPOSITION 1. In the profit-maximizing allocation $\{i^*(\theta), \eta^*(\theta)\}$, $i^*(\theta)$ is characterized by

$$
a\Phi_1\left(i(\theta)\right) = \theta + \frac{F(\theta)}{f(\theta)} \tag{3.7}
$$

which is fully-separating; i.e. $i^*_1(\theta) < 0 \;\; \forall \theta \in \left[\underline{\theta}, \bar{\theta}\right]$.

Formal theorems and proofs are relegated to the online appendix. The form of the optimal contract in Proposition 1 is similar to that of conventional contracting problems involving monetary transfer. This similarity implies that regardless of the particular type of instrument that is used to compensate agents, the essential economic reasoning that governs the principal's contract design remains the same.

Conventional wisdom posits that under information asymmetry, it is costly for the principal in coordinating agents to reach an efficient use of economic resources (known as the "first-best solution" under complete information); because the agents intend to overstate their concerns, some information rents are required to elicit their truthful revelation about their types (i.e., privacy sensitivity). Larger rents are typically required for agents who are more productive (in our case, this refers to consumers who are more tolerant to commercial use of information, or equivalently, with lower privacy sensitivity), which can be reduced by lowering the production level (i.e. the amount of information used for commercial purposes) from the less productive agents (consumers who are more privacy sensitive). Confronting this tradeoff between efficient resource use and information rents, the firm chooses to sacrifice production from the less productive agents, a finding known as the "second-best solution" in literature. $F \;\theta \;/ f \;\theta$ in equation (3.10) represents the deviation from the socially-efficient resource use, and suggests that more severe distortion is associated with the segment of consumers who are more privacy sensitive.

Compared to the absence of contract (where the personalization market is unviable), the social welfare is unambiguously improved. Proposition 1 highlights the importance of incorporating privacy protection in the contract, which requires legislative efforts from the regulator to complement the underlying infrastructure and technological capability. In the following section, we relax the assumption that the firm has complete authority over specifying both the levels of information acquisition and privacy preservation in the contract. Specifically, we consider the scenario where a regulator takes a proactive role in protecting consumers' privacy online through two alternative legislative options, namely the minimum privacy preservation policy and the maximum information acquisition policy. We investigate the respective implications of these policies on social welfare by contrasting them with the second-best market outcome. Our goal is to inform policy makers on the choice of an optimal regulatory device and the corresponding magnitude of control to induce more efficient market outcomes.

A summary of notations used in our model is presented in Table 1.

<Table 1 about here>

4. Policy Analysis

4.1 Maximum information acquisition policy

We now explore a regulation whereby the policy maker restricts the firm's ability in acquiring customer information by setting an upper bound on the amount of information to be acquired. We refer to this upper bound as an "information boundary". This option corresponds to the perspective of restricting information collection that currently prevails among the industry and policy makers (Wyatt and Vega 2010a, Wyatt and Vega 2010b). To formally represent restriction on information acquisition, we introduce the following inequality into the firm's program:

$$i(\theta) + \eta(\theta) \leq \kappa \tag{4.1}$$

Note that the information boundary most directly affects the firm's strategy for the least privacy sensitive segment. Our solution approach involves a two-stage process: in the first stage, the regulator sets a boundary on information acquisition; then conditional on this restriction, the firm designs the optimal menu in the second stage. We derive the optimal level of information acquisition from the regulator's perspective using backward induction by first considering the effects of information boundary on the firm's optimal strategy. The following proposition summarizes the firm's best response to a boundary that dictates the maximum information it can collect from each consumer.

PROPOSITION 2. The optimal response of the firm to an information boundary on acquisition, denoted by the constraint in equation (4.1) that satisfies $\kappa < I^*(\underline{\theta})$, is to offer a menu $(i^B(\theta,\kappa), \eta^B(\theta,\kappa))$ such that

1. The optimal menu consists of an identical contract for consumers who are less privacy sensitive (those with $\theta \in \left[\underline{\theta}, \hat{\theta}^B(\kappa)\right)$) while fully differentiating the more privacy sensitive consumers (those with $\theta \in \left[\hat{\theta}^B(\kappa), \overline{\theta}\right]$).

2. Compared with the scenario in the absence of regulation, the firm uses less information from all consumers $\forall \theta \in \left[\underline{\theta}, \overline{\theta}\right]$ for commercial purposes, i.e. $i^B(\theta,\kappa) < i^*(\theta)$.

3. $\hat{\theta}_1^B(\kappa) < 0$; if $\theta - \frac{F(\theta)}{f(\theta)} = 0$ for some θ, $\hat{\theta}^B(\kappa) \in \left[\underline{\theta}, \tilde{\theta}\right)$ where $\tilde{\theta} \equiv \inf\left\{\theta : \theta - \frac{F(\theta)}{f(\theta)} = 0\right\}$: lowering information boundary leads to an expansion of the bunching interval, but the expansion does not necessarily lead to a purely-bunching result.

Proposition 2 indicates that the profit-maximizing strategy for the firm is to bunch not the less productive segment but instead the most efficient types; because doing so spares the firm from having to severely suppress information use from the more privacy sensitive consumers, which would have led to larger marginal losses given that information from this segment is already underexploited.

Further, observe that it is also suboptimal for the firm to adhere to the original menu (i.e., $\left(i^*(\theta), \eta^*(\theta)\right)$) and simply truncate it for the segment given by $\left[\underline{\theta}, I^{*-1}(\kappa)\right)$ (i.e., consumers whose full-set of information under the optimal menu exceeds the boundary κ) either. Had the boundary been imposed on use (i.e. $i(\theta) \leq \xi$) rather than collection (i.e. $I(\theta) \leq \kappa$), this truncating strategy could indeed have been the optimal one (we can denote the truncating strategy by $\left(i^T(\theta,\xi), \eta^T(\theta,\xi)\right)$; see Figure 1). However, comparing the truncating strategy with the optimal menu $\left(i^B(\theta,\kappa), \eta^B(\theta,\kappa)\right)$ presented in Proposition 2 suggests that the former strategy leads to a larger bunching region ($i^B(\theta,\kappa) < i^*(\theta)$ and $\eta^B(\theta,\kappa) < \eta^*(\theta)$) for the entire market; which implies $I^{*-1}(\kappa) > \hat{\theta}^B(\kappa)$. In other words, with the truncating strategy, the firm may be sacrificing too much from the most productive segment on the desired differentiability of consumers.

<Figure 1 about here>

Suppressing information use from the less productive segment enables the firm to reduce the transfer of information rents to the more efficient segment, and to allocate a lesser amount of information out of the acquired information towards preservation (a decrease in η); the remaining subset of information can then be used to restore differentiability on the less privacy sensitive segment. In sum, this proposition demonstrates a delicate adjustment by the firm (which is characterized by Theorem 1 in the online appendix) in the face of restriction on information acquisition: suppressing information use from the most privacy sensitive consumers reserves a certain degree of differentiation for the least privacy-concerned segment, but not so much so that it triggers the aforementioned marginal losses.

Proposition 2 also states that the size of the bunching region is a function of the magnitude of the information boundary, $\hat{\theta}^B(\kappa)$. As the policy maker tightens the information boundary, a larger portion of consumers will be served an identical contract (though this may not result in *all* consumers being served the same contract; i.e. a purely bunching solution). Part 3) of Proposition 2 implies that the purely bunching outcome can occur only when $\theta > F(\theta)/f(\theta)$, a condition that characterizes a market with evenly distributed consumer types (e.g. under uniform distribution).

Anticipating the effect of its intervention on market outcome, the regulator decides whether to introduce this regulation as well as the specific level of such a boundary. Given that $\hat{\theta}^B(\kappa)$ is well-behaved, we can formulate the social planner's objective as a function of κ:

$$\begin{aligned} SW(\kappa) = &\int_{\underline{\theta}}^{\overline{\theta}} \left[\Phi\left(i^B(\theta,\kappa)\right) - i^B(\theta,\kappa) + \eta^B(\theta,\kappa)\right] f(\theta)\, d\theta \\ &+\alpha \int_{\underline{\theta}}^{\overline{\theta}} \left[a\left(i^B(\theta,\kappa) + \eta^B(\theta,\kappa)\right) - \theta i^B(\theta,\kappa)\right] f(\theta)\, d\theta \end{aligned} \tag{4.2}$$

where $\alpha > 0$ denotes the weight of the surplus of consumers relative to that of the firm assigned by the regulator in her objective. Solving the social planner's problem leads to the following proposition:

PROPOSITION 3. A maximum information acquisition policy unambiguously reduces social welfare.

The intuition behind this result is that the maximum information acquisition policy obliges the firm to gather less information from all consumers, while forcing the firm to decrease the level of privacy protection for the entire market. Therefore, both the firm's and consumer's surpluses suffer from the information boundary, resulting in a reduction in social welfare compared with the benchmark case. We shall elaborate on the underlying economic tradeoffs using a simple example in Section 4.3.

4.2 Minimum privacy preservation policy

Unlike the maximum information acquisition policy that aims to restrict information collection, a minimum privacy preservation policy mandates the firm to reserve a predetermined proportion of information (also referred to as "preservation ratio") from secondary use for all consumers. Formally, we introduce the following inequality into the firm's program[4]:

$$\eta \;\theta \;\geq \beta \;i \;\theta \;+ \eta \;\theta \tag{4.3}$$

where β denotes the mandatory ratio.

It is not as evident as the former scenario as to which portion of the market will be first affected by this intervention. Lemma 2 (in the online appendix) characterizes the property of the ratio

[4] The reason behind expressing the minimum preservation as a ratio to the total amount of information acquired by the firm, as opposed to an absolute value, is that otherwise the firm can respond by arbitrarily increasing information acquisition to counteract the effects of such a policy.

$\Delta(\theta) \equiv \dfrac{\eta(\theta)}{i(\theta)+\eta(\theta)}$ for an arbitrary incentive compatible menu. It suggests that an incentive menu with the participation constraint binding at $\bar{\theta}$ guarantees consumers with higher privacy sensitivity a larger preservation ratio. As the optimal contract $\left(i^*(\theta), \eta^*(\theta)\right)$ belongs to this class of menus, $\Delta^*(\underline{\theta})$ is found to be the lowest attainable preservation ratio based on the firm's self-interest. Any intervention with $\beta \le \Delta^*(\underline{\theta})$ is therefore irrelevant to the firm's objective, because the firm can simply adhere to the original menu $\left(i^*(\theta), \eta^*(\theta)\right)$ to maximize the profit. Hence $\Delta^*(\underline{\theta})$ is the lower bound of an effective regulatory intervention. In the subsequent discussion, we restrict our attention to cases where $\beta > \Delta^*(\underline{\theta})$.

All consumers with types lower than $\Delta^{*-1}(\beta)$ are provided with less preservation by the menu $\left(i^*(\theta), \eta^*(\theta)\right)$ than that required by regulation, and thus most directly affected by the preservation ratio. The next proposition establishes the firm's best response to the minimum privacy preservation policy.

PROPOSITION 4. The optimal response of the firm to a preservation ratio on use, denoted by the constraint in equation (4.3) that satisfies $\Delta^*(\underline{\theta}) < \beta \le (\bar{\theta} - a)/\bar{\theta}$, is to offer a menu $\left(i^P(\theta,\beta), \eta^P(\theta,\beta)\right)$ such that:

1. The optimal menu consists of an identical contract for consumers who are less privacy sensitive (those with $\left[\underline{\theta}, \hat{\theta}^P(\beta)\right)$) and fully differentiates the more privacy sensitive consumers (those with $\left[\hat{\theta}^P(\beta), \bar{\theta}\right]$).
2. Compared with the scenario in the absence of regulation, the firm commercially uses more information from more privacy sensitive consumers, i.e. $i^P(\theta,\beta) > i^*(\theta) \quad \forall \theta \in \left[\hat{\theta}^P(\beta), \bar{\theta}\right]$.

3. $\hat{\theta}_1^P(\beta) > 0$ for $\Delta^*(\underline{\theta}) < \beta \leq (\overline{\theta} - a)/\overline{\theta}$: increasing preservation ratio induces the firm to bunch a larger portion of the market. Once the mandatory ratio becomes sufficiently high (i.e. $\beta > (\overline{\theta} - a)/\overline{\theta}$), all consumers are served a single contract, and the participation constraint becomes slack.

Note that the firm's ability to differentiate consumers belonging to the least privacy sensitive segment is once again deprived due to the introduction of the regulation. Similar to the former regulatory option, the intervention influences the entire population rather than a particular segment (given by $\left[\underline{\theta}, \Delta^{*-1}(\beta)\right)$ for whom the original contract specified in $(i^*(\theta), \eta^*(\theta))$ violates this regulation. Another salient feature of this menu is that the firm engages in heavier commercial use of information from consumers with relatively high privacy sensitivity ($\theta \in \left[\hat{\theta}^P(\beta), \overline{\theta}\right]$) than without such regulation. Although it may appear to contradict the purpose of the privacy preservation policy, an increase in commercial use is in fact associated with a larger increase in compensation through "free" personalization; hence highly privacy sensitive consumers are indeed better off (observe that $U^P(\theta) = \int_\theta^{\overline{\theta}} i^P(t,\beta)\,dt > U^*(\theta) = \int_\theta^{\overline{\theta}} i^*(t)\,dt$).

A closer look at the constraint reveals that it is equivalent to a type-dependent participation constraint (i.e. the reservation value of consumers imposed by the policy is a function of their respective types) of the form $U(\theta) = a\,i(\theta) + \eta(\theta) - \theta i(\theta) \geq \left(\frac{a}{1-\beta} - \theta\right) i(\theta)$. The preservation constraint implicitly requires the firm to leave consumers with privacy sensitivity θ with at least $\left(\frac{a}{1-\beta} - \theta\right) i(\theta)$ surplus. Since $\left(\frac{a}{1-\beta} - \theta\right) i(\theta)$ is more convex than $U(\theta)$ (see the solid line in

Figure 2), this constraint starts binding for consumers with lower privacy sensitivity, who are not being adequately compensated to meet the level of surplus required by the minimum preservation policy. Ceteris paribus, to comply with the regulation, the firm needs to offer consumers in this segment a relatively high level of "free personalization" η ; which, in turn, makes the contract very attractive to the highly privacy sensitive group, creating "countervailing incentives" (i.e. high privacy sensitive types has incentives to pretend to be low privacy sensitive types, rather than the other way around as discussed in Proposition 1. Laffont and Martimort (2002), p. 104).

Though by adopting a truncating strategy (i.e. a menu that retains the same form as $i^* \theta ,\eta^* \theta$ until the region that is affected by the regulation, where the menu becomes constant and binds at the required preservation level; see Figure 2), the firm can completely eliminate such countervailing incentives, this strategy is suboptimal because it fails to take advantage of the fact that secondary use of information is more costly for those with high privacy sensitivity (marginal cost increases with θ); i.e. the firm can instead prevent consumers from understating their privacy sensitivity through increasing i for consumers with lower privacy sensitivity. Note, however, that by increasing i the firm also inadvertently tightens the participation constraint, leading to an expansion in the bunching region and larger efficiency loss. Therefore, the firm's optimal degree of the upward adjustment in i balances the two effects, which is a desirable outcome from the regulator's perspective because the underproduction problem is moderated, especially on the highly privacy sensitive segment.

<Figure 2 about here>

Comparing $U \theta$ under the two menus depicted in

Figure 2 reveals that the firm's optimal menu under the minimum privacy preservation policy leads to an increase in consumer surplus for all consumers except the least privacy sensitive ones. In other words, through a preservation ratio, the regulator can effectively manipulate the allocation of social welfare between the firm and consumers.

Proposition 4 also implies that the size of the bunching region is a function of the magnitude of the preservation ratio, $\hat{\theta}^P(\beta)$. As the regulator raises the preservation proportion, a larger segment of consumers are served an identical contract. Anticipating the effect of this policy on market outcome, the regulator decides whether or not, and to what extent, to exercise this intervention. Specifically, we can formulate the social planner's objective as a function of β:

$$\begin{aligned} SW(\beta) &= \int_{\underline{\theta}}^{\bar{\theta}} \left[\Phi\left(i^P(\theta,\beta)\right) - i^P(\theta,\beta) + \eta^P(\theta,\beta)\right] f(\theta)\, d\theta \\ &+\alpha \int_{\underline{\theta}}^{\bar{\theta}} \left[a\left(i^P(\theta,\beta) + \eta^P(\theta,\beta)\right) - \theta i^P(\theta,\beta)\right] f(\theta)\, d\theta \end{aligned} \tag{4.4}$$

Solving the social planner's problem leads to the following proposition:

PROPOSITION 5. A minimum privacy preservation policy enhances social welfare.

Proposition 5 offers important insights for a policy maker pursuing legislative options in governing the information practice of Internet firms. Our findings suggest that, although the firm has incentives to introduce privacy-preservation in the contract design, the level of preservation arising from this self-serving objective is not sufficient from the society's point of view. We find that social efficiency can be enhanced by simply introducing a minimum preservation policy by which the firm is required to preserve a portion of the acquired information from any secondary uses.

In order to demonstrate the specific effects of the two policies and the underlying economic tradeoffs, in the following section we shall illustrate their respective welfare implications using specific functional forms and uniform distribution of privacy sensitivity.

4.3 Welfare implications of the two policies

In this section, we apply the general results derived above to a simple example and draw specific guidelines for the regulator. Specifically, we assume consumer types to follow a uniform distribution, and that the firm's profit function takes the form of:

$$\Pi(i,\eta) = 2b\sqrt{i} - (i+\eta).$$

The closed-form solutions under the three cases are summarized in Table 2.

<Table 2 about here>

Since the two regulatory options are not directly comparable, we shall first derive the relationship between the values of β and κ so as to make the two policies analogous. For example, in the absence of regulation, the optimal contract for consumers of type $\theta = \frac{2}{11}(10+3\sqrt{5})$ involves gathering $I^*(\theta)|_{\theta=\frac{2}{11}(10+3\sqrt{5})} = 2$ amount of information and preserving $\Delta^*(\theta)|_{\theta=\frac{2}{11}(10+3\sqrt{5})} = \frac{3}{8}(-7+4\sqrt{5})$ proportion from secondary use. Hence $\kappa = 2$ and $\beta = \frac{3}{8}(-7+4\sqrt{5})$ can be perceived as comparable policies, because the consumer segments for which the requirements are violated under the menu $(i^*(\theta), \eta^*(\theta))$ are equivalent.

