An Economic Analysis of Peer-Disclosure in Online Social Communities

Zike Cao[†], Kai-Lung Hui[‡], and Hong Xu[‡]

July 10, 2017

Forthcoming at *Information Systems Research*

Abstract

We study a novel privacy concern, viz. peer disclosure of sensitive personal information in online social communities. We model peer disclosure as imposing a negative externality on other people. Our model encompasses the benefits from posting information, positive externalities such as recognition and entertainment benefits due to others' sharing of information, and heterogeneous privacy preferences. We find that regulation of peer disclosure is necessary. We consider two candidate regulations – nudging and quota. Nudging reduces user participation and privacy harm and sometimes improve social welfare. By contrast, imposing a quota often improves user participation, privacy protection and social welfare. Adding a nudge on top of a quota does not bring additional benefits. We show that any regulation that uniformly controls the disclosure of sensitive and non-sensitive information will not serve the triple objectives of reducing privacy harm, increasing social welfare, and increasing information contribution. We derive a necessary condition for solutions that can fulfill these three objectives. We also compare the incentives of the platform owner and social planner and draw related managerial and policy implications.

Keywords: peer disclosure; privacy; regulation; online social communities; nudging; quota

†Department of Technology and Operations Management, Rotterdam School of Management, Erasmus University, 3062 PA Rotterdam, The Netherlands. ‡Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. Cao: cao@rsm.nl, Hui: klhui@ust.hk, Xu: hxu@ust.hk.

I. Introduction

User-generated content is rife on social networking websites such as Facebook and Youtube. Such content sometimes contains information (generally referring to any text, voice recording, image, and video) about other people ("peers"), which can bring unintended fame or consequences. For example, a high school teacher was sacked in 2013 after her student posted a picture of her providing alcohol and condoms in a prom after-party on Instagram (New York Daily News 2013). A 15-year-old boy in Quebec was abused after a video of him playing a character in Star War went viral on the Internet. The boy made the video for a school project but the video was shared by his friends on the Internet without his consent (New York Times 2003). Many children today are upset because their parents share their personal pictures or videos online (New York Times 2016). Generally, a person may inflict privacy harm on other people by disclosing their personal information with the public (DiMicco and Millen 2007; Tufekci 2008; Henne and Smith 2013; Choi et al. 2015).

From a social welfare point of view, the privacy harm from peer disclosure should be balanced against the disclosure benefits. People enjoy social interaction and sharing interesting moments with friends. Friendly disclosure such as birthday greetings or achievement recognitions can bring joy to a social community. Practically, it is difficult for a social community to avoid mentioning related people in its conversations or exchanges. Hence, the pressing issue is to help users interact effectively without excessively infringing other peoples' privacy in online social communities.

The current privacy practice in online social communities mainly targets users' voluntary disclosure of their *own* information. For instance, Facebook offers users an option to restrict access to their information, including posts, profiles, and photos by other users. Users can also select whose information to view in their own timelines and remove tags about themselves from related posts and photos.[1] However, these controls do not address peer disclosure, where a user's privacy is infringed because her friends post information about her without providing her with sufficient controls over the disclosed information. Therefore, we face a novel challenge: *How does the privacy externality arising from peer disclosure of personal information affect the development of online social communities? Should we impose new policies to reduce the harm due to peer disclosure? If so, how should we design*

[1] For a comprehensive review of Facebook's privacy options, see `https://www.facebook.com/help/325807937506242/` (accessed May 2017).

the policies? Would community owners prefer such regulation?

To address these questions, we develop a stylized model that captures users' strategic decisions when they share information in an online social community. In our model, a fraction of users always participate in the community due to psychological commitment, membership, or other altruistic motivations (Hosanagar et al. 2010; Bateman et al. 2011). The other users are not committed and will strategically decide whether to join the community, taking into account the expected benefits from posting information, positive externalities from viewing posts containing others' personal information, and privacy harm resulting from the disclosure of their personal information by other people. A non-committed user would join the community if and only if she receives a higher utility from participation than staying out. The users differ in their sensitivity towards privacy.

With this model, we characterize the impacts of peer disclosure, viz. how it affects users' decisions to join the community and post information about other people. In particular, we seek economic policies that motivate users to internalize the privacy harm caused by their posts. Two broad solutions prevail in the literature of negative externalities: Indirect control and direct control. Indirect control, dating back to Pigou (1920), uses an appropriate pricing scheme that charges agents for the externalities that they impose on others. Such externality pricing has been shown to be effective in areas such as environmental protection and traffic control (Cropper and Oates 1992). A common implementation of externality pricing is to impose a tax, where the tax rate is set such that the agents would choose the efficient levels of externalities (Vickrey 1963; Sandholm 2002, 2005). In the online social community setting, we propose "nudging" as an alternative form of externality pricing. A nudge is a soft paternalistic measure that operates as a cue to remind users of the potential privacy damage that their posts could bring to others, or as an extra time delay in the form of a "cooling-off" period for users to consider withdrawing their posts. The purpose is to "nudge" users to think carefully about the privacy consequences of their posts (Acquisti 2009; Wang et al. 2013; Almuhimedi et al. 2015). A nudge has essentially a similar effect as a Pigouvian tax, where the "tax" here is non-monetary but exhibited in the form of additional time or effort in posting each piece of information.[2]

To directly control negative externalities, the classical approach is to use command-and-control regulations that directly restrict the agents' actions (Fullerton and Metcalf 2001). A typical imple-

[2] Unlike a monetary tax, the cost due to a non-monetary tax, such as a nudge, cannot be recovered and hence becomes a deadweight loss to the society.

mentation is to impose a quota or provide an allowance to each agent with the objective that the agent will generate the efficient level of externalities (Copes 1986; Calthrop and Proost 1998). Imposing a quota in online social communities is straightforward. We simply need to set a limit on the number or length of posts allowed for each user within a given time period.

The nudging and quota policies, corresponding to externality pricing and command-and-control regulation of negative externalities, are inline with the practices adopted in many industries, including cigarette and alcohol taxes, pollution permit, and road space rationing. Besides externality pricing and command-and-control regulation, the prior literature has advanced other solutions to address negative externalities, including subsidies for abatement, marketable permits, and deposit refund systems (Stavins 2011). These solutions, however, may not be applicable in online social communities, which mostly feature large numbers of users and hence the exchange of user permits is practically infeasible. It is also difficult to provide subsidies or request for deposits as most online social communities do not charge any fees to users. Another approach to address the peer disclosure externality is to deploy new technologies. For instance, using text and image processing and advanced data analytics, online social community owners may attempt to distinguish sensitive from non-sensitive personal information and directly regulate a user's disclosure of sensitive information about other people. We analyze these technical solutions in Section 4.1.

We find several unique results on nudging and quota. A nudge decreases user participation and contribution of information, but it also decreases the total privacy harm and sometimes increase social welfare by driving some users out of the community. By contrast, a quota preserves users' incentive to join the community and always increases social welfare, but it cannot encourage information contribution either. These findings exemplify the conflicting goals of enhancing social welfare and privacy protection vis-à-vis promoting community development in terms of increasing participation and information contribution. Our model provides a novel theoretical framework for analyzing the optimal policy designs in regulating online information contribution and peer disclosure.

We also find that quota dominates nudge in increasing user participation and social welfare. Contrary to the prior literature which suggests that a composite measure is more effective in addressing externalities (Roberts and Spence 1976; Christiansen and Smith 2012), we find that nudging users on top of a quota does not bring additional benefits. Furthermore, although the social planner and community owner may variously benefit from imposing a quota, they mostly prefer different quotas

because of misaligned objectives. They may prefer the same quota only when the community owner wants to grow its community size in terms of user participation. If it wants to maximize information contribution by participating users, then it will never prefer a nudge or quota. Based on this result, we derive a general condition that is necessary for any economic policy to reduce privacy harm and increase social welfare while increasing overall information contribution.

Our contributions are three-fold. First, we show that regulation is necessary when users can freely post information about other people in an online social community. To our knowledge, this is the first analysis addressing the privacy harm caused by peer disclosure on the Internet. Second, we illustrate the nuanced impacts of imposing a nudge and quota, particularly their implications on user participation, which has not been formally analyzed in prior studies (Schulze and d' Arge 1974; Weitzman 1974; Collinge and Oates 1982). Third, we uncover a novel dilemma, viz. welfare maximization and privacy protection are not aligned with community development. We provide some suggestive directions to resolve this dilemma, such as tailoring the nudge and quota for privacy-infringing posts or facilitating users to prune sensitive information related to them.

The rest of this paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the model and analyzes the impacts of imposing a nudge and quota. Section 4 derives a necessary condition for solutions that reduce privacy harm without sacrificing information contribution. Section 5 illustrates the ideas in this paper using a numerical example. Section 6 analyzes three extensions. Section 7 discusses the implications of this research and concludes the paper.

II. Related Literature

This study is closely related to the emerging stream of research that studies how peer disclosure affects consumers and that suggests possible remedial actions. Choi et al. (2015) study how embarrassing posts by friends in online social networks affect individuals' perceptions of social relationship and their consequent behavioral responses. Several studies have proposed measures to help consumers remove information shared by others without their consents (Besmer and Lipford 2010; Henne and Smith 2013). In general, the solutions involve identifying the shared information (e.g., by facial recognition technologies) and helping affected users to negotiate with the parties posting the information (e.g., by requesting for removal of infringing photos). Such solutions apply *ex post*, i.e., after the information has already been posted. Hence, they are inadequate because the damage is inflicted once the information is available on a public domain. An ideal solution should apply *ex ante*, i.e., it should incentivize people

not to haphazardly post information about others. This is the focus here.

A large body of research has studied voluntary disclosure of personal information (see, e.g., Gross and Acquisti 2005; Dwyer et al. 2007; Acquisti and Gross 2009) and its regulation (Hermalin and Katz 2006; Hui and Png 2006). This literature has variously advocated the use of "privacy nudges" (Acquisti 2009; Wang et al. 2013; Almuhimedi et al. 2015), which can be visual cues about the potential audience of a post, a time delay before the post is published, or a feedback on the potential sentiment and sensitivity of the post. The essential idea is to nudge users so that they will think twice about the privacy consequences of their posts. The focus of this literature lies in protecting consumer privacy in an online environment and the economic efficiency of information disclosure. It does not address the externalities due to information disclosure.

Prior research on privacy externalities mostly focus on marketing activities (Anderson and de Palma 2009; Anderson and Gans 2011; Johnson 2013). Seller marketing imposes a direct externality on consumers by either congesting consumers' attention span to process marketing promotions or increasing their costs to read or process the marketing. This literature has proposed solutions to help consumers address the externality. For example, Van Zandt (2004) shows that increasing senders' transmission costs using tax or other technical measures can help increase the welfare of receivers (consumers) and benefit all senders. By considering consumers' privacy harm due to seller solicitations, Hann et al. (2008) find that it is optimal to impose a charge on seller solicitations. Motivated by these suggestions, we analyze nudging as one candidate economic policy to regulate third-party externalities from peer disclosure (cf. second-party externalities from sellers).

More broadly, the economics literature has extensively analyzed the impacts of imposing a Pigouvian tax and a limit on the externality-generating activities in various contexts featuring production or consumption externalities, such as air and water pollution, smoking, and alcohol consumption (Weitzman 1974; Baumol and Oates 1988; Cropper and Oates 1992; Pizer 2002). An important consideration in this literature is entry and exit. A tax penalizes a firm and hence may force the firm to leave the industry in the long run, which can contract the industry and dampen social welfare (Schulze and d' Arge 1974; Collinge and Oates 1982; Cropper and Oates 1992). In our setting, the privacy nudge resembles a Pigouvian tax. Hence, it is important that we endogenize users' participation decisions in studying the regulation of peer disclosure in online social communities.

By contrast, limiting the externality-generating activities may have a smaller impact on participa-

tion. We consider the use of a quota as an alternative economic policy to cap or limit the externalities due to peer disclosure. Prior research has also shown that one single policy, such as imposing a tax alone, may not differentiate activities that generate different degrees of externalities. Hence, adding a direct control of the externality-generating activities may further enhance social welfare (Roberts and Spence 1976; Bennear and Stavins 2007; Christiansen and Smith 2012). For example, to address the externality due to smoking, we can apply a cigarette tax and concurrently restrict the number of outlets or limit the opening hours of outlets that sell cigarettes. We adopt a similar idea and analyze the merit of combining a nudge and a quota in this paper.

Finally, our work is related to studies of negative network externalities (Liebowitz and Margolis 1994), such as the congestion externality due to free-riding in peer-to-peer file-sharing networks (Asvanund et al. 2004). The peer disclosure externality differs from congestion externalities in that it is directly imposed at the individual user level instead of the community level. Hence, we must account for the size of user community in analyzing its impact and regulation.

III. The Model

Consider a unit mass of users who can participate and post information in an online social community. $1-\alpha$, $0<\alpha<1$, of these users are "committed" and always participate in the community. The other α users are "non-committed" and will participate if and only if they receive a higher utility from participation than staying out.[3] Among all committed and non-committed users, β, $0<\beta<1$, have high privacy sensitivity ("high types") and $1-\beta$ have low privacy sensitivity ("low types").

Each user is connected to some "peers" in the social community. A connection can be interpreted as a friendship link. We refer to a user's connected peers as "friends" and unconnected peers as "non-friends". As is the case with popular social networking websites such as Facebook or LinkedIn, the connections are undirected, i.e., two users accept each other as a friend once a connection is established. Each user i has a probability of $n_i \in [0,1]$ to establish a connection with any other user. Note that n_i can also be interpreted as the number of friends of user i because we normalize the total mass of users to be 1. We use N_i to denote the set of user i's friends and $\overline{N}_i$ as the set of user

[3] Individuals may participate in online social communities because of undisclosed self-interests, psychological commitments, or other altruistic motivations (Bateman et al. 2011). The way we model heterogeneity in user participation resembles the distinction between "altruistic" and "strategic" nodes in peer-to-peer (P2P) media distribution networks in Hosanagar et al. (2010).

i's non-friends. For a large population of users, n_i is very small. This is consistent with the case of Facebook, which has more than 700 million users but more than 95% of them have fewer than 1,000 friends (Backstrom 2011). We start with a simple setup where every user has the same number of friends, i.e., $n_i = n$ for all i. We relax this assumption in one extension later.