Figures 3 and 4 depict the amount of information used for commercial purposes and preserved from secondary use under the market support of $[2,4]$, $a = 1$ and $b = 3$.

Figure **7** describes the ratio of information preserved from secondary uses to the total amount of information acquired.

<Figures 3 to 5 about here>

Compared with the absence of regulation, the optimal menus under the two policies exhibit two distinctive features: first, the least privacy sensitive segments are bunched; second, underproduction from the least productive segments is either intensified (under maximum information acquisition policy) or moderated (under minimum privacy preservation policy). The vertical dotted lines in Figures 3 and 4 represent the marginal type of the affected consumer segment (denoted by $\hat{\theta}^T$), while the horizontal dotted lines represent the levels of commercial use and privacy preservation, respectively. Note that the horizontal dotted line and the portion of $(i^*(\theta), \eta^*(\theta))$ underneath this line constitute the truncating strategy $(i^T(\theta,\xi), \eta^T(\theta,\xi))$, which is the firm's best response to a regulation $i(\theta) \le \xi = i^*(\hat{\theta}^T)$.

In addition to highlighting these two features, it would also be useful to compare optimal menus under information boundary and preservation ratio (i.e. $(i^B(\theta,\kappa), \eta^B(\theta,\kappa))$ and $(i^P(\theta,\beta), \eta^P(\theta,\beta))$ respectively) with the truncating option (i.e. $(i^T(\theta,\xi), \eta^T(\theta,\xi))$).

Given that $i^T(\hat{\theta}^T,\xi) + \eta^T(\hat{\theta}^T,\xi) = i^B(\hat{\theta}^B,\kappa) + \eta^B(\hat{\theta}^B,\kappa) = \kappa$ (i.e., the total amount of information acquired from the most productive segment stays the same), the observation that $i^T(\hat{\theta}^T,\xi) < i^B(\hat{\theta}^B,\kappa)$ and $\eta^T(\hat{\theta}^T,\xi) > \eta^B(\hat{\theta}^B,\kappa)$ indicates that restriction on information collection allows the firm to redistribute information between commercial use and preservation so as to moderate the boundary effect on its profitability on this segment. Also observe that compared with $(i^T(\theta,\xi), \eta^T(\theta,\xi))$, $(i^P(\theta,\beta), \eta^P(\theta,\beta))$ leads to larger amount of information being acquired from the consumers, along with both heavier commercial use and larger preservation, which corresponds to the firm's efficient way to offset the countervailing incentives.

Figure **8** describes the relationship between the firm's profit and $\hat{\theta}^T$ under $\left(i^T(\theta,\xi), \eta^T(\theta,\xi)\right)$, $\left(i^B(\theta,\kappa), \eta^B(\theta,\kappa)\right)$, and $\left(i^P(\theta,\beta), \eta^P(\theta,\beta)\right)$, respectively. Note that $\xi = i^*(\hat{\theta}^T)$, $\kappa = I^*(\hat{\theta}^T)$, and $\beta = \Delta^*(\hat{\theta}^T)$. The dotted line represents the firm's original profit in the absence of the regulator's intervention. It is evident that the firm becomes worse off under all three menus, and that the truncating strategy always generates less profit for the firm compared with the firm's optimal responses under two regulatory scenarios. In addition, the menu $\left(i^P(\theta,\beta), \eta^P(\theta,\beta)\right)$ dominates both the truncated menu and the optimal menu under maximum information acquisition from the firm's perspective.

<Figure 6 about here>

To illustrate the differences in social welfare under the three optimal menus, we assume that the weighted parameter α equals $1/a$. The main purpose for such an assumption is to align the social planner's objective with the firm's first-best solution (under complete information)[5]. This can be considered as a regulator providing incentives to the firm for advancing personalization technology (an increase in a).

<Figure 7 about here>

The dotted line in **Figure 9** represents the original social welfare in the absence of the regulator's intervention. Observe that the minimum privacy preservation policy can indeed improve social welfare

[5] When $\alpha = 1/a$, the regulator's objective can be reduced to $\max_{i(\cdot)} \int_{\underline{\theta}}^{\bar{\theta}} \left(\Phi(i(\theta)) - \frac{\theta}{a} i(\theta) \right) f(\theta)\, d\theta$, which is exactly the firm's objective with complete information on θ.

(dash line in **Figure 9**). Further, the optimal regulation level under $\left(i^P(\theta,\beta), \eta^P(\theta,\beta)\right)$ falls within $\hat{\theta}^T \in \left(\underline{\theta}, \overline{\theta}\right)$, or equivalently, $\beta \in \left(\Delta^*(\underline{\theta}), (\overline{\theta} - a)/\overline{\theta}\right)$. This indicates that the regulator prefers to induce a partially-separating market outcome. Through a preservation ratio, the regulator can effectively manipulate the allocation of surplus between the firm and consumers; at optimality, more surplus is transferred to the majority of consumers. This allocation induces the firm to moderate the underproduction problem (especially on the most privacy sensitive segment), thus creating a relatively high the marginal gain from a welfare-maximizing perspective. The allocation efficiency brought about by this policy, however, does come at a cost: as the required reservation value is more convex, the firm's ability to differentiate the most productive segment is always deprived; which in turn results in larger losses associated with a higher preservation ratio (β). When the preservation ratio becomes sufficiently high, the latter effect dominates the former and results in reduction in the overall social welfare.

Interestingly, though the menu $\left(i^P(\theta,\beta), \eta^P(\theta,\beta)\right)$ still dominates both the truncated menu and the optimal menu under maximum information acquisition, the truncating strategy now becomes more favorable than $\left(i^B(\theta,\kappa), \eta^B(\theta,\kappa)\right)$ (compared with the firm's perspective). As $i(\theta) \leq \xi = i^*(\hat{\theta}^T)$ and $i(\theta) + \eta(\theta) \leq \kappa = I^*(\hat{\theta}^T)$ are considered regulations of the same intensity, the difference in social welfare between the two curves under $\left(i^T(\theta,\xi), \eta^T(\theta,\xi)\right)$ and $\left(i^B(\theta,\kappa), \eta^B(\theta,\kappa)\right)$ quantifies the additional efficiency loss due to reallocation between commercial use and preservation. This implies that the freedom to adjust information use accommodated by regulation on collection is detrimental from the society's perspective.

5. Conclusion

5.1 Implications of our work

Information technology plays a critical role in both the provision and usage of personalization by online firms and consumers. On one hand, more sophisticated technologies allow firms to deliver more and better personalized services to consumers while seamlessly sharing more information with their business partners, thus allowing for higher profit potentials; on the other hand, the very same technologies that enable firms to share larger amounts of information more quickly with more parties also exacerbate consumers' concerns for privacy. The tradeoffs between personalization and privacy can be summarized as the following predicament: without the ability to use customer information to generate revenues, firms would not have incentives to offer free personalization; without a mechanism in place to govern the usage and protection of customer data, consumers would not share their personal information. Therefore, it is of particular interests to both the industry and regulators to identify an effective strategy to not only alleviate consumer's privacy concerns, but also preserve the firm's business benefits from using customer information.

Our work argues that consumer-centric policies, such as the FTC's "Do Not Track" proposal, are prone to diminishing business viability of online personalization providers, and seeks to reconcile the personalization goals of the firm and consumers' privacy concerns. We develop a mechanism that helps alleviate consumer's privacy concerns without sacrificing online firms' business benefits from using customer information, and show that by leaving consumers partial control on how their information is used, the firm can devise a contract that serves the entire market while effectively catering to the privacy needs of different consumers.

Our proposed mechanism leads to two alternative regulatory interventions, known as the "maximum information acquisition" and "minimum privacy preservation", respectively. The former policy corresponds to restricting the amount of information a firm can collect from consumers, while the later corresponds to imposing a minimum requirement on the portion of gathered information from which the firm can use only to deliver personalized services but not for commercial purposes. Our findings suggest that imposing a requirement on minimum privacy preservation is a superior strategy to restricting the firm's ability to collect personal information, which is indeed welfare reducing.

Our work is one of the first that respond to the FTC's initiatives in pursuing legislative options in protecting consumer's online privacy and offers important guidelines to policy makers in addressing the growing concerns over online privacy. We show that while online personalization firms may have economic incentives to provide consumers with a certain level of privacy preservation, the level of protection arising from this self-serving objective may not be sufficient from the society's perspective.

An important insight from this research is that a solution to the privacy paradox cannot be attained through privacy-enhancing technologies alone; while technology architecture – such as the TAMI and HTTPA that makes contracting the usage of information feasible – may provide the capability of *implementing* a solution, the *actual* solution itself lies in the knowhow by the firm and the regulator. In other words, technology may make some decision rules implementable, but the true value lies within the decisions that induce optimal market outcomes. Our work embraces this challenge to arrive at such decisions, and points towards an important direction that, given current technologies enable verifiability of online companies' information practice, a more central role that the FTC can play is to ensure enforcement through holding the firms liable for any misuse of customer information.

5.2 Future research directions

In this paper, we provide a solution to the case where consumers' privacy concerns are sufficiently high to threaten the de-facto existence of the personalization market. Our proposed mechanism can be readily extended to examine a more general scenario where both privacy-seekers, who refrain from participating completely, and convenience-seeking consumers who prefer as much personalization as possible, coexist. Further, a sensitivity analysis on the distribution of consumer types under the generalized market conditions can reveal the effects of changes in consumers' privacy concerns on the optimal contract and equilibrium market outcome. Finally, we have assumed in our model that technology efficiency is exogenously determined; in reality, online portals often need to decide on the investment level in personalization technologies prior to engaging in an information-sharing contract with the consumers. A model that endogenizes such an investment decision and hence the corresponding efficiency by which the firm is able to generate values to their consumers would be an interesting direction to pursue.

References

Adomavicius, Gediminas and Alexander Tuzhilin, "An Architecture of e-Butler- A Consumer Centric Online Personalization Mechanism," *International Journal of Computational Intelligence and Applications*, 3, 2, (2002), 313-327.

An, Mark Yuying, "Logconcavity versus Logconvexity: A Complete Characterization," *Journal of Economic Theory*, 80, 2, (1998), 350-369.

Awad, Naveen Farag and M. S. Krishnan, "The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to be Profiled Online for Personalization," *MIS Quarterly*, 30, 1, (2006), 13-28.

Bagnoli, Mark and Ted Bergstrom, "Log-concave probability and its applications," *Economic Theory*, 26, 2, (2005), 445-469.

Baron, D. P. and R. B. Myerson, "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50, 4, (1982), 911-930.

Bélanger, France and Robert E. Crossler, "Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems," *MIS Quarterly*, (forthcoming),

Chellappa, Ramnath K. and Shivendu Shivendu, "Mechanism Design for "Free" but "No Free Disposal" Services: The Economics of Personalization Under Privacy Concerns," *Management Science*, 56, 10, (2010), 1766-1780.

Chellappa, Ramnath K. and Raymond Sin, "Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma," *Information Technology and Management*, 6, 2-3, (2005), 181-202.

FTC, "Beyond Voice: Mapping the Mobile Marketplace," Federal Trade Commission, 2009.

Giles, Martin, "Privacy 2.0 - Give a little, take a little," *Economist*, 394, 8667, (2010), 18-19.

Hann, Il-Horn, Kai-Lung Hui, Sang-Yong Tom Lee and Ivan P. L. Png, "Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach," *Journal of Management Information Systems*, 24, 2, (2007), 13-42.

Huang, Ke-Wei and Arun Sundararajan, "Pricing Digital Goods: Discontinuous Costs and Shared Infrastructure," *Information Systems Research*, (forthcoming), .

Kobsa, Alfred, "Privacy-enhanced Personalization," *Communications of the ACM*, 50, 8, (2007), 24-33.

Laffont, Jean-Jacques and David Martimort, *The Theory of Incentives - The Principal-Agent Model*, Princeton University Press, Princeton, 2002.

Laffont, Jean-Jacques and Jean Tirole, *A Theory of Incentives in Procurement and Regulation*, The MIT Press, Cambridge, 1993.

Liu, Dengpan, Sumit Sarkar and Chelliah Sriskandarajah, "Resource Allocation Policies for Personalization in Content Delivery Sites," *Information Systems Research*, 21, 2, (2010), 227-248.

Malhotra, Naresh K., Sung S. Kim and James Agarwal, "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model," *Information Systems Research*, 15, 4, (2004), 336-355.

Maskin, Eric and John Riley, "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15, 2, (1984), 171-196.

Murthi, B. P. S. and Sumit Sarkar, "The Role of the Management Sciences in Research on Personalization," *Management Science*, 49, 10, (2003), 1344-1362.

Mussa, M. and S. Rosen, "Monopoly and Product Quality," *Journal of Economics Theory*, 18, (1978), 301-317.

News, FTC, "FTC Testifies on Protecting Teen Privacy," 07/15/2010, 2010,

Rust, Roland T., P. K. Kannan and Na Peng, "The Customer Economics of Internet Privacy," *Journal of the Academy of Marketing Science*, 30, 4, (2002), 455-464.

Seneviratne, Oshani and Lalana Kagal, "Addressing Data Reuse Issues at the Protocol Level," *Proceedings of the IEEE International Symposium on Policies for Distributed Systems and Networks (POLICY)*, Pisa, 2011, 141-144.

Smith, H. Jeff, Tamara Dinev and Heng Xu, "Information Privacy Research: An Interdisciplinary Review," *MIS Quarterly*, (forthcoming), .

Sundararajan, Arun, "Managing Digital Piracy: Pricing and Protection," *Information Systems Research*, 15, 3, (2004a), 287-308.

Sundararajan, Arun, "Nonlinear Pricing of Information Goods," *Management Science*, 50, 12, (2004b), 1660-1673.

Tsai, J. Y., S. Egelman, L. Cranor and Alessandro Acquisti, "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," *Information Systems Research*, 22, 2, (2011), 254-268.

Turner, Nick and Lisa Wolfson, "Microsoft to Terminate Bing Cash-Back Program July 30 (Update3)," *Businessweek*, Issue Number, June 07 2010,

Volokh, Eugene, "Personalization and privacy," *Communications of The ACM*, 43, 8, (2000), 84-88.

Weitzner, Daniel J., Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler and Gerald Jay Sussman, "Information Accountability," *Communications of the ACM*, 51, 6, (2008), 82-87.

Wyatt, Edward and Tanzina Vega, "F.T.C. Backs Plan to Honor Privacy of Online Users," *New York Times*, Available online: http://www.nytimes.com/2010/12/02/business/media/02privacy.html?pagewanted=all, (2010a),

Wyatt, Edward and Tanzina Vega, "Stage Set for Showdown on Online Privacy," *New York Times*, Available online: http://www.nytimes.com/2010/11/10/business/media/10privacy.html, (2010b),

Appendix: Tables and Figures

Table 1: Notations

Symbol	Definition
I	The amount of information that the firm gathered through consumers' usage of personalization.
a	Personalization efficiency that measures the values that the firm creates for consumers based on a unit of personal information.
$S\ \ I$	The value of personalization generated by the firm based on I amount of customer information.
$\theta \in \left[\underline{\theta}, \overline{\theta}\right]$	Consumer's privacy attitude regarding secondary use of their personal information, distributed with p.d.f. $f\ \cdot$ and c.d.f. $F\ \cdot$.
$U\ \ I, \theta$	Surplus that consumer of type θ obtains from subscribing the personalization services, assuming all information shared is subject to secondary use.
i	The subset of information that will be used for (commercial) purposes other than provision of personalization.
η	The subset of information that is free of secondary use.
$U\ \ i, \eta, \theta$	Surplus that consumer θ obtains from subscribing a personalization service, where η units of information are reserved from secondary use.
$\Phi(i)$	The revenue function of the firm. It is assumed to be increasing (i.e., $\Phi_1\ \ i\ \ \geq 0$) and strictly concave (i.e., $\Phi_{11}\ \ i\ \ < 0$).
$\Pi\ \ i, \eta$	Firm's profit from commercially using i units of information while incurring costs to generate personalization based on $i + \eta$ units of information
$i\ \ \theta\ \ , \eta(\theta)$	The incentive compatible menu in the absence of regulation.
$i^*\ \ \theta\ \ , \eta^*(\theta)$	Profit-maximizing menu in the absence of regulation
κ	Regulatory intervention that restricts the firm to acquire at most κ units of information from each of the consumers.
$i^B\ \ \theta, \kappa\ \ , \eta^B\ \ \theta, \kappa$	Profit-maximizing menu under the maximum acquisition policy.
$\hat{\theta}^B$	The marginal consumer type that characterizes the size of the bunching region under the menu $i^B\ \ \theta, \kappa\ \ , \eta^B\ \ \theta, \kappa$.
β	Regulatory intervention that requires the firm to preserve at least β portion of information from secondary use.
$i^P\ \ \theta, \beta\ \ , \eta^P(\theta, \beta)$	Profit-maximizing menu under the minimum preservation policy.
$\hat{\theta}^P$	The marginal consumer type that characterizes the size of the bunching region under the menu $i^P\ \ \theta, \beta\ \ , \eta^P(\theta, \beta)$
ξ	Auxiliary restriction on commercial use of customer information.
$i^T\ \ \theta, \xi\ \ , \eta^T(\theta, \xi)$	Profit-maximizing menu under the auxiliary restriction on commercial use of customer information.
$\hat{\theta}^T$	The marginal consumer type that characterizes the size of the bunching region

	under the menu $i^T\ \theta,\xi\ ,\eta^T(\theta,\xi)$.