We assume the users' types are evenly distributed in each user's friends and non-friends networks. Hence, for any user i, both N_i and $\overline{N}_i$ contain a proportion α of non-committed users and a proportion β of high type users. Figure 1 depicts the composition of the population and the connection of user i in a network with 18 users, where $\alpha = 1/2$, $\beta = 1/3$, and $n_i = 1/3$.

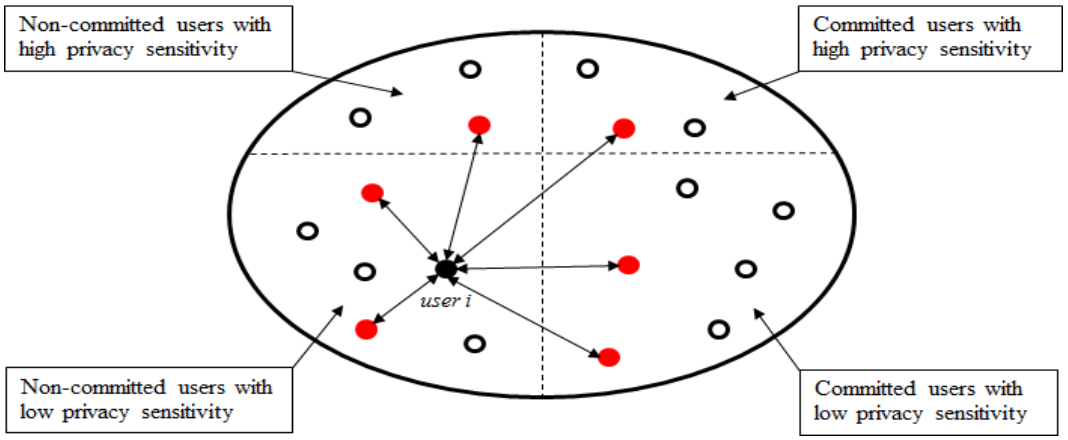

Figure 1: An example of user population and connection: total population = 18, $\alpha = 1/2$, $\beta = 1/3$, and $n_i = 1/3$.

Each participating user can post sensitive and non-sensitive information about other people. Let x_{it} and y_{it} be the amount of non-sensitive and sensitive information that user i posts about user t, $i \neq t$. The posting of sensitive information imposes a negative externality ("privacy harm") on user t. Evidently, each piece of sensitive information could cause different degrees of privacy harm, which will likely follow some statistical distribution. Without loss of generality, We use ρ_H (ρ_L) to denote the expected privacy harm that a high (low) type user suffers from each piece of sensitive information about her. We assume a participating user can post information about any other users, including non-friends and non-participating users. In practice, Facebook users can share posts or photos about anyone, including celebrities who do not have a Facebook account.

A user receives a unit benefit, v, from posting each piece of information about other people. The benefit can come from the gratification of being perceived as knowledgeable, or tangible gains from

advertising if the information attracts high viewership (e.g., garnering a large number of "likes" on Facebook). The cost for posting a piece of information, including the time and effort to acquire, edit, and upload it, varies according to the types of information and connection. We use $C_x(x_{ij}) = \frac{1}{2}c_x x_{ij}^2$, $C_y(y_{ij}) = \frac{1}{2}c_y y_{ij}^2$, $\frac{1}{\delta}C_x(x_{ik})$, and $\frac{1}{\delta}C_y(y_{ik})$ to denote the cost functions for posting non-sensitive and sensitive information about friends, i.e., $j \in N_i$, and non-friends, i.e., $k \in \overline{N}_i$.[4] We use j to index friends and k to index non-friends. The convex cost functions capture the increasing difficulty to collect and post information as the posting volume increases.

We assume it is more costly to post sensitive information than non-sensitive information, i.e., $c_x = c$ and $c_y = \frac{c}{\psi}$, $c > 0$ and $0 < \psi < 1$. Intuitively, people guard their sensitive information such as medical history or salary more carefully. People may also feel more uncomfortable in divulging embarrassing posts about others when their own identities are observable in the community. We assume that it is more difficult to post information about non-friends than friends, i.e., $0 < \delta \ll 1$, because of increased social distance and decreased level of trust towards non-friends. Realistically, people post more about their online friends who are likely to be their friends, relatives, classmates, or colleagues in their offline social circles (DiMicco and Millen 2007).

Besides the direct benefit from posting, a participating user also benefits from information posted by others. We use e to denote the *entertainment benefit* that a user enjoys from reading a piece of information unrelated to her posted by others (e.g., many people enjoy gossips about celebrities shared by others on Facebook). Similarly, we use w to denote the *recognition benefit* that a user enjoys when a piece of her non-sensitive information is posted by others (e.g., a person may enjoy pride when other people share the news that he/she has won an award).

Let s be the set of participating users. User i's expected utility from participation,

$$u_{i|s}^{in} = \int_{j\in N_i} [v(x_{ij}+y_{ij}) - C_x(x_{ij}) - C_y(y_{ij})]dj + \int_{k\in\overline{N}_i} \left[v(x_{ik}+y_{ik}) - \frac{1}{\delta}C_x(x_{ik}) - \frac{1}{\delta}C_y(y_{ik})\right] dk + e\int_{m\in s, m\neq i}\left[\int_{t\neq i}(x_{mt}+y_{mt})dt\right]dm + w\int_{m\in s,m\neq i} x_{mi}dm - \rho_i\int_{m\in s,m\neq i} y_{mi}dm. \quad (1)$$

The first two integrals are user i's expected benefits from posting about her friends and non-

[4] The linear benefit and quadratic cost functions give rise to diminishing marginal utility, which is a common feature in the literature because it is mathematically tractable and often guarantees an interior solution. It also fits real online social networks well. For example, no Facebook user would post all information about every other user, perhaps because doing so is prohibitively costly.

friends. The remaining three terms capture the externalities inflicted by other users. Specifically, the third term is the entertainment benefit, the fourth term is the recognition benefit, and the last term is the privacy harm.

Let $X_{\cdot i} \equiv \int_{m\in s, m\neq i} x_{mi} dm$ be the total quantity of non-sensitive information related to user i, $Y_{\cdot i} \equiv \int_{m\in s, m\neq i} y_{mi} dm$ be the total quantity of sensitive information related to user i, and $Q_{-i} \equiv \int_{m\in s, m\neq i} \left[\int_t (x_{mt} + y_{mt}) dt\right] dm$ be the total quantity of information posted by all participating users except user i.[5] With these notations, $\int_{m\in s, m\neq i} \left[\int_{t\neq i} (x_{mt} + y_{mt}) dt\right] dm \equiv Q_{-i} - X_{\cdot i} - Y_{\cdot i}$. Equation (1) can be rearranged as

$$\begin{aligned} u_{i|s}^{in} =& n_i \left[v x_{ij} - \frac{c x_{ij}^2}{2} + v y_{ij} - \frac{c y_{ij}^2}{2\psi} \right] + (1 - n_i) \left[v x_{ik} - \frac{c x_{ik}^2}{2\delta} + v y_{ik} - \frac{c y_{ik}^2}{2\delta\psi} \right] \\ &+ e Q_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i}, \end{aligned} \tag{2}$$

where $\omega = w - e$ and $\theta_i = \rho_i + e$ represent the recognition benefit and privacy harm *net of the entertainment value* for each piece of information. Without loss of generality, we assume $\omega > 0$, $\theta_H = 1$, and $0 < \theta_L < 1$. We refer to $eQ_{-i} + \omega X_{\cdot i}$ as the "positive externalities" that user i receives from *all* information posted by other users, and $\theta_i Y_{\cdot i}$ as the privacy harm that user i suffers from information posted about her by other users.

Note that participating users can post information about non-participating users. We assume that a non-participating user is affected by the externalities caused by the information shared in the community. Realistically, people may get exposed to information that goes viral in other media, and celebrities can be defamed by information posted in an online community even if they are not its members. In our model, conditional on s, user i's utility from staying out,

$$u_{i|s}^{out} = \epsilon \left(e Q_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i} \right), \tag{3}$$

where $\epsilon \in (0, 1)$ captures how easy it is for an outsider to get exposed to (or become aware of) information posted within the community. The larger ϵ is, the easier the information posted in

[5] Because the size of user population is a continuous measure, including user i's contribution (which is just a point in the integral) does not affect the aggregate sum, Q_{-i}. Continuous measures of user population is quite common in the literature (see, e.g., Daughety and Reinganum 2010; Conitzer et al. 2012). It reflects the realistic assumption where one single agent's decision will not affect the collective outcome of the population.

Table 1: Notations

α:	The fraction of non-committed users
β:	The fraction of users with high privacy sensitivity
n_i:	The fraction of population connected with user i, and $n_i = n$ in the main model
v:	Benefit from posting each unit of information
c:	Cost coefficient of posting non-sensitive information
ψ:	Cost ratio between posting sensitive and non-sensitive information
δ:	Cost ratio between posting information about friends and non-friends
γ:	Weighted cost ratio between posting about friends and non-friends, $\gamma = n + (1-n)\delta$
e:	Entertainment benefit from each unit of information posted by others
w	Recognition benefit from each unit of non-sensitive information posted by others
ω:	Net recognition benefit from each unit of non-sensitive information posted by others, $\omega = w - e$
ρ_i :	User i's privacy harm from each unit of sensitive information posted by others, $i \in \{L, H\}$
θ_i :	User i's net privacy harm from each unit of sensitive information posted by others, $i \in \{L, H\}$, $\theta_i = \rho_i + e$
ϵ:	The degree of information posted within community being exposed to outsiders
λ_i:	User i's average net privacy harm from each unit of information posted in the community, $i \in \{L, H\}$
τ:	Unit cost due to nudging
Λ:	Posting limit or quota

the community spreads outside the community. We can interpret $1 - \epsilon$ as the difference between the objective and subjective (perceived) information externalities, including privacy harm, faced by a non-participating user i due to the information posted in the community. Hereafter, we refer to measures scaled by ϵ as "perceived" measures and unscaled ones as "objective" measures. For example, $\theta_i Y_{\cdot i}$ is objective privacy harm whereas $\epsilon\theta_i Y_{\cdot i}$ is perceived privacy harm. Table 1 summarizes the key notations.

We study a three-stage game as shown in Figure 2. Stage 1 defines the environment, particularly whether the posting of information is regulated. In Stage 2, non-committed users decide whether to join the community. In Stage 3, participating users decide how much information to post.

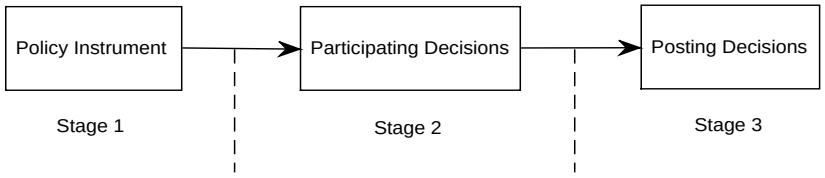

Figure 2: Timing of the Model

A. The Status Quo

We first consider a baseline setting where users can freely make participation and posting decisions without accounting for the privacy harm they inflict on others. We use backward induction to derive

the subgame perfect equilibrium. Conditional on joining the community, user i's posting decisions, x_{ij}, x_{ik}, y_{ij} and y_{ik} can be obtained from the first-order conditions of equation (2):

$$x_{ij}^{sq} = \frac{v}{c},\ y_{ij}^{sq} = \frac{\psi v}{c},\ x_{ik}^{sq} = \frac{\delta v}{c},\ y_{ik}^{sq} = \frac{\delta \psi v}{c}. \tag{4}$$

The user would simply post information based on the ratios of posting benefits over posting costs when she omits the privacy harm caused by her. Her own privacy sensitivity will not affect her posting decisions.

However, user i's participation decision depends not only on her own posting, but also on the quantity of information others post about her. We consider a rational expectations equilibrium in which users can anticipate other users' posting decisions. Substituting (4) into (2), user i's conditional utility from participation,

$$u_{i|s}^{sq,in} = \frac{\gamma(1+\psi)v^2}{2c} - \frac{\gamma(1+\psi)vs\lambda_i}{c},\ i \in \{L, H\}, \tag{5}$$

where $s \in [1-\alpha, 1]$ because committed users always participate in the community. To simplify the exposition, we let $\gamma \equiv n + (1-n)\delta$ (the weighted cost ratio of posting information about friends and non-friends) and $\lambda_i \equiv \frac{\psi\theta_i - e(1+\psi) - \omega}{1+\psi}, i \in \{L, H\}$.

The first term in (5), $\frac{\gamma(1+\psi)v^2}{2c}$, is user i's total benefit from posting. The second term, $\frac{\gamma(1+\psi)vs\lambda_i}{c}$, is the net privacy harm (i.e., privacy harm net of entertainment and recognition benefits) that user i suffers from participating in the community. In the second term, $\frac{\gamma(1+\psi)v}{c} = n(x_{ij}^{sq} + y_{ij}^{sq}) + (1-n)(x_{ik}^{sq} + y_{ik}^{sq})$ is the amount of information posted by each participating user, and so $\frac{\gamma(1+\psi)vs}{c}$ is the total quantity of information posted by the entire community. Therefore, we can interpret λ_i as user i's average net privacy harm caused by each piece of information posted in the community.[6]

User i's utility from staying out of the community follows equation (3), i.e.,

$$u_{i|s}^{sq,out} = -\epsilon \cdot \frac{\gamma(1+\psi)vs\lambda_i}{c},\ i \in \{L, H\}. \tag{6}$$

[6] A random piece of information may or may not be privacy-infringing to user i. λ_i is the expected privacy harm if a random piece of information is sensitive and related to user i minus the expected entertainment and recognition benefits otherwise.

We impose the following regularity assumption.

Assumption 1 $\theta_L > e(1+\frac{1}{\psi})+\frac{\omega}{\psi}$.

$\theta_L \leq e(1+\frac{1}{\psi})+\frac{\omega}{\psi}$ is equivalent to $\lambda_L \leq 0$, which means that all low type users will participate in the status quo because their net privacy harm is negative (meaning they benefit from other peoples' posting). Assumption 1 enables us to focus on a more realistic scenario where the privacy concern is sufficiently salient to hinder some users from participating in the community.

Given s, user i will participate if and only if $u_{i|s}^{sq,in} \geq u_{i|s}^{sq,out}$. Let s_L (s_H) be the participation rate for non-committed low (high) type users. The following results characterize users' equilibrium participation as the posting benefit, v, varies.

Lemma 1 *In the status quo equilibrium, non-committed users will participate according to the following schedule.*

v	s_L^{sq}	s_H^{sq}	s^{sq}
$(0, 2(1-\epsilon)(1-\alpha)\lambda_L)$	0	0	$1-\alpha$
$[2(1-\epsilon)(1-\alpha)\lambda_L, 2(1-\epsilon)(1-\alpha\beta)\lambda_L]$	$\frac{v/(2(1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$	0	$\frac{v}{2(1-\epsilon)\lambda_L}$
$(2(1-\epsilon)(1-\alpha\beta)\lambda_L, 2(1-\epsilon)(1-\alpha\beta)\lambda_H)$	1	0	$1-\alpha\beta$
$[2(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)\lambda_H]$	1	$\frac{v/(2(1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$	$\frac{v}{2(1-\epsilon)\lambda_H}$
$(2(1-\epsilon)\lambda_H, +\infty)$	1	1	1

Notes. $s^{sq} = (1-\alpha)+\alpha(1-\beta)s_L^{sq}+\alpha\beta s_H^{sq}$, i.e., the sum of all committed and non-committed low and high type users.

In the status quo, participating users ignore the privacy harm inflicted on others. When the posting benefit, $v < 2(1-\epsilon)(1-\alpha)\lambda_L$, both low and high type non-committed users prefer to stay out because the privacy harm outweighs the benefit from posting information. As v increases, the users will gradually participate by order of privacy sensitivity, and the participate rate increases with v. When v is sufficiently large, all users participate in the community.

Let Π be the social welfare, defined as the aggregate surplus of all users including the *perceived* privacy harm suffered by all non-participating users, Q be the total quantity of information posted in the community, and ξ be the total objective privacy harm including the harms inflicted on non-participating users.[7] The next lemma characterizes the outcomes in the status quo. Figure 3 illustrates how the outcomes in Lemmas 1 and 2 vary with v.

[7] We present the total perceived privacy harm in the Online Supplement. The analysis of objective and perceived privacy harms give similar qualitative insights.

Lemma 2 *In the status quo, the total quantity of information posted, Q, and total privacy harm, ξ, increase with v. Social welfare is negative if and only if (i) $0 < v < 2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H]$, or (ii) $2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H] < v < 2\bar{\lambda}$ and $\beta > \frac{(1-\epsilon)\lambda_H - \lambda_L}{\lambda_H - \lambda_L}$.*

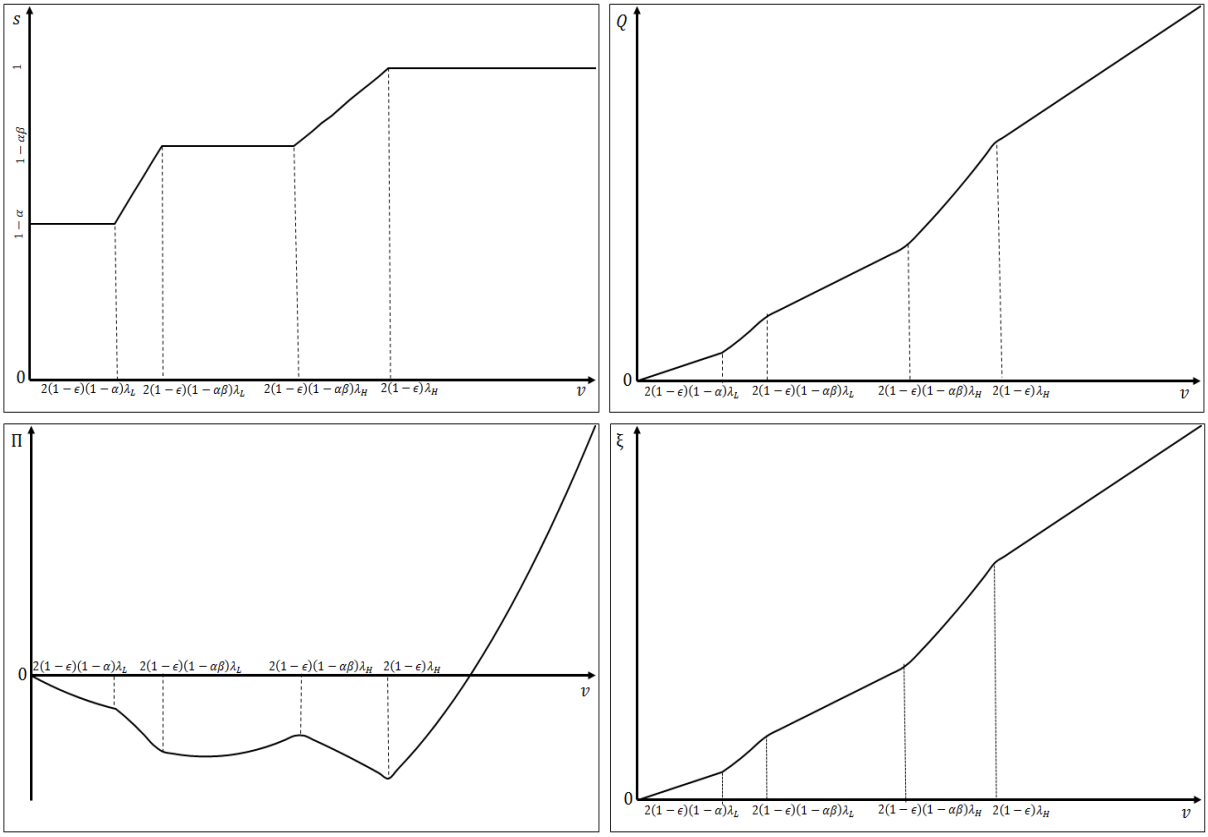

Figure 3: Equilibrium outcomes in the status quo.

Social welfare can be negative due to two reasons. Non-participating users carry negative utility since they are also (partially) affected by the privacy harm generated in the community. When the proportion of non-participating users increases as the posting benefit v decreases, the overall negative utility outweighs the positive utility from some of the participating users, therefore leading to negative social welfare. As posting benefit v increases, high type users could still have large negative utility even after they participate, which is because the positive utility from posting is lower than the privacy harm imposed on them. This could also lead to negative social welfare when there are many such highly privacy sensitive users, i.e., β is high enough.

Referring to Figure 3, one interesting observation in the status quo is that, when the social welfare is negative, a slight increase in v may further decrease social welfare. An increase in posting benefit v

induces users to post more information and more non-committed users to join the community and post information. Such additional information brings negative marginal benefit due to its privacy harm. The implication is that in some online social communities, having more users or encouraging users to post more information involving peers can be bad. Indeed, many people post unverified gossips on the Internet. Recent research has shown that online social communities or, more broadly, the Internet can help propagate wicked materials to support racial hatred, political flaming, or cyber-bully (see, e.g., Davis 1999; Keith and Martin 2005; Kowalski and Limber 2007; Bhuller et al. 2013; Chan et al. 2016).

As a benchmark, we next calculate the first-best posting decisions, in which we assume the users take into account of the externalities from their posting behavior. Without loss of generality, we let $\epsilon = 1$ and derive the first-best posting strategies as:

$$x_{ij}^{fb} = \frac{v + e + \omega}{c},\ y_{ij}^{fb} = \max\left\{0, \frac{\psi(v - \theta_j)}{c}\right\},\ x_{ik}^{fb} = \frac{\delta(v + e + \omega)}{c},\ y_{ik}^{fb} = \max\left\{0, \frac{\delta\psi(v - \theta_k)}{c}\right\}. \quad (7)$$

A simple comparison of the above posting amount with those in the status quo, as shown in equations (4) shows that users in the first best case post more non-sensitive information and less sensitive information than in the status quo, and the amount of sensitive information is contingent on the privacy sensitivity of the posted subjects.

In the following sections, we analyze a few policies and compare their impacts on users' decisions. We focus on two policies, nudge and quota, that are highly feasible to implement because these policies uniformly apply to both sensitive and non-sensitive information. We then extend the analysis to other policies, including targeted nudge, targeted quota, and information pruning, that differentiate between the two types of information, and these policies may become feasible with future advancement in technology (such as machine learning and artificial intelligence). We also provide implications for each policy from the perspectives of the social planner and the community owner.

B. Nudge

We first study the use of a privacy "nudge" as a regulation policy (Acquisti 2009; Wang et al. 2013; Almuhimedi et al. 2015). The online social community can remind users of the potential privacy infringements that their posts may cause on other people through, e.g., visual cues or warning messages. The community can also introduce an extra time delay before a post is really publicized to

allow users a "cooling-off" period during which they can always revoke the post. Such privacy nudges increase users mental efforts needed to post information and offer a chance for them to reconsider their decisions. We model the privacy nudge as imposing an additional linear posting cost on users, $\tau \in [0, v]$.[8] Intuitively, if the nudge exceeds users' posting benefit, v, then users would obtain negative benefit from posting and hence no user would post any information.

Evidently, because the nudge applies to all users and information, it increases the overall cost of posting information in the online social community. Highly drastic as it seems, mechanisms similar to a non-discriminatory nudge are commonly observed in practice. For example, one renowned solution to combat music or movie piracy is to impose a tax on all blank storage media such as CD or DVD even though they are often used for legitimate purposes such as data storage. Bill Gates has famously suggested an email tax to curb spam, which inevitably affects all legitimate uses of email. We add the superscript n to all variables in the setting with a nudging policy.

With nudging, user i's utility from participation becomes:

$$\begin{aligned} u_{i|s}^{n,in} =& n_i \left[v x_{ij} - \frac{c x_{ij}^2}{2} + v y_{ij} - \frac{c y_{ij}^2}{2\psi} - \tau(x_{ij} + y_{ij}) \right] + (1 - n_i) \left[v x_{ik} - \frac{c x_{ik}^2}{2\delta} + v y_{ik} - \frac{c y_{ik}^2}{2\delta\psi} \right. \\ & \left. - \tau(x_{ik} + y_{ik}) \right] + e Q_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i}, \end{aligned} \tag{8}$$

The first-order conditions of (8) give the following posting decisions:

$$x_{ij}^n = \frac{v - \tau}{c},\ y_{ij}^n = \frac{\psi(v - \tau)}{c},\ x_{ik}^n = \frac{\delta(v - \tau)}{c},\ y_{ik}^n = \frac{\psi\delta(v - \tau)}{c}. \tag{9}$$

Obviously, the nudge decreases the amount of information posted by participating users. However, it does not affect non-participating users. Hence, user i's utility from staying out, $u_{i|s}^{n,out}$, has the same form as equation (3). Her participating decision then depends on the comparison between $u_{i|s}^{n,in}$ and $u_{i|s}^{n,out}$. The following lemma summarizes the equilibrium participation rates.

Lemma 3 *When a nudge, $\tau \in (0, v]$, is imposed, non-committed users will participate according to the following schedule.*

By comparing Lemma 3 with Lemma 1, with nudging, a higher v is needed to encourage both

[8] The findings are qualitatively similar if we use a quadratic nudging cost.

v	s_L^n	s_H^n	s^n
$(0, 2(1-\epsilon)(1-\alpha)\lambda_L + \tau)$	0	0	$1-\alpha$
$[2(1-\epsilon)(1-\alpha)\lambda_L + \tau, 2(1-\epsilon)(1-\alpha\beta)\lambda_L + \tau]$	$\frac{\frac{v-\tau}{2(1-\epsilon)\lambda_L} - (1-\alpha)}{\alpha(1-\beta)}$	0	$\frac{v-\tau}{2(1-\epsilon)\lambda_L}$
$(2(1-\epsilon)(1-\alpha\beta)\lambda_L + \tau, 2(1-\epsilon)(1-\alpha\beta)\lambda_H + \tau)$	1	0	$1-\alpha\beta$
$[2(1-\epsilon)(1-\alpha\beta)\lambda_H + \tau, 2(1-\epsilon)\lambda_H + \tau]$	1	$\frac{\frac{v-\tau}{2(1-\epsilon)\lambda_H} - (1-\alpha\beta)}{\alpha\beta}$	$\frac{v-\tau}{2(1-\epsilon)\lambda_H}$
$(2(1-\epsilon)\lambda_H + \tau, +\infty)$	1	1	1

Notes. $s^n = (1-\alpha) + \alpha(1-\beta)s_L^n + \alpha\beta s_H^n$.

types of users to participate and post information. Not surprisingly, the nudge also leads to less information posted and hence effectively reduces the total privacy harm created by the community. The following proposition states these results formally.