Table 2: Summary of results under uniform distribution and specific functional forms

	Case 1: No Regulation	Case 2: Maximum Information Acquisition	Case 3: Minimum Privacy Preservation
$i \;\; \theta$	$\left(\frac{ab}{2\theta-\underline{\theta}}\right)^2$	$\left(\frac{ab}{2\theta-\underline{\theta}+\;\;\overline{\theta}-\underline{\theta}\;\;\hat{\lambda}^B}\right)^2$ if $\theta \geq \hat{\theta}^B$; else, $\left(\frac{ab}{2\hat{\theta}^B-\underline{\theta}+\;\;\overline{\theta}-\underline{\theta}\;\;\hat{\lambda}^B}\right)^2$	$\left(\frac{ab}{2\theta-\underline{\theta}-\;\;1-\beta\;\;\overline{\theta}-\underline{\theta}\;\;\hat{\lambda}^P}\right)^2$ if $\theta \geq \hat{\theta}^P$; else, $\left(\frac{ab}{2\hat{\theta}^P-\underline{\theta}-\;\;1-\beta\;\;\overline{\theta}-\underline{\theta}\;\;\hat{\lambda}^P}\right)^2$
$\eta \;\; \theta$	$\frac{1}{2a}\left(\begin{array}{l}\left(\frac{ab}{\sqrt{i}}+\underline{\theta}-2a\right)i \\ -ab\;\;\sqrt{\overline{i}}-\sqrt{i}\end{array}\right)$	$\frac{1}{2a}\left(\begin{array}{l}\left(\frac{ab}{\sqrt{i}}+\underline{\theta}-\;\;\overline{\theta}-\underline{\theta}\;\;\hat{\lambda}^B-2a\right)i \\ -\;\;ab\;\;\sqrt{\overline{i}}-\sqrt{i}\end{array}\right)$	$\frac{1}{2a}\left(\begin{array}{l}\left(\frac{ab}{\sqrt{i}}+\underline{\theta}+\;\;1-\beta\;\;\overline{\theta}-\underline{\theta}\;\;\hat{\lambda}^B-2a\right)i \\ -\;\;ab\;\;\sqrt{\overline{i}}-\sqrt{i}\end{array}\right)$
$\hat{\theta}$ (the size of the bunching region)	(Fully separating)	$\frac{1-\beta\;\left[2\;\;a-\;\;1-\beta\;\;\underline{\theta}\;\;\overline{\theta}+\;\;1-\beta\;\;\overline{\theta}-a\;\;\underline{\theta}\right]-\sqrt{\begin{array}{l}1-\beta\;\left[2\;\;a-\;\;1-\beta\;\;\underline{\theta}\;\;\overline{\theta}-\;\;a-\;\;1-\beta\;\;\overline{\theta}\;\;\underline{\theta}\right] \\ a-\;\;1-\beta\;\;\underline{\theta}\;\;\;\overline{\theta}\;\;1-\beta\;\;-a\end{array}}}{\left[2\;\;a-\;\;1-\beta\;\;\underline{\theta}\;\;+\;\;1-\beta\;\;\overline{\theta}-a\;\;\right]\;1-\beta}$	The first positive solution to: $\kappa\;\;\hat{\theta}^B\;\;{}^5-2\underline{\theta}\kappa\;\;\hat{\theta}^B\;\;{}^4+2\overline{\theta}\underline{\theta}\kappa\;\;\hat{\theta}^B\;\;{}^3$ $-ab^2\underline{\theta}^2\;\;\overline{\theta}-2\underline{\theta}\;\;\hat{\theta}^B-2ab^2\overline{\theta}\underline{\theta}^3=0$

Figure 1: Effects of an information boundary on the firm's optimal response, $\kappa = 2$

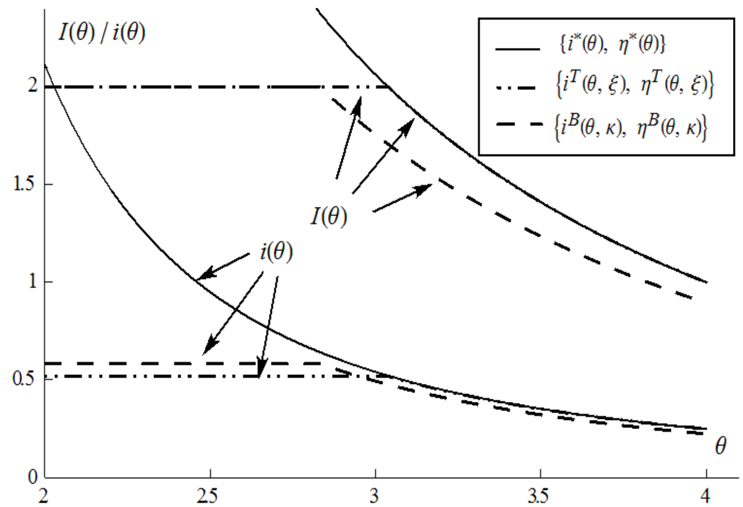


Figure 2: Effects of a preservation ratio on consumer surplus, $\beta = \frac{3}{8}\ -7 + 4\sqrt{5}$

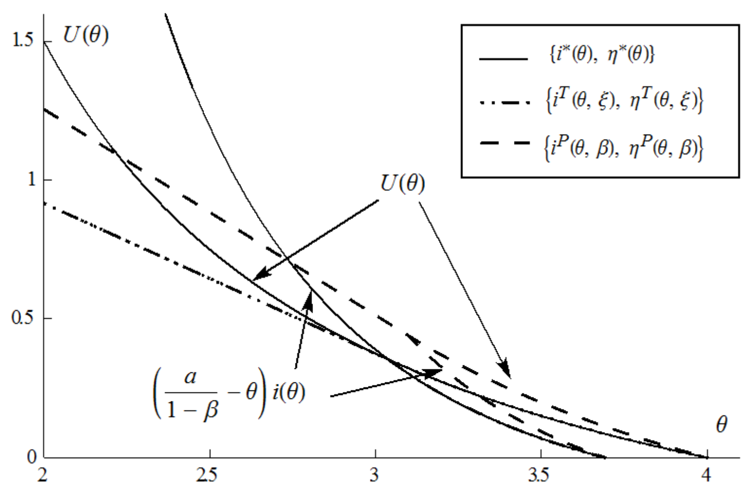


Figure 3: Effects of an information boundary on the firm's optimal response, $\kappa = 2$

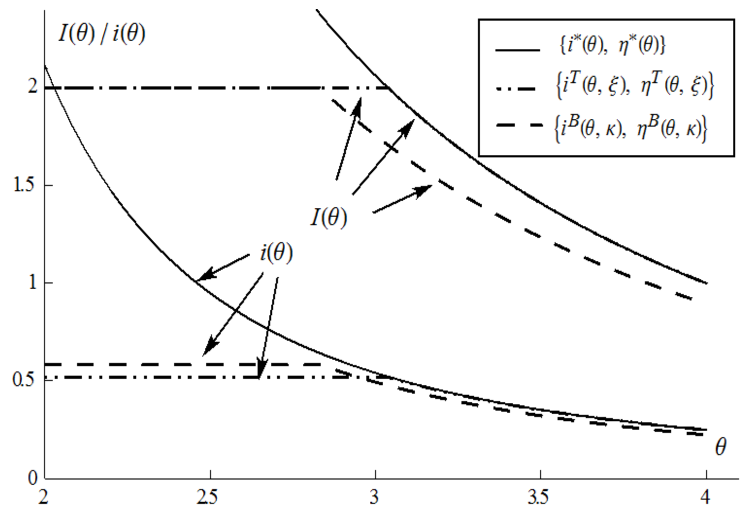


Figure 4: Effects of a preservation ratio on consumer surplus, $\beta = \frac{3}{8}\ -7+4\sqrt{5}$

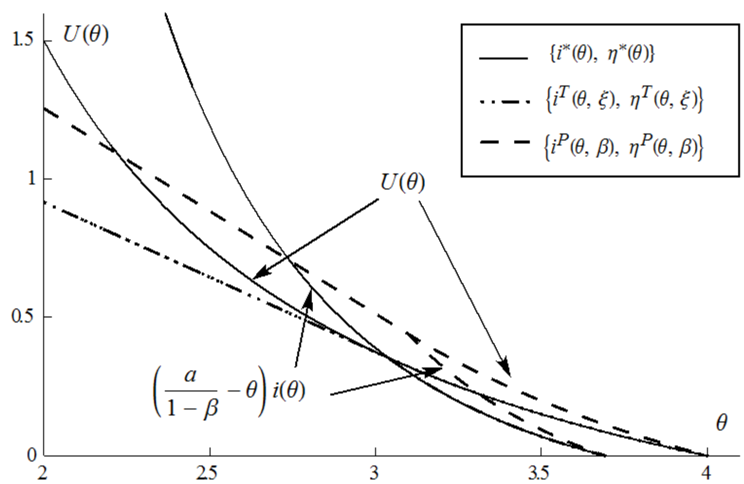


Figure 5: The amount of information subject to secondary use in the absence of regulation under maximum information acquisition ($\kappa = 2$) and minimum privacy preservation ($\beta = \frac{3}{8} \;\; -7 + 4\sqrt{5}$) policies, respectively.

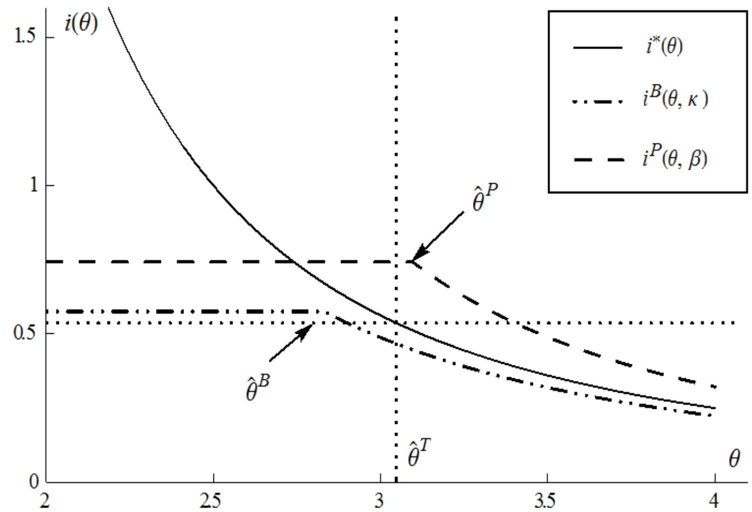


Figure 6: Privacy preservation level in the absence of regulation, under maximum information acquisition ($\kappa = 2$) and minimum privacy preservation ($\beta = \frac{3}{8} \;\; -7 + 4\sqrt{5}$) policies, respectively.

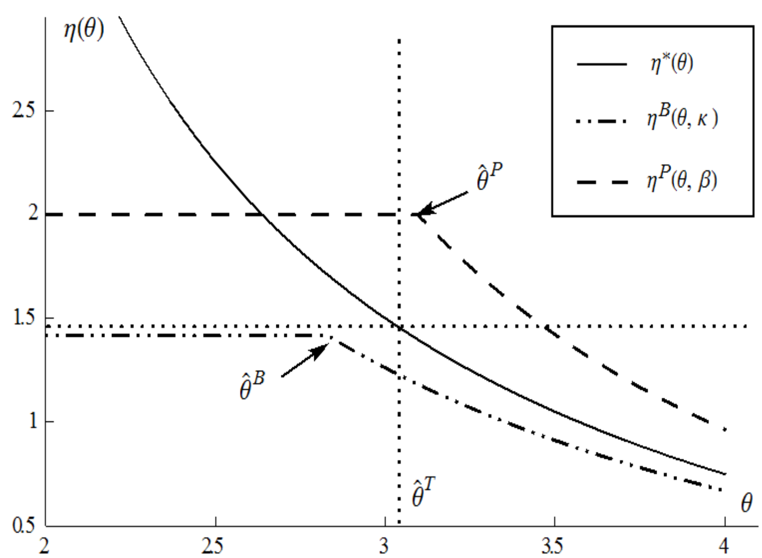


Figure 7: The ratio of information preserved from secondary uses to the total amount of information acquired under different regulatory conditions.

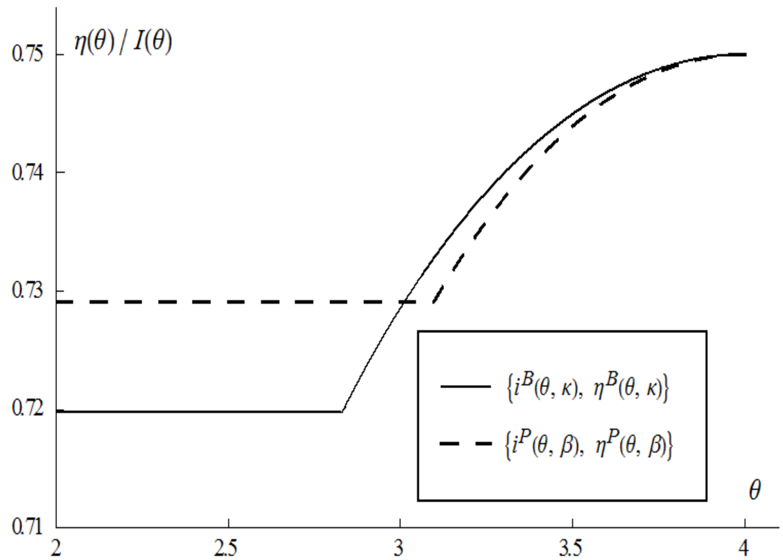


Figure 8: Firm's profit resulting from three menus under different levels of interventions

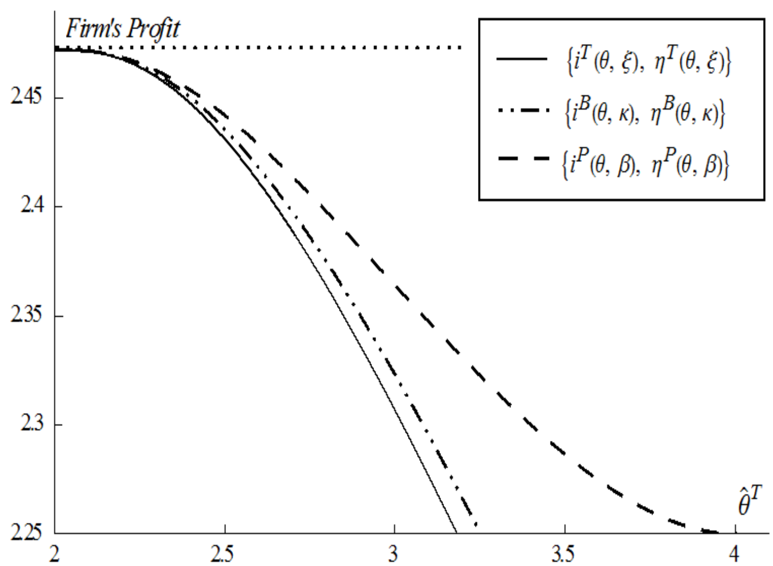
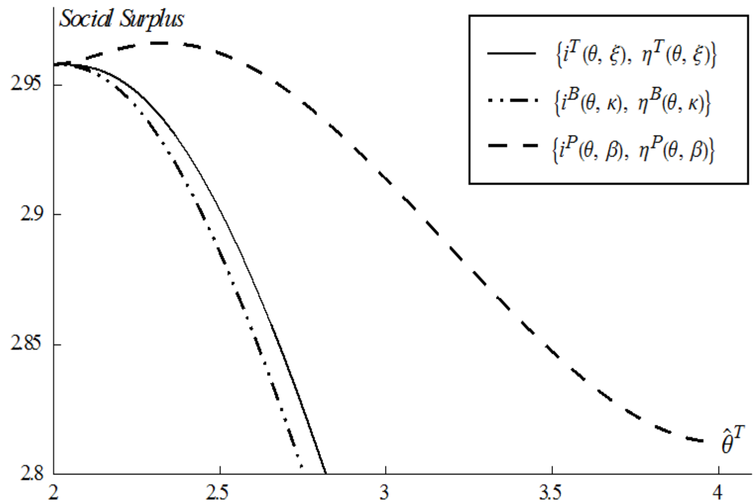


Figure 9: Social surplus resulting from three menus under different levels of interventions



Electronic Companion to the "Theory and Policy Implications of a Privacy-Preserving Contract": Theorems and Proofs

Lemma 1: An equivalent formulation of the firm's problem in (3.6) is:

$$\max_{i(\cdot)} \int_{\underline{\theta}}^{\overline{\theta}} \left[a\Phi\left(i(\theta)\right) - \theta i(\theta) - \frac{F(\theta)}{f(\theta)} i(\theta) \right] f(\theta)\, d\theta \tag{A.1}$$

subject to

$$i(\theta) \geq 0 \quad \forall \theta \in \left[\underline{\theta}, \overline{\theta}\right], \tag{A.2}$$

$$i_1(\theta) \leq 0 \quad \forall \theta \in \left[\underline{\theta}, \overline{\theta}\right], \tag{A.3}$$ ■

Proof of Lemma 1: The transformation of the objective function into a virtual surplus function is standard in literature (see Sundararajan 2004a, b). The IC implies that each consumer self-selects the contract designed for her in equilibrium. The first-order condition of maximizing $a\left(i(\tilde{\theta}) + \eta(\tilde{\theta})\right) - \theta c\left(i(\tilde{\theta})\right)$ over $\tilde{\theta}$ yields

$$a\left(i_1(\theta) + \eta_1(\theta)\right) - \theta i_1(\theta) = 0\,, \tag{A.4}$$

$\left(\left[a\left(i_1(\tilde{\theta}) + \eta_1(\tilde{\theta})\right) - \theta i_1(\tilde{\theta})\right]\Big|_{\tilde{\theta}=\theta} = 0\right.$ since $\tilde{\theta} = \theta$ is the local extremum).

The second-order condition leads to (A.3) (see Laffont and Martimont (2002), p.135 for details).

Next, we define the consumer surplus function $U(\theta) \equiv a\left(i(\theta) + \eta(\theta)\right) - \theta c\left(i(\theta)\right)$. Fully differentiating it yields: $\frac{d}{d\theta} U(\theta) = \frac{d}{d\theta}\left[a\left(i(\theta) + \eta(\theta)\right) - \theta i(\theta)\right] = a\left(i_1(\theta) + \eta_1(\theta)\right) - \theta i_1(\theta) - i(\theta) = -i(\theta)$, where the second equality results from the necessary condition (A.4). As a result, $\forall \theta \in \left[\underline{\theta}, \overline{\theta}\right]$,

$$U(\theta) = U(\overline{\theta}) + \int_{\theta}^{\overline{\theta}} i(\theta)\, dt = \int_{\theta}^{\overline{\theta}} i(\theta)\, dt \geq 0\,,$$

where $U(\overline{\theta}) = 0$. This indicates that the most privacy-sensitive consumers $\overline{\theta}$ break even from participating in the market. According to (A.4) and $U(\overline{\theta}) = 0$, it follows that

$$\eta(\theta) = \frac{1}{a}\left[\int_{\theta}^{\overline{\theta}} i(t)\, dt + \theta i(\theta)\right] - i(\theta) \tag{A.5}$$

Given this expression, we can transform the objective function as:

$$\int_{\underline{\theta}}^{\overline{\theta}}\left[\Phi(i(\theta)) - i(\theta) + \eta(\theta)\right]f(\theta)\,d\theta = \int_{\underline{\theta}}^{\overline{\theta}}\left[\Phi(i(\theta)) - \frac{1}{a}\theta i(\theta) - \frac{1}{a}\int_{\theta}^{\overline{\theta}} i(t)\,dt\right]f(\theta)\,d\theta .$$

The last term within the integral can be simplified by changing the order of integration as follows:

$$\begin{aligned}\frac{1}{a}\int_{\underline{\theta}}^{\overline{\theta}}\left[\int_{\theta}^{\overline{\theta}} c(i(t))\,dt\right]f(\theta)\,d\theta &= \frac{1}{a}\int_{\underline{\theta}}^{\overline{\theta}}\left[\int_{\theta}^{\overline{\theta}} c(i(t))\,dt\right]dF(\theta)\\ &= \frac{1}{a}\left[F(\theta)\int_{\theta}^{\overline{\theta}} c(i(t))\,dt\,\Big|_{\underline{\theta}}^{\overline{\theta}} - \int_{\underline{\theta}}^{\overline{\theta}} F(\theta)\,d\left[\int_{\theta}^{\overline{\theta}} c(i(t))\,dt\right]\right].\\ &= \frac{1}{a}\int_{\underline{\theta}}^{\overline{\theta}} F(\theta)\,c(i(\theta))\,d\theta\end{aligned}$$

Substituting the last term back into the objective function yields

$$\frac{1}{a}\int_{\underline{\theta}}^{\overline{\theta}}\left[a\Phi(i(\theta)) - \theta i(\theta) - \frac{F(\theta)}{f(\theta)} i(\theta)\right]f(\theta)\,d\theta . \blacksquare$$

Proof of Proposition 1: Maximizing (A.1) over all function $i(\cdot)$ pointwise on $\left[\underline{\theta},\overline{\theta}\right]$ gives $i^*(\theta)$ as a solution of

$$a\Phi_1(i(\theta)) = \theta + \frac{F(\theta)}{f(\theta)} \tag{A.6}$$

Fully differentiating (A.6) w.r.t. θ gives $a\Phi_{11}(i(\theta))\,i_1(\theta) = H_1(\theta)$ where $H_1(\theta) = d\left(\theta + \frac{F(\theta)}{f(\theta)}\right)/d\theta$.

Because $H_1(\theta) > 0$ (Lemma A1) and $\Phi_{11}(i) < 0$, (A.3) holds for $\forall\theta\in\left[\underline{\theta},\overline{\theta}\right]$. $\blacksquare$

Theorem 1

If $\kappa < I^*(\underline{\theta})$ in the constraint (4.1), the firm chooses to offer a menu $\left(i^B(\theta,\kappa),\eta^B(\theta,\kappa)\right)$ where

$\eta^B(\theta,\kappa) = \frac{1}{a}\left[\int_{\theta}^{\overline{\theta}} i^B(t,\kappa)\,dt + \theta i^B(\theta,\kappa)\right] - i^B(\theta,\kappa)\quad \forall\theta\in\left[\underline{\theta},\overline{\theta}\right]$ such that:

1. In the upper interval of the type space $\left[\hat{\theta}^B(\kappa),\overline{\theta}\right]$, $i^B(\theta,\kappa)$ is a solution of

$$a\Phi_1(i^B(\theta,\kappa)) = \theta + \frac{F(\theta) + \hat{\lambda}^B}{f(\theta)} \tag{A.7}$$

2. In the lower interval of the type space $\left[\underline{\theta},\hat{\theta}^B(\kappa)\right)$, $i^B(\theta,\kappa) = i^B(\hat{\theta}^B,\kappa)$.

In the above formulae, $\hat{\theta}^B(\kappa)$ and $\hat{\lambda}^B(\kappa)$ are determined by simultaneously solving $\int_{\underline{\theta}}^{\hat{\theta}^B}\left[\left(\theta+\frac{F(\theta)}{f(\theta)}\right)-\left(\hat{\theta}^B+\frac{F(\hat{\theta}^B)+\hat{\lambda}^B}{f(\hat{\theta}^B)}\right)\right]f(\theta)\,d\theta+\hat{\theta}^B\hat{\lambda}^B=0$ (A.8) and $i^B(\hat{\theta}^B,\kappa)+\eta^B(\hat{\theta}^B,\kappa)=\kappa$. (A.9). ■

Proof of Theorem 1: This proof is comprised by first conjecturing the optimal solution and then establishing it with sufficient arguments (a similar approach has been adopted by e.g., Armstrong et al. (1995)). For an incentive compatible menu $(i(\theta),\eta(\theta))$ where the participation constraint binds for the highest type (i.e., $\overline{\theta}$), if $\hat{\theta}<\overline{\theta}$ is the marginal type from which the acquisition constraint starts binding, the entire lower type space (i.e., $[\underline{\theta},\hat{\theta})$) will be bunched (i.e., $i_1(\theta)=0$). Otherwise, given (A.4) we have $I_1(\theta)=i_1(\theta)+\eta_1(\theta)=i_1(\theta)+\frac{\theta-a}{a}i_1(\theta)=\frac{\theta}{a}i_1(\theta)<0$, which implies that the constraint (4.1) is violated on $[\underline{\theta},\hat{\theta})$.

Our initial estimate is that the profit-maximizing firm leaves zero surplus to consumers with the highest privacy sensitivity; i.e. the participation constraint binds at $\overline{\theta}$ (i.e., $a\eta^B(\overline{\theta},\kappa)-(\overline{\theta}-a)i^B(\overline{\theta},\kappa)=0$). Our second conjecture is that except for bunching the lower interval of the type space due to binding of the constraint (4.1) (characterized by $\hat{\theta}^B$); the firm implements a separating menu for the remaining portion of the market. Incentive compatibility requires that an admissible $\eta^B(\theta,\kappa)$ satisfies

$$\eta^B(\theta,\kappa)=\frac{1}{a}\left[\int_{\theta}^{\overline{\theta}}i^B(t,\kappa)\,dt+\theta i^B(\theta,\kappa)\right]-i^B(\theta,\kappa) \qquad \text{(A.10)}$$

which is derived from integrating (A.4) over $[\theta,\overline{\theta}]$ for $\theta\geq\hat{\theta}^B$ and using the participation constraint to prescribe the upper limit of the integral $\eta^B(\overline{\theta},\kappa)$.

Suppose there exists a function $\mu(\theta)$ for all θ, designated as the multiplier of $i_1(\theta)\leq 0$, and a function $\nu(\theta)$ for all θ, designated as the multiplier of $\eta_1(\theta)=\frac{\theta-a}{a}i_1(\theta)$ (from equation (A.4)) (see Laffont and Martimort (2002), p.140-142; there, the optimization problem over $q(\theta)$ and $U(\theta)$ is transformed into a single control-variable problem $q_1(\theta)$ by applying the concept of information rents. In our case, we keep both $i(\theta)$ and $\eta(\theta)$, since the implication of constraint (4.1) on information rents is hardly told ex ante), then the Hamiltonian is defined as

$$H(i,\eta,i_1,\mu,\nu,\theta) = \left[\Phi(i(\theta)) - (i(\theta)+\eta(\theta))\right] f(\theta) + \mu(\theta)\, i_1(\theta) + \nu(\theta)\frac{\theta-a}{a} i_1(\theta) \qquad \text{(A.11)}$$

Suppose we can further find a non-decreasing function $\lambda^B(\theta)$ for all θ such that $\lambda^B(\theta) = \hat{\lambda}^B$ for $\theta > \hat{\theta}^B$, and a non-decreasing function $\gamma(\theta)$ for all θ such that $\gamma_1(\theta) = 0$ for $\theta < \bar{\theta}$ (given $a(i(\theta)+\eta(\theta)) - \theta i(\theta) > 0$). We can define the Lagrangian (Seierstad and Sydsæter (1987)'s theorem 3, p.336) as

$$L(i,\eta,i_1,\mu,\nu,\theta,\lambda^P,\gamma) = H(i,\eta,i_1,\mu,\nu,\theta) + \lambda^B(\theta)\left[i_1(\theta) + \frac{\theta-a}{a} i_1(\theta)\right] + \gamma(\theta)\, i(\theta)$$

where $\mu(\theta) - \lambda^B(\theta) + (a-\theta)\gamma(\theta)$ and $\nu(\theta) - \lambda^B(\theta) + a\gamma(\theta)$ are continuous.

From the Pontryagin principle, we have

$-(\mu_1(\theta) - \lambda_1^B(\theta) - \gamma(\theta) + (a-\theta)\gamma_1(\theta)) = \dfrac{\partial L}{\partial i} = \left[\Phi_1(i(\theta)) - 1\right] f(\theta) + \gamma(\theta)$,

and $-(\nu_1(\theta) - \lambda_1^B(\theta) + a\gamma_1(\theta)) = \dfrac{\partial L}{\partial \eta} = -f(\theta)$.

without loss of generality we set $\lambda^B(\underline{\theta}) = \gamma(\underline{\theta}) = 0$ (see Note 3 on p.333 Seierstad and Sydsæter (1987)). Optimality requires $\mu(\underline{\theta}) = \nu(\underline{\theta}) = \mu(\bar{\theta}) = \nu(\bar{\theta}) = 0$. Integrating the two equations and setting $\mu(\underline{\theta}) = \nu(\underline{\theta}) = 0$ yields

$$\mu(\theta) - \lambda^B(\theta) + (a-\theta)\gamma(\theta) = \int_{\underline{\theta}}^{\theta} \left[(1-\Phi_1(i(t)))f(t) - \gamma(t)\right] dt \qquad \text{(A.12)}$$

$$\text{and } \nu(\theta) - \lambda^B(\theta) + a\gamma(\theta) = F(\theta) \qquad \text{(A.13)}$$

Maximizing $H(i,\eta,i_1(\theta),\mu(\theta),\nu(\theta),\theta)$ w.r.t. $i_1(\cdot) \le 0$ requires $\mu(\theta) + \nu(\theta)\dfrac{\theta-a}{a} \ge 0$ with $i_1(\theta) = 0$ if $\mu(\theta) + \nu(\theta)\dfrac{\theta-a}{a} > 0$.

Also note that $\max_{i_1(\cdot)} H(i,\eta,i_1(\theta),\mu(\theta),\nu(\theta),\theta) = \left[\Phi(i(\theta)) - (i(\theta)+\eta(\theta))\right] f(\theta)$ is concave for v.e. θ (the Hessian matrix of $\max_{i_1(\cdot)} H(i,\eta,i_1(\theta),\mu(\theta),\nu(\theta),\theta)$ is $\begin{vmatrix} \Phi_{11}(i) f(\theta) & 0 \\ 0 & 0 \end{vmatrix} \le 0$). Moreover, $\kappa - (i(\theta)+\eta(\theta))$ and $a(i(\theta)+\eta(\theta)) - \theta i(\theta)$ are quasi-concave in i and η.

Suppose $i_1(\theta) < 0$ on the interval $\left[\hat{\theta}^P, \bar{\theta}\right]$, then $\mu(\theta) + \nu(\theta)\dfrac{\theta - a}{a} = 0$ on this interval.

Substituting $\mu(\theta)$ and $\nu(\theta)$ with (A.12) and (A.13) into $\mu(\theta) + \nu(\theta)\dfrac{\theta - a}{a}$ yields

$$
\begin{aligned}
&\mu(\theta) + \nu(\theta)\frac{\theta - a}{a} \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.14)}\\
&= \int_{\underline{\theta}}^{\theta}\left[\left(1 - \Phi_1(i(t))\right) f(t) - \gamma(t)\right]dt + \lambda^B(\theta) - (a - \theta)\gamma(\theta)\\
&+ \frac{\theta - a}{a}\left[F(\theta) + \lambda^B(\theta) - a\gamma(\theta)\right]\\
&= \int_{\underline{\theta}}^{\theta}\left[-\Phi_1(i(t)) f(t) - \gamma(t)\right]dt + \frac{\theta}{a}F(\theta) + \frac{\theta}{a}\lambda^B(\theta)\\
&= \int_{\underline{\theta}}^{\theta}\left[\left(\frac{1}{a}\left(t + \frac{F(t)}{f(t)}\right) - \Phi_1(i(t))\right) f(t) - \gamma(t)\right]dt + \frac{\theta}{a}\lambda^B(\theta)
\end{aligned}
$$

The third equality follows integration by parts.

Since $\lambda^B(\theta)$ and $\gamma(\theta)$ is continuous at $\hat{\theta}^B$, the multipliers $\mu(\theta)$ and $\nu(\theta)$ are also continuous. We thereby have $\mu(\hat{\theta}^B) + \nu(\hat{\theta}^B)\dfrac{\hat{\theta}^B - a}{a} = 0$. Consequently,

$$
\begin{aligned}
&\mu(\theta) + \nu(\theta)\frac{\theta - a}{a} - \left(\mu(\hat{\theta}^B) + \nu(\hat{\theta}^B)\frac{\hat{\theta}^B - a}{a}\right)\\
&= \int_{\underline{\theta}}^{\theta}\left[\left(\frac{1}{a}\left(t + \frac{F(t)}{f(t)}\right) - \Phi_1(i(t))\right) f(t) - \gamma(t)\right]dt + \frac{\theta}{a}\lambda^B(\theta) -\\
&- \int_{\underline{\theta}}^{\hat{\theta}^B}\left[\left(\frac{1}{a}\left(t + \frac{F(t)}{f(t)}\right) - \Phi_1(i(t))\right) f(t) - \gamma(t)\right]dt - \frac{\hat{\theta}^B}{a}\lambda^B(\hat{\theta}^B)\\
&= \int_{\hat{\theta}^B}^{\theta}\left[\frac{1}{a}\left(t + \frac{F(t)}{f(t)}\right) - \Phi_1(i(t))\right] f(t)\,dt + \frac{\theta}{a}\lambda^B(\theta) - \frac{\hat{\theta}^B}{a}\lambda^B(\hat{\theta}^B)\\
&= \int_{\hat{\theta}^P}^{\theta}\left[\frac{1}{a}\left(t + \frac{F(t) + \hat{\lambda}^B}{f(t)}\right) - \Phi_1(i(t))\right] f(t)\,dt = 0
\end{aligned}
$$

The last equality follows $\lambda^B(\theta) = \hat{\lambda}^B$ for $\theta > \hat{\theta}^B$ and $\lambda^B(\theta)$ is continuous at $\hat{\theta}^B$.

Hence $\int_{\hat{\theta}^P}^{\theta}\left[\dfrac{1}{a}\left(t + \dfrac{F(t) + \hat{\lambda}^B}{f(t)}\right) - \Phi_1(i(t))\right] f(t)\,dt = 0$ for $\forall\theta \in \left[\hat{\theta}^P, \bar{\theta}\right]$, implying (A.7) in Theorem 1. Specifically, because $\mu(\hat{\theta}^B) + \nu(\hat{\theta}^B)\dfrac{\hat{\theta}^B - a}{a} = 0$ at $\hat{\theta}^B$ and $\gamma(\theta) = 0$ for $\forall\theta \in [\underline{\theta}, \bar{\theta})$, from (A.14) it follows $\int_{\underline{\theta}}^{\hat{\theta}^B}\left(\dfrac{1}{a}\left(\theta + \dfrac{F(\theta)}{f(\theta)}\right) - \Phi_1(i(\theta))\right) f(\theta)\,d\theta + \dfrac{\hat{\theta}^B}{a}\lambda^B(\hat{\theta}^B) = 0$.

By definition of bunching, $i(\theta)$ is constant on $[\underline{\theta}, \hat{\theta}^B)$. From the continuity of the contract menu, the constant level of commercial use is also the solution to equation (A.7) at $\hat{\theta}^B$. Substituting $\Phi_1(i(\theta))$

with $\frac{1}{a}\left[\hat{\theta}^B + \frac{F(\hat{\theta}^B)+\hat{\lambda}^B}{f(\hat{\theta}^B)}\right]$ into the above equation and multiplying both sides of by a result (A.8). Equation (A.9) reflects the binding of constraint (4.1) at $\hat{\theta}^B$.