Proposition 1 *Comparing with the status quo, a nudge reduces user participation, total quantity of information posted, and total privacy harm. Further, the participation rates, total quantity of information posted, and total privacy harm decrease in the level of nudge, τ.*

Intuitively, users may be annoyed by the privacy nudge (e.g., warning messages, time delay) and hence may post less information or even drop out from the community. Our result is consistent with previous research showing that many consumers are impatient and may drop out of online communities due to inconvenience (Galletta et al. 2006; Rajamma et al. 2009; Ding et al. 2015). It is worth noting that, although a nudge can effectively reduce the total privacy harm as it discourages peer disclosure, it also leads to less activity in the community and lower posting, entertainment, and recognition benefits. Its overall impact on the community is determined by the tradeoffs between these benefits and costs. The following result characterizes when a nudge is socially preferred.

Proposition 2 *(i) A nudge can improve social welfare when social welfare is negative in the status quo. The socially optimal nudge is $\tau^* = v$, which gives social welfare of 0.*

(ii) A nudge always decreases social welfare when social welfare is positive in the status quo. The socially optimal nudge is $\tau^* = 0$.

As discussed in Lemma 2, allowing users to post information is socially undesirable when the social welfare is negative. Imposing a nudge can dissuade people from posting and hence improve social welfare. The optimal outcome is to nudge all users extensively so that they do not post any information. Social welfare will then increase from being negative to 0. A harsher nudge could

be viewed as a less user-friendly interface. The results here show that when privacy externalities predominate, an easier-to-use interface may actually hurt social welfare. By contrast, when social welfare is positive, allowing users to post information brings more benefits than harm. Imposing a nudge will only increase the cost to the community. Hence, the optimal nudge is $\tau^* = 0$.

The implication of the nudging analysis is that, if a person is concerned about privacy, perhaps she should simply not join the community. Incidentally, this implication is consistent with the Chicago School's view of how privacy should be treated, although here the privacy harm arising from peer disclosure is a form of negative externality that calls for regulation (Posner 1978, 1979, 1981; Stigler 1980).

Since a nudge always decreases user participation and posting of information, it obviously is not in the interest of the community owner to impose a nudge. However, a social planner such as the government cares more about social welfare and privacy. So, the social planner prefers nudging for all v specified in Condition (i) of Proposition 2 because it helps achieve higher social welfare and lower total privacy harm. Such discrepancy in objectives helps explain why most online social communities today do not alert users about the potential adverse consequences of peer disclosure.

C. Quota

We next consider the merit of having a quota, Λ, which is a limit on the total quantity of information that a user can post in the community. We assume $\Lambda \in (0, \frac{\gamma(1+\psi)v}{c}]$. Recall from equation (4) that a user would post $\frac{\gamma(1+\psi)v}{c}$ units of information in the status quo. Hence, when $\Lambda > \frac{\gamma(1+\psi)v}{c}$, the quota is not binding. We refer to any $\Lambda \in (0, \frac{\gamma(1+\psi)v}{c}]$ as an "effective quota" as it will affect the user's equilibrium behavior. We say that the quota is "ineffective" otherwise.

We add the superscript q to all variables in the setting with a quota. Conditional on participation, the user's posting decisions are now subject to an additional constraint

$$n(x_{ij}^q + y_{ij}^q) + (1-n)(x_{ik}^q + y_{ik}^q) \leq \Lambda. \tag{10}$$

We compute x_{ij}^q, y_{ij}^q, x_{ik}^q and y_{ik}^q by solving equation (2) with constraint (10). Because the utility function in (2) is concave and Λ can not be greater than its unique interior solution, $\frac{\gamma(1+\psi)v}{c}$, a participating user will always use up the quota, meaning constraint (10) is binding, or $n(x_{ij}^q + y_{ij}^q) + (1-n)(x_{ik}^q + y_{ik}^q) = \Lambda$. Similar to the status quo, with an effective quota, the equilibrium quantities

of information are determined by the corresponding posting costs:

$$x_{ij}^q = \frac{\Lambda}{\gamma(1+\psi)},\ y_{ij}^q = \frac{\psi\Lambda}{\gamma(1+\psi)},\ x_{ik}^q = \frac{\delta\Lambda}{\gamma(1+\psi)},\ y_{ik}^q = \frac{\delta\psi\Lambda}{\gamma(1+\psi)}. \tag{11}$$

In sharp contrast to nudging which affects the community development in obvious ways (refer to Proposition 2, the platform owner prefers not to nudge any participating users), a quota does not impose any additional cost on users. Hence, it is not obvious whether the community owner's and social planner's interests are aligned. In the following analysis, we examine the optimal quota in terms of user participation, total quantity of information posted, and social welfare.

C.1. Participation-Maximizing Quota

We first consider user participation. For online communities, particularly those at an early stage of development, having a large user base is critical to triggering network effects among users and seeking financial support from venture capitals. We use the superscript $\star$ to denote the optimal outcomes when the objective is to maximize the number of participating users. Let ι be an infinitesimally small positive number. The following proposition characterizes the optimal quota.

Proposition 3 *Imposing a quota to the status quo will weakly increase the number of participating users. The participation-optimal schedule of quota is:*

(i) When $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$*, the optimal quota is* $\Lambda^\star = \iota$*, which gives participation rates* $s_L^\star = \frac{(v-\iota)/((1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$*,* $s_H^\star = 0$*, and* $s^\star = \frac{v-\iota}{(1-\epsilon)\lambda_L}$*.*

(ii) When $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq \min\{(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)(1-\alpha\beta)\lambda_L\}$*, the optimal quota is* $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_L]}{c}\right)$*, which gives participation rates* $s_L^\star = 1$*,* $s_H^\star = 0$*, and* $s^\star = 1-\alpha\beta$*.*

(iii) When $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$*, the optimal quota is* $\Lambda^\star = \iota$*, which gives participation rates* $s_L^\star = 1$*,* $s_H^\star = \frac{(v-\iota)/((1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$*, and* $s^\star = \frac{v-\iota}{(1-\epsilon)\lambda_H}$*.*

(iv) When $(1-\epsilon)\lambda_H < v < 2(1-\epsilon)\lambda_H$*, the optimal quota is* $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}\right)$*, which gives participation rates* $s_L^\star = 1$*, and* $s_H^\star = 1$*,* $s^\star = 1$*.*

(v) For all other v*, imposing a quota will not improve the participation rate relative to the status quo.*

We illustrate Proposition 3 in Panel (A) of Figure 4. Non-committed users will participate only when they obtain higher utility from participation than staying out, which requires the positive posting benefits and externalities to outweigh the (objective) privacy harm. An effective quota limits the amount of information that each user can post. This decrease in information causes the privacy harm to decrease faster than the positive benefits, which tends to encourage users to participate in the community. As shown in Panel (A) of Figure 4, both types of users are now willing to participate with a lower posting benefit, v, when compared with the status quo.

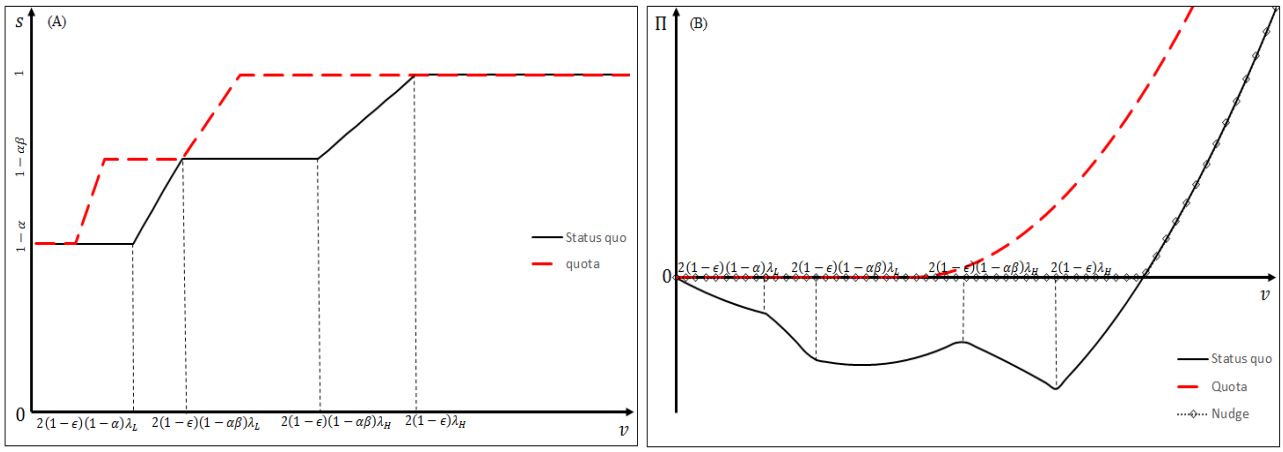

Figure 4: Participation and welfare comparison.

C.2. Information-Maximizing Quota

We next consider the total quantity of information. Some online communities rely on user activities to generate revenues. For example, Facebook places contextual advertisements related to user feeds. Some specialized communities such as cyber-lockers obtain advertising revenues by capitalizing on users' sharing of digital materials of broad interests. The following proposition characterizes the impact of the quota when the objective is to maximize total information contribution.

Proposition 4 *Imposing any effective quota,* $\Lambda < \frac{\gamma(1+\psi)v}{c}$*, will decrease the total quantity of information posted in the community.*

A quota does not increase information contribution because it limits the amount of information that a participating user can post. Although it helps attract more users to participate, the incremental gain in information due to these marginal users does not outweigh the loss due to the reduction of contribution from *every* user. Overall, Proposition 4 suggests that imposing a quota will not help an

online social community in terms of increasing information contribution.

C.3. Welfare-Maximizing Quota

We now consider social welfare. For ease of exposition, we present the optimal quota in Appendix A. The following proposition summarizes its impact.

Proposition 5 *Imposing a quota to the status quo will increase social welfare. The socially-optimal quota presented in Appendix A weakly reduces the aggregate privacy harm and increases the number of participating users.*

Recall from Proposition 2 that a nudge can improve social welfare only by nudging users out of the community. Here, a quota encourages user participation but reduces the quantity of information posted by each user. More users will join the community and contribute information but, by Proposition 4, the total quantity of information will always decrease. This implies that both the net benefit from posting information and privacy harm will decrease for each user. Nevertheless, by the utility specification in (2), the net benefit of positing information will decrease at a slower rate than the privacy harm, leading to an overall improvement in social welfare.[9]

Furthermore, Lemma 2 shows that social welfare is negative when the posting benefit, v, is not high enough because of excessive posting from all participating users. Imposing a quota will also help because, by choosing a sufficiently small quota, the social planner can effectively contain the privacy harm generated. As shown in Panel (B) of Figure 4, the social planner can always achieve a positive social welfare by choosing an appropriate quota.

C.4. Quota vs. Nudge

By the above analysis, it is straightforward to see that the quota is better than the nudge in terms of enhancing social welfare.

Proposition 6 *The socially optimal quota weakly dominates the socially optimal nudge in improving user participation and social welfare.*

By Propositions 1 and 3, a quota increases user participation, but a nudge decreases user participation. More importantly, the quota preserves users' incentives to post information. Panel (B) of

[9] Proposition 5 also holds if the net benefit from posting information increases linearly but the privacy harm increases exponentially with the amount of information posted by each user on another user. Realistically, the marginal privacy harm may increase when a user post more information about her friends – e.g., the accumulated information may allow others to track a person with a higher precision which poses a bigger privacy threat.

Figure 4 shows that the socially-optimal quota achieves higher social welfare than the socially-optimal nudge when v is sufficiently high. When v is small, allowing users to post information is not socially beneficial. Hence, both the quota and nudge apply restrictively to discourage user posting. It is worth noting that, as shown in Propositions 1 and 4, neither a nudge nor a quota can increase the total quantity of information posted in the community.

C.5. Quota Choice between the Community Owner and Social Planner

Interestingly, Propositions 3, 4 and 5 suggest that the community owner will never prefer a quota when it wants to maximize information contribution, but it may prefer a quota when it wants to grow the number of participating users. Will the community owner and social planner ever prefer the same quota? The answer is yes; a community owner who wants to maximize user participation may prefer the socially optimal quota. Figure 5 plots the optimal quotas that maximize the participation rate and social welfare. The optimal quotas overlap in some ranges of v, meaning they can serve both purposes. We present the formal conditions where both the community owner and social planner prefer the same quota in the Online Supplement.

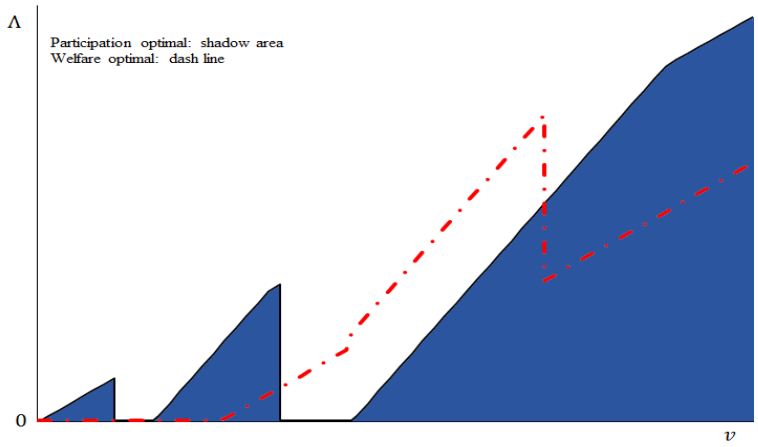

Figure 5: Participation optimal vs. welfare optimal quota.

D. Combining Nudge and Quota

We next examine if combining a nudge and a quota gives any complementary benefit. Obviously, by Proposition 2, the optimal nudge can improve social welfare only by incentivizing all users not to post any information. Adding a quota is not meaningful and will not make any difference in such a scenario. Therefore, we only consider the effect of adding a nudge on top of a quota.

Let $\tilde{\Lambda} \in (0, \frac{\gamma(1+\psi)v}{c}]$ be the free quota given to each user, beyond which a nudge $\tilde{\tau} > 0$ will be

applied to each additional piece of information posted by the user. Although the option of posting beyond the quota gives users more flexibility in posting information, the following results show that it does not bring any benefit to the community.

Proposition 7 *The optimal quota weakly dominates the composite policy of adding a nudge to the quota in terms of increasing user participation and social welfare.*

With an appropriate composite policy, i.e., $\tilde{\Lambda} < \frac{\gamma(1+\psi)(v-\tilde{\tau})}{c}$, the users can post beyond the quota if they are willing to be nudged. Such marginal posts generate more privacy harm on other people, which dissuades the sensitive users from joining the community. With fewer participating users, the community as a whole would generate less information. Hence, social welfare will decrease because of both lower user participation and less surplus received by each participating user. Intuitively, the composite policy neutralizes the benefit of the quota. The purpose of having a quota is to reduce privacy harm so that the privacy-sensitive users will find joining the community attractive. Allowing users to post beyond the quota, however, induces more privacy harm. This tends to drive the sensitive users out and hence counteracts the quota.

Note that such composite policy will not increase the total quantity of information posted in the community when compared with the status quo. This is because, by Propositions 1 and 4, both quota and nudge always decrease information contribution. Hence, any combination of them will lead to further reduction of information. Together with Proposition 7, we conclude that a composite policy cannot serve the interests of either the social planner or community owner.

IV. Solving Discord between Community Owner and Social Planner

Many online social communities generate revenues by "dollarizing" their user bases. A common business model is to serve context-based advertisements. For example, Facebook feeds advertisements to users based on their activities such as "likes", shares, and comments. The success of such context-based advertising depends greatly on whether the users are active, meaning the community owner may prefer to maximize information contribution by its users. However, our results so far show that a nudge or a quota, or their combinations, will never serve this preference of the community owner despite the fact that the social planner may prefer a nudge or quota.

In general, consider the user's utility from joining vis-à-vis staying out of the community. Sub-

tracting equation (6) from (5),

$$u_{i|s}^{sq,in-out} = \frac{\gamma(1+\psi)v^2}{2c} - (1-\epsilon)\frac{\gamma(1+\psi)vs\lambda_i}{c} = \frac{\gamma(1+\psi)v^2}{2c} - (1-\epsilon)\cdot\lambda_i\cdot Q^{sq}(s),\ i\in\{L,H\},\ (12)$$

where $Q^{sq}(s)$ is the total quantity of information posted in the community given size of participation, s. User i will join the community if and only if $u_{i|s}^{in-out}$ is non-negative. To address the privacy harm due to peer disclosure, we have to curb user posting of information about other people. This reduces the first term in equation (12). Because each user would post less now, to serve the community owner's interest in increasing the total quantity of information, we must increase the number of users participating in the community. However, given fixed $(1-\epsilon)\lambda_i$, such a requirement necessarily causes $Q^{sq}(s)$ in the second term of equation (12) to increase. Hence, taken together, $u_{i|s}^{sq,in-out}$ will decrease, meaning fewer users will want to participate in the community. This contradicts the requirement of increasing user participation.

Accordingly, any feasible welfare-maximizing solutions that also serve the community owner's interest in increasing the total quantity of information must decrease $(1-\epsilon)\lambda_i$ in the second term of equation (12). The next result formalize the necessary conditions for such solutions.

Proposition 8 *To reduce the privacy harm due to peer disclosure without decreasing the total quantity of information posted in the community, we must either reduce the average net privacy harm from posting each piece of information, λ_L or λ_H, or increase the ease for outsiders to get exposed to the information posted within the community, ϵ.*

Equation (12) and Proposition 8 are important because they highlight the intricate dilemma in the peer disclosure problem, viz. the contradictory objectives of decreasing privacy harm and maintaining information contribution. They crystallize the necessary characteristics of solutions that can address the peer disclosure problem. How can a community owner simultaneously curb peer disclosure, increase total posts, and improve social welfare? Increasing the community's visibility, ϵ, so that non-participating users are more aware of the information posted in the community is one way to entice users to join the community. However, increasing ϵ means that all non-participating users would suffer more (perceived) privacy harm, which is not desirable to the social planner. Arguably, making more non-participating people suffer is not a good way to boost participation and information contribution. Therefore, in the following discussion, we focus on policies that decrease the average

net privacy harm from each piece of information, λ_L and λ_H.

The nudge and quota analyzed in Sections B and C do not change λ_i because they penalize sensitive and non-sensitive information uniformly, causing them to decrease by the same proportion. The community owner may be able to improve the distinction of sensitive from non-sensitive information with, for example, latest photo recognition technologies or text mining and natural language processing techniques. Taking this possibility as given (i.e., we do not consider the community owner's cost of investing in technologies to detect sensitive information), we discuss some suggestive solutions to address the dilemma highlighted in Proposition 8.

A. Targeted Nudge and Quota

If distinction of sensitive from non-sensitive information is possible, then the community owner can impose a nudge or quota to target sensitive information. As a result, participating users will reduce the extent of posting sensitive information relative to non-sensitive information, reducing the overall privacy harm generated from the community. Formally, we add the superscripts "tn" and "tq" to all variables when a targeted nudge and a targeted quota are used. When sensitive information can be *perfectly* separated from non-sensitive information, equation (12) becomes

$$u_{i|s}^{tn,in-out} = \underbrace{\frac{\gamma v^2}{2c} + \frac{\gamma\psi(v-\tau)^2}{2c}}_{*} - \underbrace{(1-\epsilon)\cdot Q^{tn}(s)\cdot(\lambda_i - \zeta^{tn})}_{**},\ i \in \{L,H\},$$
$$\text{where } \zeta^{tn} = \frac{\psi\tau(\theta_i+\omega)}{(1+\psi)[v+\psi(v-\tau)]}, \tag{13}$$

and

$$u_{i|s}^{tq,in-out} = \underbrace{\frac{\gamma v^2}{2c} + v\Lambda - \frac{c\Lambda^2}{2\gamma\psi}}_{\#} - \underbrace{(1-\epsilon)\cdot Q^{tq}(s)\cdot(\lambda_i - \zeta^{tq})}_{\#\#},\ i \in \{L,H\},$$
$$\text{where } \zeta^{tq} = \frac{(\theta_i+\omega)(\gamma\psi v/c - \Lambda)}{(1+\psi)(\gamma v/c + \Lambda)}. \tag{14}$$

Note that $0 < \tau \leq v$ and $0 < \Lambda < \frac{\gamma\psi v}{c}$ because a user would post only $\frac{\gamma\psi v}{c}$ pieces of sensitive information in the status quo.

Recall $Q^{tn}(s)$ and $Q^{tq}(s)$ are the total quantities of information posted in the community given participation size, s. Comparing equations (13) and (12), $(*)$ is smaller than $\frac{\gamma(1+\psi)v^2}{2c}$ because the nudge makes posting information more costly. However, if the targeting is sufficiently accurate, i.e.,

ζ^{tn} is sufficiently large, then even if user participation increases leading to $Q^{tn}(s) > Q^{sq}(s)$, the second term in equation (13), $(**)$, can still be smaller than the second term in Equation (12), $(1-\epsilon)Q^{sq}(s)\lambda_i$. Hence, the targeted nudge reduces the average net privacy harm due to peer disclosure, meaning the contradiction highlighted by equation (12) need not exist. A similar analysis applies to a targeted quota. When ζ^{tq} is sufficiently large, a targeted quota may increase overall information contribution but suppress privacy harm and increase social welfare.

A real-life implementation of targeted nudge is the smartphone app "ReThink" (ABC News 2015). It alerts users when they try to post offensive words or phrases in social media. The core component of the app is a database of offensive trigger words and phrases. In practice, such targeted information control or nudges cannot completely accurately distinguish sensitive from non-sensitive information. They may generate false positives – non-sensitive or non-offensive information could be wrongly detected as sensitive or offensive information, or false negatives – omitting sensitive or offensive information that causes harm to others. However, as long as the targeting is sufficiently accurate, such approaches to regulating information contribution could help curb privacy harm and increase social welfare. Their deployment is aligned with the interest of community owners too because, by reducing privacy harm, more people may be willing to join online social communities which can lead to an overall increase in information contribution.

B. Information Perturbation and Pruning

In the same spirit as the series of data perturbation techniques developed to protect privacy (e.g., Li and Sarkar 2006; Menon and Sarkar 2007), another possible solution to addressing the peer disclosure problem is to help users identify and prune sensitive information (e.g., automatically blurring faces or replacing faces with emoji in photos or videos). This can directly reduce the privacy harm caused by each piece of sensitive information and thus decrease λ_i. Such information pruning may cause users to obtain less pleasure in posting information, i.e., it may decrease v. Referring to equation (12), it will decrease the user's posting benefit and net privacy harm suffered from others' posting. However, as long as the pruning of sensitive information is sufficiently accurate to the extent that it reduces the privacy harm (by reducing λ_i) more than the posting benefit (due to a decrease in v), then it can be an effective solution. Arguably, blurring or substituting the faces in a picture by non-intrusive measures could eliminate most privacy harms inflicted on the involved people without significantly hurting the poster's pleasure.

V. Numerical Example

We use a numerical example to demonstrate two results from the above discussion. First, we show that a uniform quota and uniform nudge can increase social welfare and reduce privacy harm under different levels of v as shown in Lemma 1 and Propositions 1 to 5. Second, we show that properly-constructed targeted nudge, targeted quota, and information pruning can resolve the conflict between social planner and community owner as characterized in Section IV, by effectively regulating privacy harm while increasing the number of participating users and social welfare.

Let $\alpha = 0.7$, $\beta = 0.5$, $n = 0.1$, $\psi = 0.5$, $\delta = 0.5$, $\epsilon = 0.1$, $e = 0.1$, $\omega = 0.01$, $c = 1$, $\theta_L = 0.5$, and $\theta_H = 1$. These values describe an online social community with many strategic users facing high costs of posting sensitive information and posting about strangers in a small friendship network. Users in the community enjoy some recognition benefits. Users not in the community face a small chance of being affected by the privacy externality. As specified in Lemma 1, the equilibrium outcomes in the status quo differ in five ranges of v. We choose the average v in each of these five ranges and construct five sets of outcomes in Table 2.[10] In each panel of Table 2, we compare the outcomes in the status quo with the outcomes under different regulations, including uniform nudge, uniform quota, targeted nudge, targeted quota, and information pruning. For illustrative purpose, we set the uniform nudge as 30% of the posting benefit, v, and the uniform quota as 70% of the posting volume in the status quo. We set the targeted nudge as 60% of v and targeted quota as 50% of the posting volume of sensitive information in the status quo.[11] Figure 6 plots the results under the different levels of v as shown in Table 2.

Consistent with Proposition 1, a uniform nudge (UN) reduces user participation because it makes information contribution less beneficial. The total quantity of information in the community declines, leading to less privacy harm. It may increase social welfare by driving some marginal users out as Proposition 2 suggests. However, such welfare improvement is possible only when social welfare is negative in the status quo, i.e., when $v = 0.016$ or 0.051. When social welfare is positive in the status quo, i.e., when $v = 0.168$, 0.337, or 0.458, nudging decreases social welfare.

---

[10] For the last equilibrium in Lemma 1 where both types of users participate in the community, the range of v can extend to infinity. We choose the last v in Table 2 as $2(1-\epsilon)\lambda_H + 0.5$.

[11] Note that these are not optimal nudges and quotas. The optimal nudges and quotas differ in settings with different v's. We choose these values just to illustrate that the uniform and targeted nudges and quotas indeed carry the properties and functions as analyzed in the Propositions.

Table 2: Numerical example

(i) $v = 0.016$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	0.300	0.004	-0.018	0.100
Uniform Nudge:	0.300	0.003	-0.013	0.070
Uniform Quota:	0.300	0.003	-0.012	0.070
Targeted Nudge:	0.650	0.007	-0.003	0.100
Targeted Quota:	0.650	0.007	0.004	0.088
Information Pruning:	0.300	0.004	-0.007	0.064
(ii) $v = 0.051$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	0.475	0.020	-0.074	0.503
Uniform Nudge:	0.333	0.010	-0.036	0.246
Uniform Quota:	0.618	0.018	-0.067	0.457
Targeted Nudge:	0.650	0.021	0.022	0.316
Targeted Quota:	0.650	0.021	0.055	0.278
Information Pruning:	0.650	0.025	-0.032	0.442
(iii) $v = 0.168$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	0.650	0.090	0.107	2.248
Uniform Nudge:	0.650	0.063	-0.083	1.574
Uniform Quota:	0.650	0.063	0.233	1.574
Targeted Nudge:	0.890	0.096	0.387	1.416
Targeted Quota:	1	0.105	0.757	1.396
Information Pruning:	0.740	0.092	0.279	1.643
(iv) $v = 0.337$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	0.825	0.229	1.390	5.727
Uniform Nudge:	0.650	0.126	0.580	3.159
Uniform Quota:	1	0.194	1.467	4.860
Targeted Nudge:	1	0.217	2.229	3.193
Targeted Quota:	1	0.210	3.579	2.802
Information Pruning:	1	0.250	1.994	4.457
(v) $v = 0.458$:	s	Q	Π ($\times 10^{-2}$)	ξ ($\times 10^{-2}$)
Status Quo:	1	0.378	3.237	9.446
Uniform Nudge:	0.786	0.208	1.261	5.196
Uniform Quota:	1	0.264	4.083	6.612
Targeted Nudge:	1	0.295	4.546	4.345
Targeted Quota:	1	0.286	6.885	3.813
Information Pruning:	1	0.340	4.572	6.064

Notes. s: participation size; Q: total posts; Π: social welfare; ξ: total privacy harm.

By Proposition 3, a uniform quota (UQ) helps some users obtain a higher surplus and encourages them to join the community. In our example, when $v = 0.051$ or 0.337, imposing a quota can increase user participation. It also enhances social welfare in all the scenarios as suggested in Proposition 5. However, consistent with Proposition 4, it always decreases the total quantity of information posted in the community.