$\theta + \frac{F(\theta)+\hat{\lambda}^B}{f(\theta)}$ in equation (A.7) is increasing on $\left[\hat{\theta}^B,\overline{\theta}\right]$, the establishment of which relies on Lemma A2 and the fact that $\hat{\theta}^B$ lies on the increasing region of $\theta + \frac{F(\theta)+\hat{\lambda}^B}{f(\theta)}$ from equation (A.8). We need to then identify a $\gamma(\overline{\theta}) \geq 0$ (so that our conjecture that the participation constraint binds at $\overline{\theta}$ is correct) such that $\mu(\overline{\theta}) = \nu(\overline{\theta}) = 0$ and $\lambda^B(\theta)$ is non-decreasing on $\left[\underline{\theta},\hat{\theta}^B\right]$. Rearranging (A.8) yields

$$\hat{\lambda}^B = \frac{F(\hat{\theta}^B)^2}{f(\hat{\theta}^B)\hat{\theta}^B - F(\hat{\theta}^B)} \tag{A.15}$$

From (A.13) and $\nu(\overline{\theta}) = 0$, $a\gamma(\overline{\theta}) - \hat{\lambda}^B = 1$, which implies that

$$a\gamma(\overline{\theta}) = 1 + \hat{\lambda}^B > 0 \tag{A.16}$$

From (A.12), $\mu(\overline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}}\left(1-\Phi_1(i(\theta))\right)f(\theta)\,d\theta + \hat{\lambda}^B - (a-\overline{\theta})\gamma(\overline{\theta})$, such that

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}} \Phi_1(i(\theta))f(\theta)\,d\theta \\
&= \int_{\underline{\theta}}^{\hat{\theta}^B} \Phi_1(i(\theta))f(\theta)\,d\theta + \int_{\hat{\theta}^B}^{\overline{\theta}} \Phi_1(i(\theta))f(\theta)\,d\theta \\
&= \frac{1}{a}\int_{\underline{\theta}}^{\hat{\theta}^B}\left(\hat{\theta}^P + \frac{F(\hat{\theta}^B)+\hat{\lambda}^B}{f(\hat{\theta}^B)}\right)f(\theta)\,d\theta + \frac{1}{a}\int_{\hat{\theta}^B}^{\overline{\theta}}\left(\theta + \frac{F(\theta)+\hat{\lambda}^B}{f(\theta)}\right)f(\theta)\,d\theta. \\
&= \frac{1}{a}\left[\left(\hat{\theta}^P + \frac{F(\hat{\theta}^B)+\hat{\lambda}^B}{f(\hat{\theta}^B)}\right)F(\hat{\theta}^B) + \overline{\theta} - \hat{\theta}^B F(\hat{\theta}^B) + (\overline{\theta} - \hat{\theta}^P)\hat{\lambda}^B\right] \\
&= \frac{1}{a}\left[\frac{F(\hat{\theta}^B)+\hat{\lambda}^B}{f(\hat{\theta}^B)}F(\hat{\theta}^B) + \overline{\theta} + (\overline{\theta} - \hat{\theta}^P)\hat{\lambda}^B\right]
\end{aligned}$$

Substituting (A.15) and (A.16) in to the above equation, $\mu(\overline{\theta}) = 0$. The proof is complete. ■

Proof of Proposition 2: Part 1 of Proposition 2 is the direct result of Theorem 1. On $\left[\hat{\theta}^B(\kappa),\overline{\theta}\right]$, $a\Phi_1(i^B(\theta,\kappa)) = \theta + \frac{F(\theta)+\hat{\lambda}^B}{f(\theta)}$, and $a\Phi_1(i^*(\theta)) = \theta + \frac{F(\theta)}{f(\theta)}$; given $\hat{\lambda}^B > 0$ and $\Phi_{11}(i) < 0$, $i^B(\theta,\kappa) < i^*(\theta)$; on $\left[\underline{\theta},\hat{\theta}^B(\kappa)\right)$, $i^B(\theta,\kappa) = i^B(\hat{\theta}^B,\kappa)$ but $i^*(\theta) \geq i^*(\hat{\theta}^B)$, therefore $i^B(\theta,\kappa) < i^*(\theta)$.

Equations (A.8) and (A.9) simultaneously determine $\hat\theta^B$ and $\hat\lambda^B$, both of which are functions of the preservation level, κ. According to Theorem 1, $i^B(\theta,\kappa)$ is determined by equation (A.7) on the separating interval $\left[\hat\theta(\beta),\bar\theta\right]$. Fully differentiating (A.7) w.r.t. κ gives

$$a\Phi_{11}(i^B(\theta,\kappa))\, i_2^B(\theta,\kappa) = \frac{\hat\lambda_1^B(\kappa)}{f(\theta)} \qquad \text{(A.17)}$$

Similarly, $i_1^P(\theta,\beta)$ satisfies

$$a\Phi_{11}(i^B(\theta,\kappa))\, i_1^B(\theta,\kappa) = H_1^B(\theta) \qquad \text{(A.18)},$$

where $H_1^B(\theta) \equiv d\left(\theta + \frac{F(\theta)+\hat\lambda^B}{f(\theta)}\right)/d\theta$.

Fully differentiating (A.8) and (A.9) w.r.t. κ yields

$$\left[-H_1^B(\hat\theta^B)\,\hat\theta_1^B(\kappa) - \frac{\hat\lambda_1^B(\kappa)}{f(\hat\theta^B)}\right]F(\hat\theta^B) + \hat\theta^B(\kappa)\,\hat\lambda_1^B(\kappa) = 0\,, \text{(A.19)}$$

$$\text{and } \int_{\hat\theta^B}^{\bar\theta} i_2^B(t,\kappa)\,dt + \left[i_2^B(\hat\theta^B,\kappa) + i_1^B(\hat\theta^B,\kappa)\,\hat\theta_1^B(\kappa)\right]\hat\theta^B(\kappa) = a\,, \text{(A.20)}$$

respectively.

Substituting (A.17) and (A.18) into (A.20) yields

$$\int_{\hat\theta^B}^{\bar\theta}\frac{\hat\lambda_1^B(\kappa)}{\Phi_{11}(i^B(t,\kappa))\,f(t)}\,dt + \left[\frac{\hat\lambda_1^B(\kappa)}{\Phi_{11}(i^B(\hat\theta^B,\kappa))\,f(\hat\theta^B)} + \frac{H_1^B(\hat\theta^B)}{\Phi_{11}(i^B(\hat\theta^B,\kappa))}\hat\theta_1^B(\kappa)\right]\hat\theta^B(\kappa) = a^2 \qquad \text{(A.21)}$$

Reorganizing (A.19), we have

$$\hat\lambda_1^B(\kappa) = \frac{H_1^B(\hat\theta^B)\,\hat\theta_1^B(\kappa)\,F(\hat\theta^B)}{\hat\theta^B(\kappa) - \frac{F(\hat\theta^B)}{f(\hat\theta^B)}} \qquad \text{(A.22)}$$

To verify whether this first order derivative is well-defined, we integrate (A.8) by parts, which yields

$$\hat\theta^B F(\hat\theta^B) - \left(\hat\theta^B + \frac{F(\hat\theta^B)+\hat\lambda^B}{f(\hat\theta^B)}\right)F(\hat\theta^B) + \hat\theta^B(\kappa)\,\hat\lambda^B(\kappa) = 0$$

which leads to $\hat{\lambda}^B(\kappa) = \dfrac{F(\hat{\theta}^B)}{f(\hat{\theta}^B)} F(\hat{\theta}^B) \Big/ \left(\hat{\theta}^B - \dfrac{F(\hat{\theta}^B)}{f(\hat{\theta}^B)}\right)$. By definition, $\hat{\lambda}^B(\kappa) > 0$. Therefore, the solution of $\hat{\theta}^B$ satisfies $\hat{\theta}^B(\kappa) - \dfrac{F(\hat{\theta}^B)}{f(\hat{\theta}^B)} > 0$. Consequently, $\hat{\lambda}_1^B(\kappa)$ is well defined. Substituting (A.22) into (A.21) yields

$$\left[\int_{\hat{\theta}^B}^{\bar{\theta}} \frac{F(\hat{\theta}^B)}{\Phi_{11}(i^B(t,\kappa))f(t)} dt + \frac{\left[\hat{\theta}^B(\kappa)\right]^2}{\Phi_{11}(i^B(\hat{\theta}^B,\kappa))}\right]\hat{\theta}_1^B(\kappa) = \frac{a^2\left(\hat{\theta}^B(\kappa) - \dfrac{F(\hat{\theta}^B)}{f(\hat{\theta}^B)}\right)}{H_1^B(\hat{\theta}^B)}$$

$\hat{\theta}_1^B(\kappa) < 0$ follows $\hat{\theta}^B(\kappa) - \dfrac{F(\hat{\theta}^B)}{f(\hat{\theta}^B)} > 0$, $H_1^B(\hat{\theta}^B) > 0$ (see Lemma A2), and $\Phi_{11}(i) < 0$. According to (A.22), $\hat{\lambda}_1^B(\kappa) < 0$.

$\dfrac{d}{d\theta}(\theta f(\theta) - F(\theta)) = \theta f_1(\theta)$ implies that the monotonicity of $\theta f(\theta) - F(\theta)$ is the same as that of $f(\theta)$: it is first monotonically increasing, and, beyond a tipping point, is monotonically decreasing (a log-concave function is always unimodal (An 1998)). Since $\underline{\theta} f(\underline{\theta}) - F(\underline{\theta}) > 0$, $\theta f(\theta) - F(\theta) = 0$ can only occur on the decreasing interval of $f(\theta)$.

If $\theta f(\theta) - F(\theta) > 0$ on the entire type space, $\varnothing = \left\{\theta : \theta - \dfrac{F(\theta)}{f(\theta)} = 0\right\}$, and the upper bound of $\hat{\theta}^B(\kappa)$ is $\bar{\theta}$. Otherwise, by unimodality of $f(\theta)$, $\theta f(\theta) - F(\theta) \le 0$ for $\theta \in [\tilde{\theta}, \bar{\theta}]$. It implies that $\hat{\theta}^B(\kappa) \in [\tilde{\theta}, \bar{\theta}]$ cannot be the solution to equation (A.8) and equation (A.9). ■

Proof of Proposition 3: For a partially-bunching menu, equation (4.2) can be represented by

$$\begin{aligned} SW(\kappa) &= \int_{\underline{\theta}}^{\hat{\theta}^B}\left[\Phi(i^B(\hat{\theta}^B,\kappa)) + (\alpha a - 1)i^B(\hat{\theta}^B,\kappa) + \eta^B(\hat{\theta}^B,\kappa) - \alpha\theta i^B(\hat{\theta}^B,\kappa)\right] f(\theta)\, d\theta \\ &+ \int_{\hat{\theta}^B}^{\bar{\theta}}\left[\Phi(i^B(\theta,\kappa)) + (\alpha a - 1)i^B(\theta,\kappa) + \eta^B(\theta,\kappa) - \alpha\theta i^B(\theta,\kappa)\right] f(\theta)\, d\theta \end{aligned}$$

Substituting (A.10) into the above equation produces

$$
\begin{aligned}
SW(\kappa) &= \int_{\underline{\theta}}^{\hat{\theta}^B}\left[\Phi\left(i^B\left(\hat{\theta}^B,\kappa\right)\right)+\left(\alpha-\frac{1}{a}\right)\int_{\hat{\theta}^B}^{\bar{\theta}} i^B(t,\kappa)\,dt+\hat{\theta}^B i^B\left(\hat{\theta}^B,\kappa\right)-\alpha\theta i^B\left(\hat{\theta}^B,\kappa\right)\right]f(\theta)\,d\theta \\
&+\int_{\hat{\theta}^B}^{\bar{\theta}}\left[\Phi\left(i^B(\theta,\kappa)\right)+\left(\alpha-\frac{1}{a}\right)\int_{\theta}^{\bar{\theta}} i^B(t,\kappa)\,dt+\theta i^B(\theta,\kappa)-\alpha\theta i^B(\theta,\kappa)\right]f(\theta)\,d\theta \\
&=\int_{\underline{\theta}}^{\hat{\theta}^B}\left[\Phi\left(i^B\left(\hat{\theta}^B,\kappa\right)\right)+\left(\alpha-\frac{1}{a}\right)\int_{\hat{\theta}^B}^{\bar{\theta}} i^B(t,\kappa)\,dt+\left(\alpha\left(\hat{\theta}^B-\theta\right)-\frac{\hat{\theta}^B}{a}\right)i^B\left(\hat{\theta}^B,\kappa\right)\right]f(\theta)\,d\theta \\
&+\int_{\hat{\theta}^B}^{\bar{\theta}}\left[\Phi\left(i^B(\theta,\kappa)\right)-\frac{\theta}{a}i^B(\theta,\kappa)+\left(\alpha-\frac{1}{a}\right)\int_{\theta}^{\bar{\theta}} i^B(t,\kappa)\,dt\right]f(\theta)\,d\theta
\end{aligned}
$$

Fully differentiating $SW(\kappa)$ w.r.t. κ yields

$$
\begin{aligned}
\frac{d}{d\kappa}SW(\kappa) &= \int_{\underline{\theta}}^{\hat{\theta}^B}\left[\begin{array}{l}\Phi_1\left(i^B\left(\hat{\theta}^B,\kappa\right)\right)\left(i_1^B\left(\hat{\theta}^B,\kappa\right)\hat{\theta}_1^B(\kappa)+i_2^B\left(\hat{\theta}^B,\kappa\right)\right)\\ +\left(\alpha-\dfrac{1}{a}\right)\displaystyle\int_{\hat{\theta}^B}^{\bar{\theta}} i_2^B(t,\kappa)\,dt \\ +\left(\alpha\left(\hat{\theta}^B-\theta\right)-\dfrac{\hat{\theta}^B}{a}\right)\left(i_1^B\left(\hat{\theta}^B,\kappa\right)\hat{\theta}_1^B(\kappa)+i_2^B\left(\hat{\theta}^B,\kappa\right)\right)\end{array}\right]f(\theta)\,d\theta \\
&+\int_{\hat{\theta}^B}^{\bar{\theta}}\left[\Phi_1\left(i^B(\theta,\kappa)\right)i_2^B(\theta,\kappa)-\frac{\theta}{a}i_2^B(\theta,\kappa)+\left(\alpha-\frac{1}{a}\right)\int_{\theta}^{\bar{\theta}} i_2^B(t,\kappa)\,dt\right]f(\theta)\,d\theta \\
&=\int_{\underline{\theta}}^{\hat{\theta}^B}\left(\Phi_1\left(i^B\left(\hat{\theta}^B,\kappa\right)\right)+\alpha\left(\hat{\theta}^B-\theta\right)-\frac{\hat{\theta}^B}{a}\right)\left(i_1^B\left(\hat{\theta}^B,\kappa\right)\hat{\theta}_1^B(\kappa)+i_2^B\left(\hat{\theta}^B,\kappa\right)\right)f(\theta)\,d\theta \\
&+\left(\alpha-\frac{1}{a}\right)F\left(\hat{\theta}^B\right)\int_{\hat{\theta}^B}^{\bar{\theta}} i_2^B(t,\kappa)\,dt+\left(\alpha-\frac{1}{a}\right)\left[-F\left(\hat{\theta}^B\right)\int_{\hat{\theta}^B}^{\bar{\theta}} i_2^B(t,\kappa)\,dt+\int_{\hat{\theta}^B}^{\bar{\theta}} F(\theta)\,i_2^B(\theta,\kappa)\,d\theta\right] \\
&+\int_{\hat{\theta}^B}^{\bar{\theta}}\left(\Phi_1\left(i^B(\theta,\kappa)\right)-\frac{\theta}{a}\right)i_2^B(\theta,\kappa)f(\theta)\,d\theta \\
&=\int_{\underline{\theta}}^{\hat{\theta}^B}\left(\Phi_1\left(i^B\left(\hat{\theta}^B,\kappa\right)\right)+\alpha\left(\hat{\theta}^B-\theta\right)-\frac{\hat{\theta}^B}{a}\right)\left(i_1^B\left(\hat{\theta}^B,\kappa\right)\hat{\theta}_1^B(\kappa)+i_2^B\left(\hat{\theta}^B,\kappa\right)\right)f(\theta)\,d\theta \\
&+\left(\alpha-\frac{1}{a}\right)\int_{\hat{\theta}^B}^{\bar{\theta}} F(\theta)\,i_2^B(\theta,\kappa)\,d\theta \\
&+\int_{\hat{\theta}^B}^{\bar{\theta}}\left(\Phi_1\left(i^B(\theta,\kappa)\right)-\frac{\theta}{a}\right)i_2^B(\theta,\kappa)f(\theta)\,d\theta .
\end{aligned}
$$

The second equality is a direct result from integration by parts.

Substituting (A.7) into the above equation yields

$$
\begin{aligned}
\frac{d}{d\kappa}SW(\kappa) &= \int_{\underline{\theta}}^{\hat{\theta}^B}\left(\Phi_1\left(i^B\left(\hat{\theta}^B,\kappa\right)\right)+\alpha\left(\hat{\theta}^B-\theta\right)-\frac{\hat{\theta}^B}{a}\right)\left(i_1^B\left(\hat{\theta}^B,\kappa\right)\hat{\theta}_1^B(\kappa)+i_2^B\left(\hat{\theta}^B,\kappa\right)\right)f(\theta)\,d\theta \\
&+\left(\alpha-\frac{1}{a}\right)\int_{\hat{\theta}^B}^{\bar{\theta}} F(\theta)\,i_2^B(\theta,\kappa)\,d\theta+\int_{\hat{\theta}^B}^{\bar{\theta}}\left(\Phi_1\left(i^B(\theta,\kappa)\right)-\frac{\theta}{a}\right)i_2^B(\theta,\kappa)f(\theta)\,d\theta \\
&=\int_{\underline{\theta}}^{\hat{\theta}^B}\left(\frac{1}{a}\frac{F\left(\hat{\theta}^B\right)+\hat{\lambda}^B}{f\left(\hat{\theta}^B\right)}+\alpha\left(\hat{\theta}^B-\theta\right)\right)\left(i_1^B\left(\hat{\theta}^B,\kappa\right)\hat{\theta}_1^B(\kappa)+i_2^B\left(\hat{\theta}^B,\kappa\right)\right)f(\theta)\,d\theta \\
&+\int_{\hat{\theta}^B}^{\bar{\theta}}\left(\frac{\hat{\lambda}^B}{a}+\alpha F(\theta)\right)i_2^B(\theta,\kappa)\,d\theta .
\end{aligned}
$$

From Proposition 2, we know $\hat{\theta}_1^B(\kappa) < 0$ and $\hat{\lambda}_1^B(\kappa) < 0$ (from equation (A.22)). According to (A.17) and (A.18), $i_1^B(\theta,\kappa) < 0$ and $i_2^B(\theta,\kappa) > 0$. Hence, we have $\frac{d}{d\kappa}SW(\kappa) > 0$. This result is independent of the value of α. ■

Lemma 2. For an incentive compatible menu $(i(\theta),\eta(\theta))$, the participation constraint binds for the highest type (i.e., $\bar{\theta}$), $\Delta(\theta) \equiv \frac{\eta(\theta)}{i(\theta)+\eta(\theta)}$ is increasing on $\{\theta \mid i_1(\theta) < 0\}$ on $[\underline{\theta},\bar{\theta})$. ■

Proof of Lemma 2:

$$
\begin{aligned}
\frac{d}{d\theta}\Delta(\theta) &= \frac{d}{d\theta}\left[\frac{\eta(\theta)}{i(\theta)+\eta(\theta)}\right] \\
&= \frac{\eta_1(\theta)\left(i(\theta)+\eta(\theta)\right)-\eta(\theta)\left(\eta_1(\theta)+i_1(\theta)\right)}{\left(i(\theta)+\eta(\theta)\right)^2} \\
&= \frac{\eta_1(\theta)i(\theta)-\eta(\theta)i_1(\theta)}{\left(i(\theta)+\eta(\theta)\right)^2} \\
&= \frac{\frac{\theta-a}{a}i(\theta)-\eta(\theta)}{\left(i(\theta)+\eta(\theta)\right)^2}i_1(\theta)
\end{aligned}
.
$$

The last equality results from condition (A.4), which is necessary for incentive compatibility.