One dilemma highlighted in Section IV and particularly equation (12) is that a non-discriminatory nudge or quota cannot simultaneously increase social welfare, decrease privacy harm, and increase information contribution. We simulate the impacts of imposing a targeted nudge (TN) and targeted

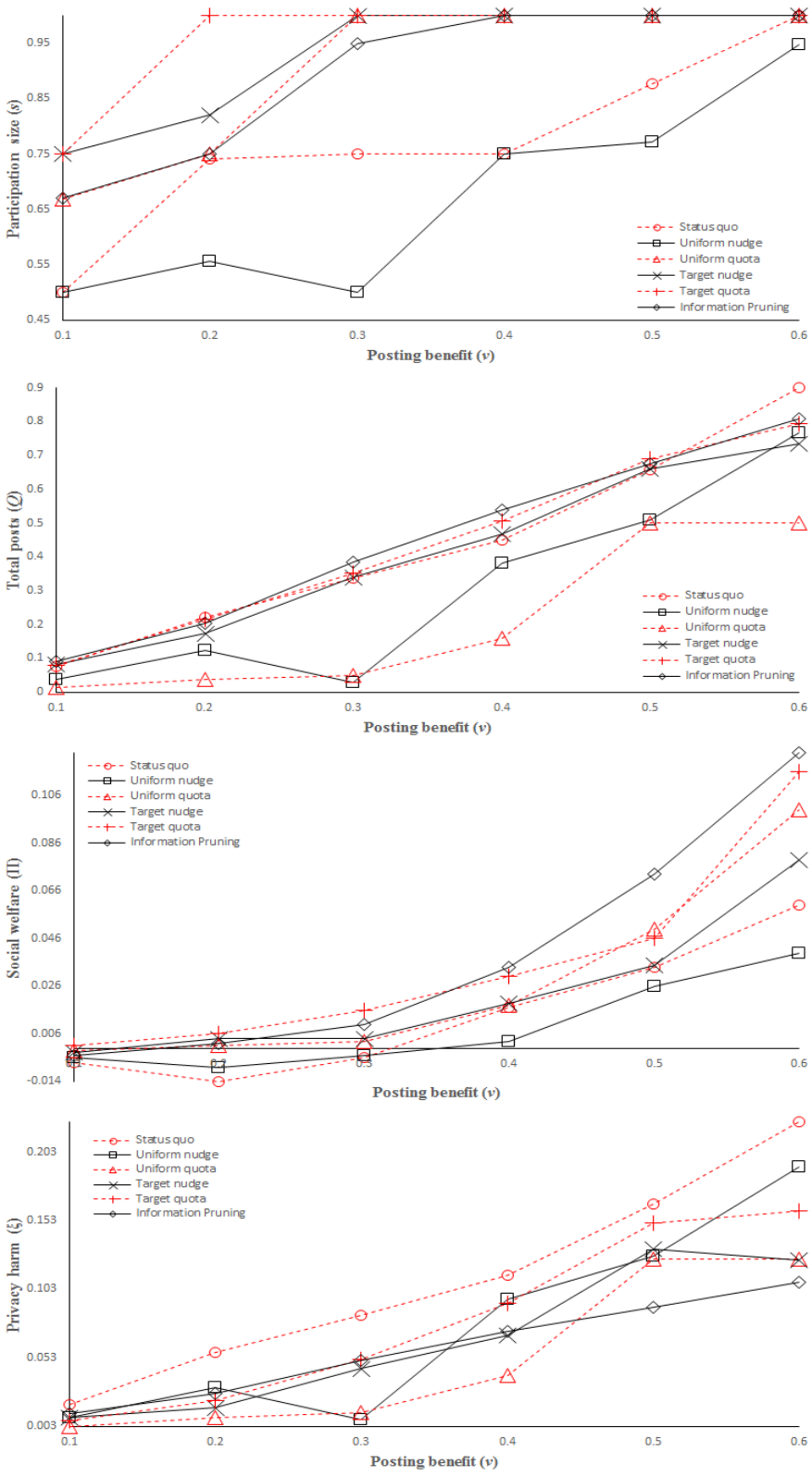

Figure 6: Numerical Example

quota (TQ) as analyzed in Section 4.1. We allow the targeting technologies to be imperfect, viz. non-sensitive information can be misclassified as sensitive ("false positive") and wrongly suppressed, and sensitive information can be misclassified as non-sensitive ("false negative"). We set the probability of such mis-targeting to be 10% (please refer to the Online Supplement for how we derive numerical results for targeted nudge and quota). Furthermore, to illustrate the effect of information pruning (IP) as analyzed in Section 4.2, we multiply v and λ_i by discount factors of 90% and 50%.

As shown in Table 2 and Figure 6, all three targeting and pruning measures, TN, TQ, and IP, can improve social welfare and reduce the privacy harm. They also increase total information contribution in many scenarios (specifically, TN and TQ increase total posts when $v = 0.016$, 0.051, and 0.168, and IP increases total posts when $v = 0.051$, 0.168, and 0.337). This is achieved by attracting more users to join the community. These numerical results are consistent with our analysis in Section IV, that targeted information control (TN or TQ) and information pruning (IP) can help align the interests of the social planner and community owner. They can lead to a win-win situation – increase social welfare, decrease privacy harm, and increase total information contribution.

VI. Extensions

We assess the robustness of the above analysis by relaxing several key assumptions.

(i) Heterogeneity in "friendship". In this extension, we relax the assumption that all users must have the same expected number of "friends". We allow for heterogeneity in users' number of friends and assume that n_i is uniformly distributed between 0 and 1, i.e., $n_i \in U[0,1]$. To ensure tractability, we let $\theta_i = 1 - n_i \in [0,1]$, meaning that users who are less sensitive about privacy have more friends. The justification is that users who are more privacy sensitive tend to minimize exposure of their personal information, and restricting their friendship network is one effective means to reduce such exposure. Previous research has shown that reciprocity is salient in social networking websites (e.g., Cha et al. 2009; Kumar et al. 2010; Weng et al. 2010), suggesting that people who share more information with others tend to have more friends, which is consistent with our assumption. To simplify the analysis and focus on the impact of heterogeneity in "friendship", we assume that all users are non-committed strategic users in this extension, i.e., $\alpha = 1$. All other setups remain the same as in the main model. The objective function for any user i with n_i friends is specified in Equation (2). We characterize the equilibrium in the status quo in the following lemma.

Lemma 4 *In the status quo, (1) when* $0 < v < (1-\epsilon)(1+\frac{1}{\delta})\frac{\psi-e(1+\psi)-\omega}{1+\psi}$*, users with* $\theta < \theta^o$ *participate in the community and users with* $\theta > \theta^o$ *stay out of the community, where* θ^o *is the solution to* $v = \frac{(1-\epsilon)[\psi\theta^o-e(1+\psi)-\omega]\theta^o[2-(1-\delta)\theta^o]}{(1+\psi)[1-(1-\delta)\theta^o]}$*; (2) when* $v \geq (1-\epsilon)(1+\frac{1}{\delta})\frac{\psi-e(1+\psi)-\omega}{1+\psi}$*, all users participate in the community.*

It is easy to verify that θ^o increases in v, meaning that the participation rate is increasing in the posting benefit. This is consistent with our findings from the main model.

We next consider the impact of a nudge. Imposing a nudge reduces a user's benefit from posting each piece of information. Hence, the impact of a nudge is similar to that of reducing the posting benefit, v. As the participation rate is increasing in v, we expect that a nudge would decrease the participation rate. Furthermore, using a numerical example reported in the Online Supplement, we show that a nudge has similar impacts on the community's participation rate, total information contribution, total privacy harm, and aggregate user surplus as in the main model.

The analysis of a quota is less straightforward because a uniform quota is no longer appropriate when users adopt different posting strategies based on their number of friends. We consider a quota of the following form: $\Lambda_i = f \cdot \frac{\gamma_i(1+\psi)v}{c}$, where $f \in [0,1]$ and $\frac{\gamma_i(1+\psi)v}{c}$ is the total amount of information that user i would post in the status quo. In other words, each user faces a different quota contingent on their numbers of friends. We show in the Online Supplement that there exists a quota that (weakly) increases the participation rate and decreases total information contribution and total privacy harm caused by the community, and hence exhibits similar effects as the quota in the main model. We further use numerical analysis (reported in the Online Supplement) to show that a properly-designed quota improves the social welfare under different parameters.

(ii) Nonlinear externality. We consider a scenario when the externalities, e, ω, and θ increase with the number of participating users. Realistically, the impact of disclosing sensitive information may increase with the size of audience. Having more audience increases the chance that the information resonates with interested friends or acquaintances. Let $e(s) = es$, $\omega(s) = \omega s$, $\theta_L(s) \equiv \theta_L s$, and $\theta_H(s) \equiv \theta_H s = s$, where $s \in (0,1)$ is the fraction of participating users. With these changes, the *total* privacy harm that a user suffers becomes a convex function.

We report the detailed equilibrium outcomes in the Online Supplement. The following lemma characterizes users' participation incentives.

Lemma 5 *When the externalities increase with the number of participating users, the non-committed*

users are more likely to participate in the status quo.

Recall non-committed low type users join the community at a lower v than non-committed high type users. When low type users deliberate their participation decisions, they enjoy less externality because the size of participating users is small, i.e., $s < 1$. Hence, low type users will suffer less privacy harm because the community is still small and so will join when v is lower. When v increases, some non-committed high type users will gradually join. For these early high type users, they also suffer less privacy harm and thus have more incentives to join. Here again, all of our earlier results regarding the merits of the regulations continue to apply.

(iii) Unintended Disclosure. In our model, users make posting decisions about sensitive and non-sensitive information, meaning they intentionally divulge others' sensitive information. In the Online Supplement, we analyze a variant in which users make posting decisions about only one set of information containing both sensitive and non-sensitive information. In other words, they disclose others' sensitive information unintentionally. We assume an exogenous fraction (unknown to the user) of the posted information causes privacy harm on others. All results in the main model continue to apply in this new setting.

VII. Discussion and Implications

Using a stylized model, we show that regulation is necessary to control peer disclosure in an online social community. Depending on the benefit from posting information, the community may have too many participants and the participants may post too much information about other people. Although many countries legislate explicit privacy laws to protect consumer privacy, most of these regulations focus on the merchant-consumer relationship involving direct privacy infringement but not third-party privacy harm such as peer disclosure.[12] In privacy disputes resulting from peer disclosure, individuals may pursue a defamation or libel lawsuit against the insulting party. However, such cases are rare because not every sensitive statement on social media can be considered as the basis for a defamation or libel lawsuit. For example, the mere disclosure of the whereabouts of a person may cause the person to lose her job because of dereliction of duties, but such disclosure does not constitute any defamation or libel. Furthermore, suing people for offensive or disturbing messages may have a negative impact

[12] For example, the European Parliament and Council Directive 95/46/EC states that it "shall not apply to the processing of personal data...by a natural person in the course of a purely personal or household activity".

on free speech (The Telegraph 2012). Practically, it is not feasible to stipulate what a person can say about her friends and peers. Defining and sanctioning peer disclosure could be immensely difficult or costly too. Hence, explicit legislation is not likely to be a practicable solution.

We propose two implementable policies: a nudge and quota. A carefully selected nudge and quota can help enhance user welfare, but they work differently. A nudge helps by driving users who are concerned about privacy out of the community and suppressing those who participate from posting information. A quota helps by reducing the amount of information posted by each user and hence reducing the privacy harm and preserving participation incentives. We show that the community owner will never prefer nudging. It mostly does not prefer a quota either except when it wants to grow its user base. We present an important necessary condition, Proposition 8, for any regulation to achieve the triple objectives of enhancing social welfare, reducing privacy harm, and increasing information contribution. Based on this condition, we propose three solutions that selectively target different kinds of information.

One immediate insight from our analysis in Sections 3.1 and 3.2 is that, lacking any regulation, some users should simply not join online social communities with mediocre benefits from information sharing (cf. the magnitude of privacy harm from peer disclosure). These users should be excluded not because they are "harmful" to other people. Instead, it is because they are more vulnerable to the privacy harm from peer disclosure. The practical implication is that if a person is sensitive about privacy, she should not join online social communities with particularly intimate themes, such as those promoting extra-marital affairs or socially improper behaviors such as drug consumption. Similarly, users who are not ready for politically-charged discussion or abusive comments with real personal identities should stay out from communities predominated by users who like to post information about, confront, or abuse others.

Notwithstanding this insight, regulation is necessary to enhance the collective welfare of users in online social communities. Our nuanced consideration of user participation and information contribution helps explain why almost no online social community is eager to nudge users despite the fact that nudging has been repeatedly advocated (Acquisti 2009; Acquisti et al. 2013; Wang et al. 2013). It also highlights the disadvantage of imposing a quota, that it preserves users' incentives to join the community but decreases overall information contribution.

Ideally, we want to limit the privacy harm due to peer disclosure but encourage users to par-

ticipate in the community and contribute more information. Section 4 discusses three solutions that either impose a nudge or quota selectively to sensitive information, or perturb or prune the sensitive information while retaining its informational benefits. Section 5 shows that these solutions can increase social welfare and are aligned with the community owner's interest too. However, they require sophisticated technologies that can identify and target sensitive personal information reasonably accurately. Such technologies may be impracticable or too expensive and hence not all owners of online social communities are willing to develop and deploy them.