Since the participation constraint is binding for the market highest type, i.e., $U(\bar{\theta}) = 0$, substituting (A.5) into the above formula yields

$$
\frac{d}{d\theta}\Delta(\theta) = \frac{-\int_{\theta}^{\bar{\theta}} i(t)\,dt}{\left(i(\theta)+\eta(\theta)\right)^2}i_1(\theta)
$$

$\frac{d}{d\theta}\Delta(\theta) > 0$ for all $\theta \in [\underline{\theta},\bar{\theta})$ and $\theta \in \{\theta \mid i_1(\theta) < 0\}$ follows from (A.2); moreover, $\frac{d}{d\theta}\Delta(\theta) = 0$ for all $\theta \in [\underline{\theta},\bar{\theta})$ and $\theta \in \{\theta \mid i_1(\theta) = 0\}$. ■

Theorem 2

If $\Delta^*(\underline{\theta}) < \beta \leq (\bar{\theta}-a)/\bar{\theta}$ in constraint (4.3), the profit-maximizing firm chooses to offer a menu $(i^P(\theta,\beta),\eta^P(\theta,\beta))$ where $\eta^P(\theta,\beta) = \frac{1}{a}\left[\int_{\theta}^{\bar{\theta}} i^P(t,\beta)\,dt + \theta i^P(\theta,\beta)\right] - i^P(\theta,\beta)$ $\forall\theta \in [\underline{\theta},\bar{\theta}]$ such that:

1. In the upper interval of the type space $\left[\hat{\theta}^P(\beta),\bar{\theta}\right]$, $i^P(\theta,\beta)$ is a solution of

$$a\Phi_1 \ \ i^P \ \ \theta,\beta \ \ = \theta + \frac{F \ \ \theta \ \ - \ \ 1-\beta \ \ \hat{\lambda}^P}{f \ \ \theta} ; \qquad \text{(A.23)}$$

2. In the lower interval of the type space $\left[\underline{\theta},\hat{\theta}^P \ \ \beta \ \ \ , i^P \ \ \theta,\beta \ \ = i^P \ \ \hat{\theta}^P,\beta \ \ .$

In all the above formulae, $\hat{\theta}^P \ \ \beta$ and $\hat{\lambda}^P \ \ \beta$ are determined by simultaneously solving

$$\int_{\underline{\theta}}^{\hat{\theta}^P} \left[\left(\theta + \frac{F \ \ \theta}{f \ \ \theta} \right) - \left(\hat{\theta}^P + \frac{F \ \ \hat{\theta}^P \ \ - \ \ 1-\beta \ \ \hat{\lambda}^P}{f \ \ \hat{\theta}^P} \right) \right] f \ \ \theta \ \ d\theta \qquad \text{(A.24)}$$
$$+ \ \ a - \hat{\theta}^P \ \ 1-\beta \ \ \ \hat{\lambda}^P = 0$$

and $1-\beta \ \ \eta^P \ \ \hat{\theta}^P,\beta \ \ = \beta i^P \ \ \hat{\theta}^P,\beta \ \ .$ (A.25) ■

Proof of Theorem 2: According to lemma 2, for any incentive compatible menu where the participation constraint binds for the highest type $\bar{\theta}$, $\frac{d}{d\theta}\Delta \ \ \theta \ \ \geq 0$.

Suppose that $\hat{\theta} < \bar{\theta}$ is the highest type from which the preservation constraint (4.3) starts binding. If $i_1 \ \ \theta \ \ < 0$ on $\left[\underline{\theta},\hat{\theta}\right.$, $\frac{d}{d\theta}\Delta \ \ \theta \ \ > 0$, which means that constraint (4.3) would be violated. Hence $i_1 \ \ \theta \ \ = 0$ on $\left[\underline{\theta},\hat{\theta}\right.$.

When $\beta > \frac{\bar{\theta} - a}{\bar{\theta}}$, (4.3) implies that the firm cannot reduce consumer surplus of any type to the zero level. Then $U \ \ \theta \ \ \geq 0$ holds $\forall \theta \in \left[\underline{\theta},\bar{\theta}\right]$.

Similar to the proof of theorem 1, we guess that the profit-maximizing firm would leave zero surplus to consumers with the highest privacy sensitivity (i.e., $a\eta^P \ \ \bar{\theta},\beta \ \ - \ \ \bar{\theta} - a \ \ i^P \ \ \bar{\theta},\beta \ \ = 0$); except for bunching the lower interval of the type space (due to the fact that preservation constraint (4.3) binds), the firm implements a separating menu for the remaining portion of the market. Within the separating region, incentive compatibility requires that an admissible $\eta^P \ \ \theta,\beta$ has the following relationship with $i^P \ \ \theta,\beta$,

$$\eta^P \ \ \theta,\beta \ \ = \frac{1}{a}\left[\int_{\theta}^{\bar{\theta}} i^P \ \ t,\beta \ \ dt + \theta i^P \ \ \theta,\beta \ \ \right] - i^P \ \ \theta,\beta \qquad \text{(A.26)}$$

which is derived from integrating (A.4) over $\left[\theta,\bar{\theta}\right]$ for $\theta \geq \hat{\theta}^P$ and considering that the IR binds in prescribing the upper limit of the integral $\eta^P \ \ \bar{\theta},\beta \ \ .$

Suppose we can find a function $\mu(\theta)$ for all θ, designated as the multiplier of $i_1(\theta) \le 0$, and a function $\nu(\theta)$ for all θ, designated as the multiplier of $\eta_1(\theta) = \dfrac{\theta - a}{a} i_1(\theta)$ (by equation (A.4)). We can then define the Hamiltonian as equation (A.11). Suppose we can further find a non-decreasing function $\lambda^P(\theta)$ for all θ such that $\lambda^P(\theta) = \hat{\lambda}^P$ for $\theta > \hat{\theta}^P$ (because $1 - (\beta\eta(\theta) - \beta i(\theta)) > 0$), and a non-decreasing function $\gamma(\theta)$ for all θ such that $\gamma_1(\theta) = 0$ for $\theta < \bar{\theta}$ (since $a(i(\theta) + \eta(\theta)) - \theta i(\theta) > 0$), we can define the Lagrangian as

$$L(i, \eta, i_1, \mu, \nu, \theta, \lambda^P, \gamma) = H(i, \eta, i_1, \mu, \nu, \theta) - \lambda^P(\theta)\left[1 - \left(\beta \frac{\theta - a}{a} i_1(\theta) - \beta i_1(\theta)\right)\right] + \gamma(\theta) i(\theta)$$

where $\mu(\theta) - \beta\lambda^P(\theta) + (a - \theta)\gamma(\theta)$ and $\nu(\theta) + (1 - \beta)\lambda^P(\theta) + a\gamma(\theta)$ is continuous.

From the Pontryagin principle, we have

$$-\left(\mu_1(\theta) - \beta\lambda_1^P(\theta) - \gamma(\theta) + (a - \theta)\gamma_1(\theta)\right) = \frac{\partial L}{\partial i} = \left[\Phi_1(i(\theta)) - 1\right] f(\theta) + \gamma(\theta)$$

$$\text{and } -\left(\nu_1(\theta) + (1 - \beta)\lambda_1^P(\theta) + a\gamma_1(\theta)\right) = \frac{\partial L}{\partial \eta} = -f(\theta)$$

Without loss of generality we set $\lambda^P(\underline{\theta}) = \gamma(\underline{\theta}) = 0$. Optimality requires $\mu(\underline{\theta}) = \nu(\underline{\theta}) = \mu(\bar{\theta}) = \nu(\bar{\theta}) = 0$. Integrating the two equations and setting $\mu(\underline{\theta}) = \nu(\underline{\theta}) = 0$ yields

$$\mu(\theta) - \beta\lambda^P(\theta) + (a - \theta)\gamma(\theta) = \int_{\underline{\theta}}^{\theta} \left[(1 - \Phi_1(i(t))) f(t) - \gamma(t)\right] dt \qquad \text{(A.27)}$$

$$\text{and } \nu(\theta) + (1 - \beta)\lambda^P(\theta) + a\gamma(\theta) = F(\theta) \qquad \text{(A.28)}$$

Maximizing $H(i, \eta, i_1(\theta), \mu(\theta), \nu(\theta), \theta)$ w.r.t. $i_1(\cdot) \le 0$ requires $\mu(\theta) + \nu(\theta)\dfrac{\theta - a}{a} \ge 0$ with $i_1(\theta) = 0$ if $\mu(\theta) + \nu(\theta)\dfrac{\theta - a}{a} > 0$.

Also note that $\max_{i_1(\cdot)} H(i, \eta, i_1(\theta), \mu(\theta), \nu(\theta), \theta) = \left[\Phi(i(\theta)) - (i(\theta) + \eta(\theta))\right] f(\theta)$ is concave (see Theorem 1). Moreover, $\eta(\theta) - \beta(i(\theta) + \eta(\theta))$ and $a(i(\theta) + \eta(\theta)) - \theta i(\theta)$ are clearly quasi-concave in i and η.

Suppose $i_1(\theta) < 0$ on the interval $\left[\hat{\theta}^P, \bar{\theta}\right]$, then $\mu(\theta) + \nu(\theta)\dfrac{\theta - a}{a} = 0$ on this interval. Substituting $\mu(\theta)$ and $\nu(\theta)$ with (A.27) and (A.28) yields

$$
\begin{aligned}
&\mu(\theta)+\nu(\theta)\frac{\theta-a}{a} &\text{(A.29)}\\
&=\int_{\underline{\theta}}^{\theta}\left[\left(1-\Phi_1\left(i(t)\right)\right)f(t)-\gamma(t)\right]dt+\beta\lambda^P(\theta)(-a-\theta)\gamma(\theta)\\
&+\frac{\theta-a}{a}\left[F(\theta)-(1-\beta)\lambda^P(\theta)-a\gamma(\theta)\right]\\
&=\int_{\underline{\theta}}^{\theta}\left[\left(-\Phi_1\left(i(t)\right)\right)f(t)-\gamma(t)\right]dt+\lambda^P(\theta)+\frac{\theta}{a}F(\theta)-\frac{\theta}{a}(1-\beta)\lambda^P(\theta)\\
&=\int_{\underline{\theta}}^{\theta}\left[\left(\frac{1}{a}\left(t+\frac{F(t)}{f(t)}\right)-\Phi_1\left(i(t)\right)\right)f(t)-\gamma(t)\right]dt+\lambda^P(\theta)-\frac{\theta}{a}(1-\beta)\lambda^P(\theta)
\end{aligned}
$$

The last equality follows directly from integration by parts.

Since $\lambda^P(\theta)$ and $\gamma(\theta)$ is continuous at $\hat{\theta}^P$, the multipliers $\mu(\theta)$ and $\nu(\theta)$ are also continuous. We thereby have $\mu(\hat{\theta}^P)+\nu(\hat{\theta}^P)\frac{\hat{\theta}^P-a}{a}=0$. Consequently,

$$
\begin{aligned}
&\mu(\theta)+\nu(\theta)\frac{\theta-a}{a}-\left(\mu(\hat{\theta}^P)+\nu(\hat{\theta}^P)\frac{\hat{\theta}^P-a}{a}\right)\\
&=\int_{\underline{\theta}}^{\theta}\left[\left(\frac{1}{a}\left(t+\frac{F(t)}{f(t)}\right)-\Phi_1\left(i(t)\right)\right)f(t)-\gamma(t)\right]dt+\lambda^P(\theta)-\frac{\theta}{a}(1-\beta)\lambda^P(\theta)-\\
&-\int_{\underline{\theta}}^{\hat{\theta}^P}\left[\left(\frac{1}{a}\left(t+\frac{F(t)}{f(t)}\right)-\Phi_1\left(i(t)\right)\right)f(t)-\gamma(t)\right]dt-\lambda^P(\hat{\theta}^P)+\frac{\hat{\theta}^P}{a}(1-\beta)\lambda^P(\hat{\theta}^P)\\
&=\int_{\hat{\theta}^P}^{\theta}\left[\left(\frac{1}{a}\left(t+\frac{F(t)}{f(t)}\right)-\Phi_1\left(i(t)\right)\right)f(t)\right]dt-\frac{\theta}{a}(1-\beta)\lambda^P(\theta)+\frac{\hat{\theta}^P}{a}(1-\beta)\lambda^P(\hat{\theta}^P)\\
&=\int_{\hat{\theta}^P}^{\theta}\left[\frac{1}{a}\left(t+\frac{F(t)-(1-\beta)\hat{\lambda}^P}{f(t)}\right)-\Phi_1\left(i(t)\right)\right]f(t)\,dt=0
\end{aligned}
$$

The last equality follows $\lambda^P(\theta)=\hat{\lambda}^P$ for $\theta>\hat{\theta}^P$ and $\lambda^P(\theta)$ is continuous at $\hat{\theta}^P$. Hence $\int_{\hat{\theta}^P}^{\theta}\left[\frac{1}{a}\left(t+\frac{F(t)-(1-\beta)\hat{\lambda}^P}{f(t)}\right)-\Phi_1\left(i(t)\right)\right]f(t)\,dt=0$ for $\forall\theta\in\left[\hat{\theta}^P,\overline{\theta}\right]$, implying (A.23) in Theorem 2. In particular, since $\mu(\hat{\theta}^P)+\nu(\hat{\theta}^P)\frac{\hat{\theta}^P-a}{a}=0$ and $\gamma(\theta)=0$ for $\forall\theta\in\left[\underline{\theta},\overline{\theta}\right)$, from (A.29) it follows that

$$
\int_{\underline{\theta}}^{\hat{\theta}^P}\left[\frac{1}{a}\left(t+\frac{F(t)}{f(t)}\right)-\Phi_1\left(i(t)\right)\right]f(t)\,dt+\left(1-\frac{\hat{\theta}^P}{a}(1-\beta)\right)\hat{\lambda}^P=0
$$

By definition of bunching, $i(\theta)$ is constant on $\left[\underline{\theta},\hat{\theta}^P\right)$. From the continuity of the contract menu, the constant level of commercial use is also the solution to equation (A.23) at $\hat{\theta}^P$. Substituting $\Phi_1\left(i(\theta)\right)$ with $\frac{1}{a}\left[\hat{\theta}^P+\frac{F(\hat{\theta}^P)-(1-\beta)\hat{\lambda}^P}{f(\hat{\theta}^P)}\right]$ into the above equation and multiplying both sides of by a result (A.24). Equation (A.25) reflects the binding of constraint (4.3) at $\hat{\theta}^P$.

We need to then identify a $\gamma(\overline{\theta}) > 0$ (so that our conjecture that the participation constraint binds at $\overline{\theta}$ is correct) such that $\mu(\overline{\theta}) = \nu(\overline{\theta}) = 0$ and $\lambda^P(\theta)$ are non-decreasing on $\left[\underline{\theta}, \hat{\theta}^P\right]$. From (A.25) and (A.26),

$(1-\beta)\int_{\hat{\theta}^P}^{\overline{\theta}} i^P(t,\beta)\,dt = \left(a-(1-\beta)\hat{\theta}^P\right) i^P\left(\hat{\theta}^P,\beta\right)$, which implies $a-(1-\beta)\hat{\theta}^P > 0$ for $\hat{\theta}^P < \overline{\theta}$

Rearranging (A.24), $$\hat{\lambda}^P = \frac{F\left(\hat{\theta}^P\right)^2}{f\left(\hat{\theta}^P\right)\left[a-(1-\beta)\hat{\theta}^P + (1-\beta)\dfrac{F\left(\hat{\theta}^P\right)}{f\left(\hat{\theta}^P\right)}\right]} \qquad \text{(A.30)}$$

From (A.28) and $\nu(\overline{\theta}) = 0$, $(1-\beta)\hat{\lambda}^P + a\gamma(\overline{\theta}) = 1$, which implies

$$a\gamma(\overline{\theta}) = 1-(1-\beta)\hat{\lambda}^P = 1-\frac{(1-\beta)\dfrac{F\left(\hat{\theta}^P\right)}{f\left(\hat{\theta}^P\right)}}{\left[a-(1-\beta)\hat{\theta}^P + (1-\beta)\dfrac{F\left(\hat{\theta}^P\right)}{f\left(\hat{\theta}^P\right)}\right]}F\left(\hat{\theta}^P\right) > 0 \qquad \text{(A.31)}$$

The inequality follows $0 < \dfrac{(1-\beta)\dfrac{F\left(\hat{\theta}^P\right)}{f\left(\hat{\theta}^P\right)}}{\left[a-(1-\beta)\hat{\theta}^P + (1-\beta)\dfrac{F\left(\hat{\theta}^P\right)}{f\left(\hat{\theta}^P\right)}\right]} < 1$ and $0 < F\left(\hat{\theta}^P\right) < 1$ for $\hat{\theta}^P \in \left(\underline{\theta}, \overline{\theta}\right)$.

From (A.27), $\mu(\overline{\theta}) = 1-\int_{\underline{\theta}}^{\overline{\theta}} \Phi_1(i(\theta))f(\theta)\,d\theta + \beta\hat{\lambda}^P - (a-\overline{\theta})\gamma(\overline{\theta})$, where

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}} \Phi_1(i(\theta))f(\theta)\,d\theta \\
&= \int_{\underline{\theta}}^{\hat{\theta}^P} \Phi_1(i(\theta))f(\theta)\,d\theta + \int_{\hat{\theta}^P}^{\overline{\theta}} \Phi_1(i(\theta))f(\theta)\,d\theta \\
&= \frac{1}{a}\int_{\underline{\theta}}^{\hat{\theta}^P}\left(\hat{\theta}^P + \frac{F\left(\hat{\theta}^P\right)-(1-\beta)\hat{\lambda}^P}{f\left(\hat{\theta}^P\right)}\right)f(\theta)\,d\theta + \frac{1}{a}\int_{\hat{\theta}^P}^{\overline{\theta}}\left(\theta + \frac{F(\theta)-(1-\beta)\hat{\lambda}^P}{f(\theta)}\right)f(\theta)\,d\theta \\
&= \frac{1}{a}\left[\left(\hat{\theta}^P + \frac{F\left(\hat{\theta}^P\right)-(1-\beta)\hat{\lambda}^P}{f\left(\hat{\theta}^P\right)}\right)F\left(\hat{\theta}^P\right) + \overline{\theta} - \hat{\theta}^P F\left(\hat{\theta}^P\right) - \left(\overline{\theta}-\hat{\theta}^P\right)(1-\beta)\hat{\lambda}^P\right] \\
&= \frac{1}{a}\left[\frac{F\left(\hat{\theta}^P\right)-(1-\beta)\hat{\lambda}^P}{f\left(\hat{\theta}^P\right)}F\left(\hat{\theta}^P\right) + \overline{\theta} - \left(\overline{\theta}-\hat{\theta}^P\right)(1-\beta)\hat{\lambda}^P\right]
\end{aligned}$$

Substituting $\hat{\lambda}^P$ with equation (A.30) and $\gamma(\overline{\theta})$ with equation (A.31) into the above equation, we have $\mu(\overline{\theta}) = 0$. It can be established that $\theta + \dfrac{F(\theta) - (1-\beta)\hat{\lambda}^P}{f(\theta)}$ in equation (A.23) is increasing on $\left[\hat{\theta}^P, \overline{\theta}\right]$ (Lemma A2) as $(1-\beta)\hat{\lambda}^P < 1$ from (A.31).

When $\beta > (\overline{\theta} - a)/\overline{\theta}$, we have $\eta(\overline{\theta}) \geq \beta\left(i(\overline{\theta}) + \eta(\overline{\theta})\right)$ for type $\overline{\theta}$. The participation constraint for type $\overline{\theta}$ becomes slack; i.e., $a\left(i(\overline{\theta}) + \eta(\overline{\theta})\right) - \overline{\theta} i(\overline{\theta}) > 0$ because $\left(1 - \dfrac{\overline{\theta} - a}{\overline{\theta}}\right)\left(i(\overline{\theta}) + \eta(\overline{\theta})\right) - i(\overline{\theta}) > (1-\beta)\left(i(\overline{\theta}) + \eta(\overline{\theta})\right) - i(\overline{\theta}) \geq 0$.

The first inequality follows $\beta > (\overline{\theta} - a)/\overline{\theta}$ and the second follows $\eta(\overline{\theta}) \geq \beta\left(i(\overline{\theta}) + \eta(\overline{\theta})\right)$. Then the firm's response characterized in this theorem does not apply to the scenario where the mandatory ratio is strictly larger than $(\overline{\theta} - a)/\overline{\theta}$. The proof is complete. ■

Proof of Proposition 4: Part 1 of Proposition 4 is the direct result of Theorem 2. Substituting (A.30) into $\dfrac{F(\hat{\theta}^P) - (1-\beta)\hat{\lambda}^P}{f(\hat{\theta}^P)}$ yields

$$\frac{F(\hat{\theta}^P) - (1-\beta)\hat{\lambda}^P}{f(\hat{\theta}^P)} = \frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\left[\frac{a - (1-\beta)\hat{\theta}^P}{a - (1-\beta)\hat{\theta}^P + (1-\beta)\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}}\right]$$

Since $a - (1-\beta)\hat{\theta}^P \geq 0$ for $\hat{\theta}^P \in (\underline{\theta}, \overline{\theta}]$ from (A.25) and (A.26),

$$0 \leq \frac{F(\hat{\theta}^P) - (1-\beta)\hat{\lambda}^P}{f(\hat{\theta}^P)} = \frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\left[\frac{a - (1-\beta)\hat{\theta}^P}{a - (1-\beta)\hat{\theta}^P + (1-\beta)\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}}\right] < \frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)},$$ where the inequalities follows $0 \leq \dfrac{a - (1-\beta)\hat{\theta}^P}{a - (1-\beta)\hat{\theta}^P + (1-\beta)\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}} < 1$ for $\hat{\theta}^P \in (\underline{\theta}, \overline{\theta}]$.

For $\theta \in \left[\hat{\theta}^P(\beta), \bar{\theta}\right]$, $\theta + \frac{F(\theta)}{f(\theta)} > \theta + \frac{F(\theta) - (1-\beta)\hat{\lambda}^P}{f(\theta)}$. Relating it to (A.6) and (A.23), and the assumption that $\Phi_{11}(i) < 0$, we have $i^P(\theta,\beta) > i^*(\theta)$. Equations (A.24) and (A.25) simultaneously determine $\hat{\theta}^P$ and $\hat{\lambda}^P$, both of which are functions of the preservation level, β.

According to Theorem 1, $i^P(\theta,\beta)$ is determined by equation (A.23) on the separating interval $\left[\hat{\theta}(\beta), \bar{\theta}\right]$. Fully differentiating (A.23) w.r.t. β gives

$$af(\theta)\Phi_{11}(i^P(\theta,\beta))\, i_2^P(\theta,\beta) = -(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P(\beta) \tag{A.32}$$

Similarly, $i_1^P(\theta,\beta)$ satisfies

$$a\Phi_{11}(i^P(\theta,\beta))\, i_1^P(\theta,\beta) = H_1^P(\theta) \tag{A.33}$$,

where $H_1^P(\theta) \equiv d\left(\theta + \frac{F(\theta) - (1-\beta)\hat{\lambda}^P}{f(\theta)}\right) / d\theta$.

Fully differentiating (A.24) and (A.25) w.r.t. β yields

$$\begin{aligned}\left(\hat{\theta}^P - \frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\hat{\lambda}^P + \left[a + \frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}(1-\beta) - \hat{\theta}^P(1-\beta)\right]\hat{\lambda}_1^P(\beta) \\ -H_1^P(\hat{\theta}^P)F(\hat{\theta}^P)\hat{\theta}_1^P(\beta) = 0\end{aligned} \tag{A.34}$$

and
$$\begin{aligned}-\left[\int_{\hat{\theta}^P}^{\bar{\theta}} i^P(t,\beta)\,dt + \hat{\theta}^P i^P(\hat{\theta}^P,\beta)\right] + (1-\beta)\int_{\hat{\theta}^P}^{\bar{\theta}} i_2^P(t,\beta)\,dt + \\ ((1-\beta)\hat{\theta}^P - a)\left[i_1^P(\hat{\theta}^P,\beta)\hat{\theta}_1^P(\beta) + i_2^P(\hat{\theta}^P,\beta)\right] = 0\end{aligned} \tag{A.35}$$

respectively.

Substituting (A.32) and (A.33) into (A.35) yields

$$\begin{aligned}-\left[\int_{\hat{\theta}^P}^{\bar{\theta}} i^P(t,\beta)\,dt + \hat{\theta}^P i^P(\hat{\theta}^P,\beta)\right] + (1-\beta)\left[-(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P\right]\int_{\hat{\theta}^P}^{\bar{\theta}} \frac{1}{af(t)\Phi_{11}(i^P(t,\beta))}dt \\ + ((1-\beta)\hat{\theta}^P - a)\frac{1}{a\Phi_{11}(i^P(\hat{\theta}^P,\beta))}\left[H_1^P(\hat{\theta}^P)\hat{\theta}_1^P(\beta) + \frac{-(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P}{f(\hat{\theta}^P)}\right] = 0\end{aligned} \tag{A.36}$$

Given (A.34), we have

$$\hat{\lambda}_1^P(\beta) = \frac{H_1^P(\hat{\theta}^P)F(\hat{\theta}^P)\hat{\theta}_1^P(\beta) - \left[\hat{\theta}^P(\beta) - \frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right]\hat{\lambda}^P(\beta)}{\left[a - (1-\beta)\hat{\theta}^P(\beta) + (1-\beta)\frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right]} \tag{A.37}$$

(According to(A.30), $\hat{\lambda}_1^P(\beta)$ is well defined when $\hat{\lambda}^P(\beta)$ is well defined).

Substituting (A.37) and $\int_{\hat{\theta}^P}^{\bar{\theta}} i^P(t,\beta)\,dt + \hat{\theta}^P i^P(\hat{\theta}^P,\beta) = \dfrac{a i^P(\hat{\theta}^P,\beta)}{1-\beta}$ (according to (A.25)) into (A.36) yields

$$\left[-\frac{\left(\dfrac{a}{1-\beta}-\hat{\theta}^P\right)^2}{a\Phi_{11}\left(i^P(\hat{\theta}^P,\beta)\right)} - F(\hat{\theta}^P)\int_{\hat{\theta}^P}^{\bar{\theta}} \frac{1}{af(t)\Phi_{11}\left(i^P(t,\beta)\right)}dt\right]\frac{(1-\beta)H_1^P(\hat{\theta}^P)}{\left[\dfrac{a}{1-\beta}-\hat{\theta}^P+\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right]}\hat{\theta}_1^P(\beta)$$

$$= -\frac{a\hat{\lambda}^P}{\left[\dfrac{a}{1-\beta}-\hat{\theta}^P+\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right]}\int_{\hat{\theta}^P}^{\bar{\theta}} \frac{1}{af(t)\Phi_{11}\left(i^P(t,\beta)\right)}dt + \frac{a}{1-\beta}\left[\frac{(1-\beta)\hat{\lambda}^P}{af(\hat{\theta}^P)\Phi_{11}\left(i^P(\hat{\theta}^P,\beta)\right)}\frac{\dfrac{a}{(1-\beta)}-\hat{\theta}^P}{\dfrac{a}{(1-\beta)}-\hat{\theta}^P+\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}}+i^P(\hat{\theta}^P,\beta)\right]$$

Since $\Phi_{11}(\cdot)<0$, $\dfrac{a}{1-\beta}-\hat{\theta}^P+\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}>0$, and $H_1^P(\hat{\theta}^P)>0$, the term on the left hand side of the equation is positive for $\beta<\dfrac{\bar{\theta}-a}{\bar{\theta}}$. Then what remains to be checked is the sign of $\dfrac{(1-\beta)\hat{\lambda}^P}{af(\hat{\theta}^P)\Phi_{11}\left(i^P(\hat{\theta}^P,\beta)\right)}+i^P(\hat{\theta}^P,\beta)$ on the right hand side of the equation. According to (A.23), $i^P(\hat{\theta}^P,\beta)=\Phi_1^{-1}\left[\dfrac{1}{a}\left(\hat{\theta}^P+\dfrac{F(\hat{\theta}^P)-(1-\beta)\hat{\lambda}^P(\beta)}{f(\hat{\theta}^P)}\right)\right]$. Note that when $\Phi_{111}(i)\le 0$,

$$\Phi_1^{-1}\left[\frac{1}{a}\left(\hat{\theta}^P+\frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\right]-\Phi_1^{-1}\left[\frac{1}{a}\left(\hat{\theta}^P+\frac{F(\hat{\theta}^P)-(1-\beta)\hat{\lambda}^P(\beta)}{f(\hat{\theta}^P)}\right)\right]\le\frac{1}{\Phi_{11}\left(i^P(\hat{\theta}^P,\beta)\right)}\frac{1}{a}\frac{(1-\beta)\hat{\lambda}^P}{af(\hat{\theta}^P)}$$

which implies that $\dfrac{(1-\beta)\hat{\lambda}^P}{af(\hat{\theta}^P)\Phi_{11}\left(i^P(\hat{\theta}^P,\beta)\right)}+i^P(\hat{\theta}^P,\beta)\ge\Phi_1^{-1}\left[\dfrac{1}{a}\left(\hat{\theta}^P+\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\right]$ with $\Phi_1(0)>\dfrac{1}{a}\left(\bar{\theta}+\dfrac{1}{F(\bar{\theta})}\right)$, $\Phi_1^{-1}\left[\dfrac{1}{a}\left(\hat{\theta}^P+\dfrac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\right]>0$. ($\Phi_{111}(\cdot)\le 0$ is a sufficient condition; some functions with $\Phi_{111}(i)>0$, e.g., $\Phi(i)=2b\sqrt{i}$, can still ensure $\hat{\theta}_1^P(\beta)>0$ (see the numerical example)).

When $\beta>(\bar{\theta}-a)/\bar{\theta}$, The participation constraint is slack for all consumer types, i.e., $a(i(\theta)+\eta(\theta))-\theta i(\theta)>0$; in addition, incentive compatibility requires that the entire market be served an identical contract (Lemma 2). ■

Proof of Proposition 5: For a partially-bunching menu, (4.4) can be represented as:

$$
\begin{aligned}
SW(\beta) &= \int_{\underline{\theta}}^{\overline{\theta}} \left[\Phi\left(i^P(\theta,\beta)\right) - i^P(\theta,\beta) + \eta^P(\theta,\beta) \right] f(\theta)\, d\theta + \\
&\quad \alpha \int_{\underline{\theta}}^{\overline{\theta}} \left[a\, i^P(\theta,\beta) + \eta^P(\theta,\beta) - \theta i^P(\theta,\beta) \right] f(\theta)\, d\theta \\
&= \int_{\underline{\theta}}^{\hat{\theta}^P} \left[\Phi\left(i^P(\hat{\theta}^P,\beta)\right) + (\alpha a - 1)\, i^P(\hat{\theta}^P,\beta) + \eta^P(\hat{\theta}^P,\beta) - \alpha\theta i^P(\hat{\theta}^P,\beta) \right] f(\theta)\, d\theta \\
&\quad + \int_{\hat{\theta}^P}^{\overline{\theta}} \left[\Phi\left(i^P(\theta,\beta)\right) + (\alpha a - 1)\, i^P(\theta,\beta) + \eta^P(\theta,\beta) - \alpha\theta i^P(\theta,\beta) \right] f(\theta)\, d\theta
\end{aligned}
.
$$

Substituting (A.26) into the above equation yields

$$
\begin{aligned}
SW(\beta) &= \int_{\underline{\theta}}^{\hat{\theta}^P} \left[\Phi\left(i^P(\hat{\theta}^P,\beta)\right) + \left(\alpha - \frac{1}{a}\right) \int_{\hat{\theta}^P}^{\overline{\theta}} i^P(t,\beta)\, dt + \hat{\theta}^P i^P(\hat{\theta}^P,\beta) - \alpha\theta i^P(\hat{\theta}^P,\beta) \right] f(\theta)\, d\theta \\
&\quad + \int_{\hat{\theta}^P}^{\overline{\theta}} \left[\Phi\left(i^P(\theta,\beta)\right) + \left(\alpha - \frac{1}{a}\right) \int_{\theta}^{\overline{\theta}} i^P(t,\beta)\, dt + \theta i^P(\theta,\beta) - \alpha\theta i^P(\theta,\beta) \right] f(\theta)\, d\theta \\
&= \int_{\underline{\theta}}^{\hat{\theta}^P} \left[\Phi\left(i^P(\hat{\theta}^P,\beta)\right) + \left(\alpha\left(\hat{\theta}^P - \theta\right) - \frac{\hat{\theta}^P}{a}\right) i^P(\hat{\theta}^P,\beta) + \left(\alpha - \frac{1}{a}\right) \int_{\hat{\theta}^P}^{\overline{\theta}} i^P(t,\beta)\, dt \right] f(\theta)\, d\theta \\
&\quad + \int_{\hat{\theta}^P}^{\overline{\theta}} \left[\Phi\left(i^P(\theta,\beta)\right) - \frac{\theta}{a} i^P(\theta,\beta) + \left(\alpha - \frac{1}{a}\right) \int_{\theta}^{\overline{\theta}} i^P(t,\beta)\, dt \right] f(\theta)\, d\theta
\end{aligned}
.
$$