Lacking an extrinsic motivation to address the privacy harm, how can the community owner be motivated to adopt these regulatory policies (either impose a uniform nudge or quota at the cost of less information contribution, or invest in relevant technologies to target sensitive information)? A promising direction is to attribute part of the damage from the privacy harm to the community owner so that it has incentives to address the harm suffered by privacy-sensitive users. For example, the regulator can help victims take legal actions and seek compensation from the owner of an online social community when the privacy damage from peer disclosure is excessive. Such legal sanctions against platform owners who do not directly impose the damage is not without precedents. For example, the U.S. government has shut down MegaUpload.com, a cyberlocker helping users to download movies or music shared by other users, because it has "contributed" to the infringement of the copyright of affected intellectual property owners. It is common for plaintiffs in defamation cases to sue also the media for vicariously contributing to the damage caused by other people (a situation highly similar to "peer disclosure" in our setting). If the platform owner can be held liable for the information disclosed by its users about their peers, then it should have a stronger, vested incentive to regulate user behaviors.[13]

Although our analysis is framed on peer "disclosure", our insights extend to other settings where users impose negative externalities on peers. One example is game invitations on online social communities. Increasingly, games designed for mobile devices encourage players to send invitations to friends before granting additional game credits to the players. This promotional tactic has caused many players to invite friends to try the games, which arguably creates annoyance and inconvenience

[13] Because we have not developed a utility function for the platform owner, we cannot explicitly analyze how this allocation of damage would affect the extent of regulation and the equilibrium outcomes. We do not model the platform owner's utility because its decision may not be driven purely by economic considerations. We defer the study of platform owner's decisions and utility to future research.

to peers. Although this practice does not involve disclosure, it intrudes the private space of the peers and so threatens the seclusion aspect of privacy (Stigler 1980). As such, our analysis directly applies. Nudging or imposing a limit on such "invitations" would help enhance the aggregate user welfare. In fact, we contend that platform owners may have a higher incentive to nudge or cap such game invitations because they are less vested in these promotions.

Numerous instances of externality-curbing policies exist in other peer-to-peer (P2P) applications. For example, BitTorrent, a popular P2P file sharing protocol, reduces network traffic congestion by slowing down the download speed for free riders (Hosanagar et al. 2010). Online music streaming services such as Spotify and Pandora have been pressed by music labels to limit free streaming access to theirs users in order to alleviate the negative externalities imposed on paying users and musicians (Gigaom.com 2013; The Independent 2015). Facebook has experimented with charging fees to users who bombard celebrities with unwanted messages in order to limit the negative externalities generated from such harassment (The Independent 2013). Our framework provides a basis for analyzing the optimal choices of policy instruments in these applications.

A. Implementation

Given our conclusion that regulation is necessary and the identification of the properties of a good regulatory policy, this paper has made an important first step towards improving the privacy and welfare for users to participate in online social communities. The next step is how to implement the right nudges or quotas. Lacking an accurate account of privacy externality, we offer the following guidance.

The first step is to measure the size and privacy sensitivity of users. It may not be feasible to directly poll users about their privacy preferences. Users may not respond to such polls and, even if they do, their responses are likely biased because people tend to exaggerate their privacy needs (Harper and Singleton 2001; Hui et al. 2007; Vasalou et al. 2011). One alternative way is to infer user preferences from their activities. For example, Facebook can track the frequency of users untagging their names from photos shared by their friends. Google can track delisting requests related to identity removal. Such data can help construct users' privacy profiles. A potential challenge with this approach is how to account for selection bias as we can only observe the behavior of participating users. This selection bias poses a smaller threat when the size of participating users is large relative to non-participating users, which is likely the case for large communities such as Facebook or Google.

Nonetheless, measuring the privacy preferences of people not taking part in online social communities is a good topic for future research.

To set the nudge or quota levels, the community owner can gauge users' posting benefit, v, using standard marketing techniques such as conjoint analysis (e.g., Hann et al. 2007; Krasnova et al. 2009) or field experiments (e.g., Hui et al. 2007; Beresford et al. 2012). Users' posting cost, c, consisting of the cost to collect and post peers' information, can be calibrated based on the nature of the community and technological sophistication of the users. For example, users in a community of indecent affairs or paparazzis are likely to incur a higher posting cost as the underlying content is more privacy sensitive and difficult to obtain. By contrast, for a community targeting the mass market such as Facebook, users tend to bear a much lower information collection and posting costs.

The implementation of a nudge or quota also requires quantification of personal information because different types of personal information vary in privacy sensitivity. Previous research has attempted to quantify the value and sensitivity of personal data (e.g., Hui et al. 2007; Hann et al. 2007). Similar methodologies can be extended to other personal data such as photos or videos. A nudge or quota can then be applied directly to each piece of "unitized" information.

Nudging can take different forms in practice, such as warning messages or visual cues. As a user gains experience, they may omit and become unresponsive to privacy nudges. To ensure the salience of the privacy nudges, the community owner can include a time delay whenever a nudge is applied, for example, by forcing users to read the warning messages. It can also regularly change how and when the warning messages or cues are displayed, or the content of the messages or cues itself. By doing so, users will less likely skip the privacy cues.

B. Limitations and Future Research

Our analysis has several limitations. First, for ease of tractability, we assume two types of users, which allows us to show the responses of users with differing privacy sensitivity to the regulations. Future research should extend the analysis to more heterogeneous users. It may also model homophily which could affect how users form friendship and engage in peer disclosure.

Second, we consider the community owner's interest in maximizing user participation or posting of information, but we have not developed its objective function. Constructing such a function may help us gain a holistic view of social welfare and the equilibrium behaviors. The challenge lies in how to reasonably capture the differing objectives of online social communities.

Third, this analysis is confined to one online social community and does not consider "multi-homing" (Koh and Fichman 2014). An interesting extension is to allow users to choose between communities and study how their privacy preferences and peer disclosure interact.

Finally, because of complexity with interplays of many factors such as user commitment, privacy concern, information sensitivity, costs, and the modeling of friendship network structure, we cannot derive unambiguous comparative static analysis. Hence, we cannot conclude if, for example, having more committed or privacy-sensitive users will favor a nudge or a quota, or whether regulation is more or less important when the user demographic changes in a particular direction. Developing a more parsimonious model may help overcome this difficulty. However, balancing parsimony and richness of insights is an obvious challenge.

VIII. Concluding Remarks

In 2017, the number of active users on Facebook, WeChat, Instagram, Twitter, and Pinterest were 1.97 billion, 889 million, 600 million, 319 million, and 150 million respectively (Statistica 2017). There are around 3.2 billion Internet users. This mean that Facebook alone has more than 60% of penetration. Evidently, these online platforms present people with novel avenues for social interaction. New research is necessary to uncover the implications of such interaction with unprecedented reach and scale.

This paper analyzes one novel behavior in online social communities, viz. peer disclosure of personal information. The sharing of information among friends is mostly not regulated, but its consequence is starting to surface. For example, there has been cases when people were sacked from work because of friends' posting of their improper behavior in online social communities, and crime syndicates have used data harvested from online social networks to track targeted victims. Studying the implications of peer disclosure and its regulation is an important first step towards shaping a healthy online environment for social interaction. This study serves such a purpose.

We find that regulation is necessary. The choice of regulation depends on how privacy is treated in the jurisdiction. A nudge is helpful if privacy is absolutely preferred, whereas a quota is better if we focus on economic utility and are willing to trade privacy for the pleasure in sharing information. Most importantly, having more users need not be good for the society. Any thoughtful analysis of the benefits of online social communities should consider the pros and cons of user participation and exit, and the related benefits and damages, in a holistic framework.

References

ABC News. 2015. 15-year-old's "rethink" app aims to prevent cyberbullying. URL `http://abcnews.go.com/Lifestyle/15-year-olds-rethink-app-aims-prevent-cyberbullying/story?id=33329748`.

Acquisti, Alessandro. 2009. Nudging privacy: The behavioral economics of personal information. *IEEE Security & Privacy* **7**(6) 82–85.

Acquisti, Alessandro, Laura Brandimarte, Idris Adjerid. 2013. Gone in 15 seconds: The limit of privacy transparency and control. *IEEE Security & Privacy* **11**(4) 72–74.

Acquisti, Alessandro, Ralph Gross. 2009. Predicting social security numbers from public data. *Proceedings of the National Academy of Sciences* **106**(27) 10975–10980.

Almuhimedi, Hazim, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, Yuvraj Agarwal. 2015. Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15, ACM, New York, NY, USA, 787–796.

Anderson, Simon P., Andre de Palma. 2009. Information congestion. *The RAND Journal of Economics* **40**(4) 688–709.

Anderson, Simon P., Joshua Gans. 2011. Platform siphoning: Ad-avoidance and media content. *American Economic Journal: Microeconomics* **3**(4) 1–34.

Asvanund, Atip, Karen Clay, Ramayya Krishnan, Michael D. Smith. 2004. An empirical analysis of network externalities in peer-to-peer music-sharing networks. *Information Systems Research* **15**(2) 155–174.

Backstrom, Lars. 2011. Anatomy of facebook. *Facebook Data Team*. URL `https://www.facebook.com/notes/facebook-data-team/anatomy-of-facebook/10150388519243859`.

Bateman, Patrick J., Peter H. Gray, Brian S. Butler. 2011. Research note – The impact of community commitment on participation in online communities. *Information Systems Research* **22**(4) 841–854.

Baumol, William J., Wallace E. Oates. 1988. *The Theory of Environmental Policy (2nd Edition)*. Cambridge University Press, Cambridge.

Bennear, Lori Snyder, Robert N. Stavins. 2007. Second-best theory and the use of multiple policy instruments. *Environmental and Resource Economics* **37**(1) 111–129.

Beresford, Alastair R., Dorothea Kubler, Soren Preibusch. 2012. Unwillingness to pay for privacy: A field experiment. *Economics Letters* **117**(1) 25–27.

Besmer, Andrew, Heather R. Lipford. 2010. Moving beyond untagging: Photo privacy in a tagged world. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 1563–1572.

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, Magne Mogstad. 2013. Broadband Internet: An information superhighway to sex crime? *Review of Economic Studies* **80**(4) 1237–1266.

Calthrop, Edward, Stef Proost. 1998. Road transport externalities. *Environmental and Resource Economics* **11**(3) 335–348.

Cha, Meeyoung, Alan Mislove, Krishna P. Gummadi. 2009. A measurement-driven analysis of information propagation in the flickr social network. *Proceedings of the 18th international conference on World wide web, ACM* 721–730.

Chan, Jason, Anindya Ghose, Robert Seamans. 2016. The Internet and racial hate crime: Offline spillovers from online access. *MIS Quarterly* **40**(2) 381–403.

Choi, Ben C. F., Zhenhui (Jack) Jiang, Bo Xiao, Sung S. Kim. 2015. Embarrassing exposures in online social networks: An integrated perspective of privacy invasion and relationship bonding. *Information Systems Research* **26**(4) 675–694.

Christiansen, Vidar, Stephen Smith. 2012. Externality-correcting taxes and regulation. *The Scandinavian Journal of Economics* **114**(2) 358–383.

Collinge, Robert A., Wallace E. Oates. 1982. Efficiency in pollution control in the short and long runs: A system of rental emission permits. *The Canadian Journal of Economics* **15**(2) 346–354.

Conitzer, Vincent, Curtis R. Taylor, Liad Wagman. 2012. Hide and seek: Costly consumer privacy in a market with repeat purchases. *Marketing Science* **31**(2) 277–292.

Copes, Parzival. 1986. A critical review of the individual quota as a device in fisheries management. *Land economics* **62**(3) 278–291.

Cropper, Maureen L., Wallace E. Oates. 1992. Environmental economics: A survey. *Journal of Economic Literature* **30**(2) 675–740.

Daughety, Andrew F., Jennifer F. Reinganum. 2010. Public goods, social pressure, and the choice between privacy and publicity. *American Economic Journal: Microeconomics* **2**(2) 191–221.

Davis, Richard. 1999. *The Web of Politics: The Internet's Impact on the American Political System*. Oxford University Press, New York.

DiMicco, Joan Morris, David R. Millen. 2007. Identity management: Multiple presentations of self in Facebook. *Proceedings of International ACM Conference on Supporting Group Work* 383–386.

Ding, Amy Wenxuan, Shibo Li, Patrali Chatterjee. 2015. Learning user real-time intent for optimal dynamic web page transformation. *Information Systems Research* **26**(2) 339–359.

Dwyer, Catherine, Starr Roxanne Hiltz, Katia Passerini. 2007. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *Proceedings of the American Conference on Information Systems* 339.

Fullerton, Don, Gilbert E. Metcalf. 2001. Environmental controls, scarcity rents, and pre-existing distortions. *Journal of Public Economicw* **80**(2) 249–267.

Galletta, Dennis F., Raymond M. Henry, Scott McCoy, Peter Polak. 2006. When the wait isn't so bad: The interacting effects of website delay, familiarity, and breadth. *Information Systems Research* **17**(1) 20–37.

Gigaom.com. 2013. Pandora caps monthly free tunes on mobiles to 40 hours. URL `https://gigaom.com/2013/02/28/pandora-caps-monthly-free-tunes-on-mobiles-to-40-hours/`.

Gross, Ralph, Alessandro Acquisti. 2005. Information revelation and privacy in online social networks: The Facebook case. *Proceedings of the 2005 ACM workshop on Privacy in the electronic society* 71–80.

Hann, Il-Horn, Kai-Lung Hui, Sang-Yong Tom Lee, Ivan P.L. Png. 2007. Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems* **24**(2) 13–42.

Hann, Il-Horn, Kai-Lung Hui, Sang-Yong, Ivan P. L. Png. 2008. Consumer privacy and marketing avoidance: A static model. *Management Science* **54**(6) 1094–1103.

Harper, Jim, Solveig Singleton. 2001. With a grain of salt: What consumer privacy surveys don't tell us. *Competitive Enterprise Institute* 1–18.

Henne, Benjamin, Matthew Smith. 2013. Awareness about photos on the web and how privacy-privacy-tradeoffs could help. *International Conference on Financial Cryptography and Data Security*. Springer, 131–148.

Hermalin, Benjamin E., Michael L. Katz. 2006. Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics* **4**(2) 209–239.