Fully differentiating $SW(\beta)$ w.r.t. β yields

$$
\begin{aligned}
\frac{d}{d\beta} SW(\beta) &= \int_{\underline{\theta}}^{\hat{\theta}^P} \left[\begin{aligned} &\Phi_1\left(i^P(\hat{\theta}^P,\beta)\right)\left(i_1^P(\hat{\theta}^P,\beta)\,\hat{\theta}_1^P(\beta) + i_2^P(\hat{\theta}^P,\beta)\right) \\ &+ \left(\alpha\left(\hat{\theta}^P - \theta\right) - \frac{\hat{\theta}^P}{a}\right)\left(i_1^P(\hat{\theta}^P,\beta)\,\hat{\theta}_1^P(\beta) + i_2^P(\hat{\theta}^P,\beta)\right) + \left(\alpha - \frac{1}{a}\right) \int_{\hat{\theta}^P}^{\overline{\theta}} i_2^P(t,\beta)\, dt \end{aligned} \right] f(\theta)\, d\theta \\
&\quad + \int_{\hat{\theta}^P}^{\overline{\theta}} \left[\Phi_1\left(i^P(\theta,\beta)\right) i_2^P(\theta,\beta) - \frac{\theta}{a} i_2^P(\theta,\beta) + \left(\alpha - \frac{1}{a}\right) \int_{\theta}^{\overline{\theta}} i_2^P(t,\beta)\, dt \right] f(\theta)\, d\theta \\
&= \int_{\underline{\theta}}^{\hat{\theta}^P} \left[\begin{aligned} &\left(\Phi_1\left(i^P(\hat{\theta}^P,\beta)\right) + \alpha\left(\hat{\theta}^P - \theta\right) - \frac{\hat{\theta}^P}{a}\right)\left(i_1^P(\hat{\theta}^P,\beta)\,\hat{\theta}_1^P(\beta) + i_2^P(\hat{\theta}^P,\beta)\right) \\ &+ \left(\alpha - \frac{1}{a}\right) \int_{\hat{\theta}^P}^{\overline{\theta}} i_2^P(t,\beta)\, dt \end{aligned} \right] f(\theta)\, d\theta \\
&\quad + \int_{\hat{\theta}^P}^{\overline{\theta}} \left[\left(\Phi_1\left(i^P(\theta,\beta)\right) - \frac{\theta}{a}\right) i_2^P(\theta,\beta) + \left(\alpha - \frac{1}{a}\right) \int_{\theta}^{\overline{\theta}} i_2^P(t,\beta)\, dt \right] f(\theta)\, d\theta \\
&= \int_{\underline{\theta}}^{\hat{\theta}^P} \left(\Phi_1\left(i^P(\hat{\theta}^P,\beta)\right) + \alpha\left(\hat{\theta}^P - \theta\right) - \frac{\hat{\theta}^P}{a}\right)\left(i_1^P(\hat{\theta}^P,\beta)\,\hat{\theta}_1^P(\beta) + i_2^P(\hat{\theta}^P,\beta)\right) f(\theta)\, d\theta \\
&\quad + \left(\alpha - \frac{1}{a}\right) F\left(\hat{\theta}^P\right) \int_{\hat{\theta}^P}^{\overline{\theta}} i_2^P(t,\beta)\, dt + \left(\alpha - \frac{1}{a}\right) \left(-F\left(\hat{\theta}^P\right) \int_{\hat{\theta}^P}^{\overline{\theta}} i_2^P(t,\beta)\, dt + \int_{\hat{\theta}^P}^{\overline{\theta}} F(\theta)\, i_2^P(\theta,\beta)\, d\theta \right) \\
&\quad + \int_{\hat{\theta}^P}^{\overline{\theta}} \left(\Phi_1\left(i^P(\theta,\beta)\right) - \frac{\theta}{a}\right) i_2^P(\theta,\beta) f(\theta)\, d\theta \\
&= \int_{\underline{\theta}}^{\hat{\theta}^P} \left(\Phi_1\left(i^P(\hat{\theta}^P,\beta)\right) + \alpha\left(\hat{\theta}^P - \theta\right) - \frac{\hat{\theta}^P}{a}\right)\left(i_1^P(\hat{\theta}^P,\beta)\,\hat{\theta}_1^P(\beta) + i_2^P(\hat{\theta}^P,\beta)\right) f(\theta)\, d\theta \\
&\quad + \left(\alpha - \frac{1}{a}\right) \int_{\hat{\theta}^P}^{\overline{\theta}} F(\theta)\, i_2^P(\theta,\beta)\, d\theta + \int_{\hat{\theta}^P}^{\overline{\theta}} \left(\Phi_1\left(i^P(\theta,\beta)\right) - \frac{\theta}{a}\right) i_2^P(\theta,\beta) f(\theta)\, d\theta
\end{aligned}
.
$$

The third equality follows integration by parts. Substituting (A.23) into the above equation yields

$$\begin{aligned}\frac{d}{d\beta}SW(\beta) &= \int_{\underline{\theta}}^{\hat{\theta}^P}\left(\frac{1}{a}\frac{F(\hat{\theta}^P) - (1-\beta)\hat{\lambda}^P}{f(\hat{\theta}^P)} + \alpha(\hat{\theta}^P - \theta)\right)\left(i_1^P(\hat{\theta}^P,\beta)\hat{\theta}_1^P(\beta) + i_2^P(\hat{\theta}^P,\beta)\right) f(\theta)\,d\theta \\ &+\int_{\hat{\theta}^P}^{\bar{\theta}}\left(\alpha F(\theta) - \frac{1}{a}(1-\beta)\hat{\lambda}^P\right) i_2^P(\theta,\beta)\,d\theta\end{aligned}$$

Substituting (A.32) and (A.33) into the above equation yields

$$\begin{aligned}&\frac{d}{d\beta}SW(\beta) \\ &= \int_{\underline{\theta}}^{\hat{\theta}^P}\left(\frac{1}{a}\frac{F(\hat{\theta}^P) - (1-\beta)\hat{\lambda}^P}{f(\hat{\theta}^P)} + \alpha(\hat{\theta}^P - \theta)\right)\left(\frac{H_1^P(\hat{\theta}^P)}{a\Phi_{11}(i^P(\hat{\theta}^P,\beta))}\hat{\theta}_1^P(\beta) + \frac{-(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P(\beta)}{af(\hat{\theta}^P)\Phi_{11}(i^P(\hat{\theta}^P,\beta))}\right) f(\theta)\,d\theta \\ &+\int_{\hat{\theta}^P}^{\bar{\theta}}\left(\alpha F(\theta) - \frac{1}{a}(1-\beta)\hat{\lambda}^P\right)\frac{-(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P(\beta)}{af(\theta)\Phi_{11}(i^P(\theta,\beta))}\,d\theta\end{aligned}$$

(A.38)

When $\beta = \Delta^*(\underline{\theta})$, $\hat{\lambda}^P(\beta)|_{\beta=\Delta^*(\underline{\theta})} = 0$ and $\hat{\theta}^P(\beta)|_{\beta=\Delta^*(\underline{\theta})} = \underline{\theta}$, and $\hat{\lambda}_1^P(\beta)|_{\beta=\Delta^*(\underline{\theta})} = 0$ and $\hat{\theta}_1^P(\beta)|_{\beta=\Delta^*(\underline{\theta})} > 0$. Evaluating (A.38) at $\beta = \Delta^*(\underline{\theta})$, $\frac{d}{d\beta}SW(\beta)|_{\beta=\Delta^*(\underline{\theta})} = 0$. Further differentiating (A.38) w.r.t. β yields:

$$\begin{aligned}&\frac{d^2}{d\beta^2}SW(\beta) \\ &= \left[\begin{array}{l}\left(\frac{1}{a}\frac{F(\hat{\theta}^P) - (1-\beta)\hat{\lambda}^P}{f(\hat{\theta}^P)} + \alpha(\hat{\theta}^P - \hat{\theta}^P)\right)\left(\frac{H_1^P(\hat{\theta}^P)}{a\Phi_{11}(i^P(\hat{\theta}^P,\beta))}\hat{\theta}_1^P(\beta) + \frac{-(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P(\beta)}{af(\hat{\theta}^P)\Phi_{11}(i^P(\hat{\theta}^P,\beta))}\right) f(\hat{\theta}^P) \\ -\left(\alpha F(\hat{\theta}^P) - \frac{1}{a}(1-\beta)\hat{\lambda}^P\right)\frac{-(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P(\beta)}{af(\hat{\theta}^P)\Phi_{11}(i^P(\hat{\theta}^P,\beta))}\end{array}\right]\hat{\theta}_1^P(\beta) \\ &+\int_{\underline{\theta}}^{\hat{\theta}^P}\frac{d}{d\beta}\left[\left(\frac{1}{a}\frac{F(\hat{\theta}^P) - (1-\beta)\hat{\lambda}^P}{f(\hat{\theta}^P)} + \alpha(\hat{\theta}^P - \theta)\right)\left(\frac{H_1^P(\hat{\theta}^P)}{a\Phi_{11}(i^P(\hat{\theta}^P,\beta))}\hat{\theta}_1^P(\beta) + \frac{-(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P(\beta)}{af(\hat{\theta}^P)\Phi_{11}(i^P(\hat{\theta}^P,\beta))}\right) f(\theta)\right]d\theta \\ &+\int_{\hat{\theta}^P}^{\bar{\theta}}\frac{d}{d\beta}\left[\left(\alpha F(\theta) - \frac{1}{a}(1-\beta)\hat{\lambda}^P\right)\frac{-(1-\beta)\hat{\lambda}_1^P(\beta) + \hat{\lambda}^P(\beta)}{af(\theta)\Phi_{11}(i^P(\theta,\beta))}\right]d\theta\end{aligned}$$

.

Again, we evaluate $\frac{d^2}{d\beta^2}SW(\beta)$ at $\beta = \Delta^*(\underline{\theta})$. The expression in the first bracket is 0. The first integral also equals 0 since the integrated function is bounded at $\beta = \Delta^*(\underline{\theta})$. The last component within the integral is given by:

$$\frac{d}{d\beta}\left[\left(\alpha F(\theta)-\frac{1}{a}(1-\beta)\hat{\lambda}^P\right)\frac{-(1-\beta)\hat{\lambda}_1^P(\beta)+\hat{\lambda}^P(\beta)}{af(\theta)\Phi_{11}(i^P(\theta,\beta))}\right]$$

$$=\frac{1}{a}\hat{\lambda}^P\frac{-(1-\beta)\hat{\lambda}_1^P(\beta)+\hat{\lambda}^P(\beta)}{af(\theta)\Phi_{11}(i^P(\theta,\beta))}+\frac{1}{a}\left(\alpha\frac{F(\theta)}{f(\theta)}-\frac{1}{a}\frac{(1-\beta)\hat{\lambda}^P}{f(\theta)}\right)$$

$$\frac{\left(-(1-\beta)\hat{\lambda}_{11}^P(\beta)+2\hat{\lambda}_1^P(\beta)\right)\Phi_{11}(i^P(\theta,\beta))-\left(-(1-\beta)\hat{\lambda}_1^P(\beta)+\hat{\lambda}^P(\beta)\right)\Phi_{111}(i^P(\theta,\beta))\,i_2^P(\theta,\beta)}{\left[\Phi_{11}(i^P(\theta,\beta))\right]^2}$$

$$=-\frac{\alpha}{a}\frac{F(\theta)}{f(\theta)}\frac{(1-\beta)\hat{\lambda}_{11}^P(\beta)}{\Phi_{11}(i^P(\theta,\beta))}$$

Hence $\frac{d^2}{d\beta^2}SW(\beta)|_{\beta=\Delta^*(\underline{\theta})}=-\int_{\hat{\theta}^P}^{\bar{\theta}}\left[\frac{\alpha}{a}\frac{F(\theta)}{f(\theta)}\frac{(1-\beta)\hat{\lambda}_{11}^P(\beta)}{\Phi_{11}(i^P(\theta,\beta))}\right]d\theta$. The sign of $\frac{d^2}{d\beta^2}SW(\beta)|_{\beta=\Delta^*(\underline{\theta})}$ depends on the sign of $\hat{\lambda}_{11}^P(\beta)$. According to equation (A.37),

$$\hat{\lambda}_{11}^P(\beta)=\frac{1}{\left[(1-\beta)\left(\frac{a}{1-\beta}-\hat{\theta}^P+\frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\right]^2}$$

$$\left\{\begin{array}{l}\frac{d}{d\beta}\left[H_1^P(\hat{\theta}^P)F(\hat{\theta}^P)\hat{\theta}_1^P(\beta)-\left(\hat{\theta}^P(\beta)-\frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\hat{\lambda}^P(\beta)\right](1-\beta)\left(\frac{a}{1-\beta}-\hat{\theta}^P+\frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\\ -\left[H_1^P(\hat{\theta}^P)F(\hat{\theta}^P)\hat{\theta}_1^P(\beta)-\left(\hat{\theta}^P(\beta)-\frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\hat{\lambda}^P(\beta)\right]\frac{d}{d\beta}\left(\frac{a}{(1-\beta)}-\hat{\theta}^P+\frac{F(\hat{\theta}^P)}{f(\hat{\theta}^P)}\right)\end{array}\right\} \quad \text{(A.39)}$$

Evaluating it at $\beta=\Delta^*(\underline{\theta})$, we have $\hat{\lambda}_{11}^P(\beta)|_{\beta=\Delta^*(\underline{\theta})}=\frac{H_1^P(\hat{\theta}^P)f(\hat{\theta}^P)\left[\hat{\theta}_1^P(\beta)\right]^2}{a-(1-\beta)\underline{\theta}}>0$. This result is independent of the value of α. ■

Lemma A1: if the density function $f(\theta)$ is log-concave and non-zero everywhere on $[\underline{\theta},\bar{\theta}]$, $\frac{F(\theta)}{f(\theta)}$ is increasing in θ. ■

Proof of Lemma A1: The function $f(\cdot)$ is log-concave if and only if $\frac{d^2}{d\theta^2}(\ln f(\theta))=\frac{d}{d\theta}\left[\frac{f_1(\theta)}{f(\theta)}\right]\leq 0$.

Then for all $\theta\in[\underline{\theta},\bar{\theta}]$, $\frac{f_1(\theta)}{f(\theta)}F(\theta)=\frac{f_1(\theta)}{f(\theta)}\int_{\underline{\theta}}^{\theta}f(t)dt\leq\int_{\underline{\theta}}^{\theta}\frac{f_1(t)}{f(t)}f(t)dt=f(\theta)-f(\underline{\theta})\leq f(\theta)$. Hence, $f_1(\theta)F(\theta)-f(\theta)^2\leq 0$. It implies $\frac{d}{d\theta}\left[\frac{F(\theta)}{f(\theta)}\right]\geq 0$. ■

Lemma A2: Define the auxiliary function $P(\theta,\mu) = \theta + \dfrac{F(\theta)+\mu}{f(\theta)}$ for $\theta \in [\underline{\theta},\overline{\theta}]$ where μ is an arbitrary constant. The function $P(\theta,\mu)$ is quasi-convex in θ on $[\underline{\theta},\overline{\theta}]$ if $\mu \geq 0$, and is increasing in θ on $[\underline{\theta},\overline{\theta}]$ if $-1 \leq \mu < 0$. ■

Proof of Lemma A2: Since $f(\theta)$ is continuously differentiable and positive everywhere on the support, $P_1(\theta,\mu) = 2 - \dfrac{(F(\theta)+\mu) f_1(\theta)}{f(\theta)^2}$ is continuous and well-defined.

1) $\mu \geq 0$.

Consider an arbitrary extremum (providing that one exists) $\widehat{\theta}$ of $P(\theta,\mu)$. Note that

$$\begin{aligned} f(\theta)^3 P_{11}(\theta,\mu) &= f(\theta)^3 \frac{d}{d\theta}\left[\frac{2f(\theta)^2 - (F(\theta)+\mu)f_1(\theta)}{f(\theta)^2}\right] \\ &= \frac{\left(4f(\theta)f_1(\theta) - (F(\theta)+\mu)f_{11}(\theta) - f_1(\theta)f(\theta)\right)f(\theta)^2 - 2\left(2f(\theta)^2 - (F(\theta)+\mu)f_1(\theta)\right)f(\theta)f_1(\theta)}{f(\theta)} \\ &= -f(\theta)^2 f_1(\theta) + \left(2f_1(\theta)^2 - f_{11}(\theta)f(\theta)\right)(F(\theta)+\mu) \end{aligned}$$

The condition that defines an extremum is given by $P_1(\widehat{\theta},\mu) = 2 - \dfrac{(F(\widehat{\theta})+\mu) f_1(\widehat{\theta})}{f(\widehat{\theta})^2} = 0$, which can be the case only when $f_1(\widehat{\theta}) > 0$ (since $F(\theta)+\mu \geq 0$ for all $\theta \in [\underline{\theta},\overline{\theta}]$). Therefore, $F(\widehat{\theta})+\mu = \dfrac{2f(\widehat{\theta})^2}{f_1(\widehat{\theta})}$ at the extremum. Substituting this into the above second-order expression, we have

$$P_{11}(\widehat{\theta},\mu) = \frac{1}{f(\widehat{\theta})f_1(\widehat{\theta})}\left(f_1(\widehat{\theta})^2 + 2\left(f_1(\widehat{\theta})^2 - f_{11}(\widehat{\theta})f(\widehat{\theta})\right)\right) > 0,$$

where the inequality follows from the log-concavity and everywhere positivity of $f(\theta)$ and $f_1(\widehat{\theta}) > 0$.

Therefore, for $\mu \geq 0$, if there exists any extremum of $P(\theta,\mu)$, it must be a minimum. It implies that once $P_1(\theta,\mu) \geq 0$, it remains so for all higher values of θ.

2) $-1 \leq \mu < 0$.

We can define $\hat{\theta}$ as the point at which $F(\hat{\theta}) + \mu = 0$. Since $f(\theta)$ is everywhere positive, $F(\hat{\theta})$ is strictly increasing and $\hat{\theta}$ is uniquely determined. If there exists an extremum $\widehat{\theta}$ in the interval $[0,\hat{\theta})$,

$P_1(\widehat{\theta},\mu) = 2 - \dfrac{(F(\widehat{\theta})+\mu) f_1(\widehat{\theta})}{f(\widehat{\theta})^2} = 0$ requires $f_1(\widehat{\theta}) < 0$, then it implies that $P_{11}(\widehat{\theta},\mu) = \dfrac{1}{f(\widehat{\theta}) f_1(\widehat{\theta})}\left[f_1(\widehat{\theta})^2 + 2 f_1(\widehat{\theta})^2 - f_{11}(\widehat{\theta}) f(\widehat{\theta}) \right] < 0$. In other words, this extremum must also be a maximum.

Then at the point $\hat{\theta}$ where $F(\hat{\theta}) + \mu = 0$, $P_1(\hat{\theta},\mu) < 0$ by continuity. However, $P_1(\hat{\theta},\mu) = 2 - \dfrac{(F(\hat{\theta})+\mu) f_1(\hat{\theta})}{f(\hat{\theta})^2} > 0$. This results in contradiction, implying that there cannot be an extremum (a decreasing region) in the interval $[0,\hat{\theta})$. In the interval $[\hat{\theta},\overline{\theta}]$, $F(\hat{\theta}) + \mu > 0$. The first part of this lemma implies that $P(\theta,\mu)$ is quasi-convex in this interval. $P_1(\hat{\theta},\mu) > 0$ implies $P(\theta,\mu) > 0$ on $[\hat{\theta},\overline{\theta}]$.■