Hosanagar, Kartik, Peng Han, Yong Tan. 2010. Diffusion models for peer-to-peer (P2P) media distribution: On the impact of decentralized, constrained supply. *Information Systems Research* **21**(2) 271–287.

Hui, Kai-Lung, Ivan P. L. Png. 2006. The economics of privacy. *Handbook of Economics and Information Systems* 471–497.

Hui, Kai-Lung, Hock Hai Teo, Sang-Yong Tom Lee. 2007. The value of privacy assurance: An exploratory field experiment. *MIS Quarterly* **31**(1) 19–33.

Johnson, Justin P. 2013. Targeted advertising and advertising avoidance. *The RAND Journal of Economics* **44**(1) 128–144.

Keith, Susan, Michelle E. Martin. 2005. Cyber-bullying: Creating a culture of respect in a cyber world. *Reclaiming children and youth* **13**(4) 224–228.

Koh, Tat-Koon, Mark Fichman. 2014. Multi-homing users' preferences for two-sided exchange networks. *MIS Quarterly* **38**(4) 977–996.

Kowalski, Robin M., Susan P. Limber. 2007. Electronic bullying among middle school students. *Journal of Adolescent Health* **41**(6) S22–S30.

Krasnova, Hanna, Thomas Hildebrand, Oliver Guenther. 2009. Investigating the value of privacy in online social networks: Conjoint analysis. *International Conference on Information Systems*.

Kumar, Ravi, Jasmine Novak, Andrew Tomkins. 2010. Structure and evolution of online social networks. *Link mining: models, algorithms, and applications. Springer New York* 337–357.

Li, Xiao-Bai, Sumit Sarkar. 2006. Privacy protection in data mining: A perturbation approach for categorical data. *Information Systems Research* **17**(3) 254–270.

Liebowitz, S. J., Stephen E. Margolis. 1994. Network externality: An uncommon tragedy. *The Journal of Economic Perspectives* **8**(2) 133–150.

Menon, Syam, Sumit Sarkar. 2007. Minimizing information loss and preserving privacy. *Management Science* **53**(1) 101–116.

New York Daily News. 2013. Florida teacher fired after she rented party penthouse for students that included alcohol, condoms. URL `http://www.nydailynews.com/news/national/fla-teacher-fired-allegedly-giving-students-alcohol-condoms-article-1.1488471`.

New York Times. 2003. Fame is no laughing matter for the 'star wars kid'. URL `http://www.nytimes.com/2003/05/19/business/compressed-data-fame-is-no-laughing-matter-for-the-star-wars-kid.html`.

New York Times. 2016. Don't post about me on social media, children say. URL `http://well.blogs.nytimes.com/2016/03/08/dont-post-about-me-on-social-media-children-say`.

Pigou, Arthur C. 1920. The economics of welfare. *McMillan&Co., London* .

Pizer, William A. 2002. Combining price and quantity controls to mitigate global climate change. *Journal of Public Economics* **85**(3) 409–434.

Posner, Richard A. 1978. An economic theory of privacy. *Regulation* **9**(3) 19–26.

Posner, Richard A. 1979. Privacy, secrecy, and reputation. *Buffalo Law Review* **28** 1–55.

Posner, Richard A. 1981. The economics of privacy. *American Economic Review* **71**(2) 405–409.

Rajamma, Rajasree K., Audhesh K. Paswan, Muhammad M. Hossain. 2009. Why do shoppers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience. *Journal of Product & Brand Management* **18**(3) 188–197.

Roberts, Marc J., Michael Spence. 1976. Effluent charges and licenses under uncertainty. *Journal of Public Economics* **5**(3-4) 193–208.

Sandholm, William H. 2002. Evolutionary imlementation and congestion pricing. *The Review of Economic Studies* **69** 667–689.

Sandholm, William H. 2005. Negative externalities and evolutionary implementation. *The Review of Economic Studies* **72**(3) 885–915.

Schulze, William, Ralph C. d' Arge. 1974. The coase proposition, information constraints, and long-run equilibrium. *The American Economic Review* **64**(4) 763–772.

Statistica. 2017. Most famous social network sites worldwide as of April 2017, ranked by number of active users (in millions). URL https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Stavins, Robert N. 2011. The problem of the commons: Still unsettled after 100 years. *The American Economic Review* **101**(1) 81–108.

Stigler, George J. 1980. An introduction to privacy in economics and politics. *Journal of Legal Studies* **9**(4) 623–644.

The Independent. 2013. Facebook now charges you for messages sent to celebrities and people you aren't friends with. URL http://www.independent.co.uk/news/uk/home-news/facebook-now-charges-you-for-messages-sent-to-celebrities-and-people-you-arent-friends-with-8563299.html.

The Independent. 2015. Spotify reportedly under pressure from music labels to limit free streaming. URL http://www.independent.co.uk/arts-entertainment/music/news/spotify-reportedly-under-pressure-from-music-labels-to-limit-free-streaming-10126106.html.

The Telegraph. 2012. Drunk Twitter users unlikely to face criminal prosecution. URL http://www.telegraph.co.uk/technology/twitter/9754007/Drunk-Twitter-users-unlikely-to-face-criminal-prosecution.html.

Tufekci, Zeynep. 2008. Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science Technology and Society* **28**(1) 20–36.

Van Zandt, Timothy. 2004. Information overload in a network of targeted communication. *The RAND Journal of Economics* **35**(3) 542–560.

Vasalou, Asimina, Alastair J. Gill, Fadhila Mazanderani, Chrysanthi Papoutsi, Adam Joinson. 2011. Privacy dictionary: A new resource for the automated content analysis of privacy. *Journal of The American Soceity for Information Science and Technology* **62**(11) 2095–2105.

Vickrey, William S. 1963. Pricing in urban and suburban transport. *The American Economic Review* **53**(2) 452–465.

Wang, Yang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, Lorrie Faith Cranor. 2013. Privacy nudges for social media: An exploratory facebook study. *Proceedings of the 22nd International World Wide Web Conference Committee (IW3C2)* 763–770.

Weitzman, Martin L. 1974. Prices vs. quantities. *The Review of Economic Studies* **41**(4) 477–491.

Weng, Jianshu, Ee-Peng Lim, Jing Jiang, Qi He. 2010. Twitterrank: Finding topic-sensitive influential twitterers. *Proceedings of the third ACM international conference on Web search and data (WSDM), ACM* 261–270.

Appendix. A

This appendix presents the optimal quota that maximizes the aggregate user surplus. We use the superscript $*$ to denote the associated outcome. ι is an infinitesimally small positive number.

(i) When $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$:

(i.1) If $0 < v \leq (1-\epsilon)(1-\alpha)\lambda_L$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 0$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha$.

(i.2) If $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$, $\Lambda^* = \iota$, leading to $s_L^{q*} = \frac{(v-\iota)/((1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$, $s_H^{q*} = 0$, and $s^{q*} = \frac{v-\iota}{(1-\epsilon)\lambda_L}$.

(i.3) If $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha\beta$.

(i.4) If $\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H < v \leq (1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)[v-\bar{\lambda}+(1-\epsilon)\alpha\beta\lambda_H]}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha\beta$.

(i.5) If $(1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda} < v \leq 2(1-\epsilon)(1-\alpha\beta)\lambda_H$, $\Lambda^* = \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha\beta$.

(i.6) If $2(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq 2(1-\epsilon)\lambda_H - \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)v}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = \frac{v/(2(1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$, and $s^{q*} = \frac{v}{2(1-\epsilon)\lambda_H}$.

(i.7) If $v > \max\{2(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)\lambda_H - \bar{\lambda}\}$, $\Lambda^* = \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 1$, and $s^{q*} = 1$.

(ii) When $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$:

(ii.1) If $0 < v \leq (1-\epsilon)(1-\alpha)\lambda_L$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 0$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha$.

(ii.2) If $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$, $\Lambda^* = \iota$, leading to $s_L^{q*} = \frac{(v-\iota)/((1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$, $s_H^{q*} = 0$, and $s^{q*} = \frac{v-\iota}{(1-\epsilon)\lambda_L}$.

(ii.3) If $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 0$, and $s^{q*} = 1-\alpha\beta$.

(ii.4) If $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 1$, $s_H^{q*} = \frac{(v-\iota)/((1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$, and $s^{q*} = \frac{v-\iota}{(1-\epsilon)\lambda_H}$.

(ii.5) If $(1-\epsilon)\lambda_H < v \leq \bar{\lambda}$, $\Lambda^* = \iota$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 1$, and $s^{q*} = 1$.

(ii.6) If $v > \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}$, leading to $s_L^{q*} = 1$, $s_H^{q*} = 1$, and $s^{q*} = 1$.

An Economic Analysis of Peer-Disclosure in Online Social Communities

Online Supplement

I. Omitted Proofs in the Main Model

Proof of Lemma 1. Note that $u_{i|s}^{sq,in} > u_{i|s}^{sq,out}$ is equivalent to $u_{i|s}^{sq,in-out} \equiv u_{i|s}^{sq,in} - u_{i|s}^{sq,out} > 0$. We can easily derive $u_{i|s}^{sq,in-out} = \frac{\gamma(1+\psi)v}{2c}[v - 2(1-\epsilon)s\lambda_i]$, $i \in \{L, H\}$. Note that $u_{i|s}^{sq,in-out}$ decreases in s and λ_i. Then any potential equilibrium can be characterized by one of the following 5 cases: (1) $u_{H|s}^{sq,in-out} < u_{L|s}^{sq,in-out} < 0$, (2) $u_{H|s}^{sq,in-out} < u_{L|s}^{sq,in-out} = 0$, (3) $u_{H|s}^{sq,in-out} < 0 < u_{L|s}^{sq,in-out}$, (4) $u_{H|s}^{sq,in-out} = 0 < u_{L|s}^{sq,in-out}$, or (5) $0 < u_{H|s}^{sq,in-out} < u_{L|s}^{sq,in-out}$. In the following, we derive the conditions for each equilibrium. Q denotes the total information, Π denotes the social welfare, ξ denotes the total *objective* privacy damage and ξ' denotes the total *perceived* privacy damage. The superscript sq indicates the status quo.

(1) If $u_{H|s}^{sq,in-out} < u_{L|s}^{sq,in-out} < 0$, then no non-committed user will prefer to join, implying the equilibrium s is $s^{sq} = 1 - \alpha$ (i.e., only committed users join). Substituting $s = 1 - \alpha$ into $u_{H|s}^{sq,in-out} < u_{L|s}^{sq,in-out} < 0$, we get $v < 2(1-\epsilon)(1-\alpha)\lambda_L$, which is the sufficient and necessary condition for the existence of this equilibrium. Accordingly, $Q^{sq} = \frac{(1-\alpha)\gamma(1+\psi)v}{c}$, $\Pi^{sq} = \frac{(1-\alpha)\gamma(1+\psi)v}{2c}[v - 2(1-\alpha+\epsilon\alpha)\bar{\lambda}]$, $\xi^{sq} = \frac{(1-\alpha)\gamma\psi\bar{\theta}v}{c}$, and $\xi'^{sq} = \frac{(1-\alpha)\gamma\psi v}{c}[(1-\alpha+\epsilon\alpha)\bar{\theta}]$.

(2) If $u_{H|s}^{sq,in-out} < u_{L|s}^{sq,in-out} = 0$, then low-type non-committed users are indifferent between joining and not joining while high-type non-committed users prefer not to join, implying in equilibrium $s^{sq} \in [1-\alpha, 1-\alpha\beta]$. Meanwhile, the condition $u_{L|s}^{sq,in-out} = 0$ gives $s^{sq} = \frac{v}{2(1-\epsilon)\lambda_L}$. So $\frac{v}{2(1-\epsilon)\lambda_L} \in [1-\alpha, 1-\alpha\beta]$ leads to $2(1-\epsilon)(1-\alpha)\lambda_L \le v \le 2(1-\epsilon)(1-\alpha\beta)\lambda_L$, which is the sufficient and necessary condition for the existence of this equilibrium. In this equilibrium, a fraction, s_L^{sq}, of low-type non-committed users choose to join while the rest of low-type non-committed users choose not to join. And $s_L^{sq} = \frac{s^{sq}-(1-\alpha)}{\alpha(1-\beta)} = \frac{v/(2(1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$. Accordingly, $Q^{sq} = \frac{\gamma(1+\psi)v^2}{2c(1-\epsilon)\lambda_L}$, $\Pi^{sq} = \frac{r(1+\psi)v^2}{2c}\left[(1-\alpha) - \frac{(1-\alpha+\epsilon\alpha)\bar{\lambda}}{(1-\epsilon)\lambda_L}\right]$, $\xi^{sq} = \frac{\gamma\psi\bar{\theta}v^2}{2c(1-\epsilon)\lambda_L}$, and $\xi'^{sq} = \frac{\gamma\psi v^2}{2c(1-\epsilon)\lambda_L}[(1-\alpha+\epsilon\alpha)\bar{\theta} + \alpha(1-\beta)(1-\epsilon)s_L^{sq}\theta_L]$.

(3) if $u_{H|s}^{sq,in-out} < 0 < u_{L|s}^{sq,in-out}$, then low-type non-committed users prefer to join while high-type non-committed users prefer not to join, implying in equilibrium $s^{sq} = 1 - \alpha\beta$. Substituting $s = 1 - \alpha\beta$ into $u_{H|s}^{sq,in-out} < 0 < u_{L|s}^{sq,in-out}$, we obtain $2(1-\epsilon)(1-\alpha\beta)\lambda_L < v < 2(1-\epsilon)(1-\alpha\beta)\lambda_H$, which is the sufficient and necessary condition for the existence of this equilibrium. Accordingly, $Q^{sq} = \frac{(1-\alpha\beta)\gamma(1+\psi)v}{c}$, $\Pi^{sq} = \frac{(1-\alpha\beta)\gamma(1+\psi)v}{2c}\{v - 2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H]\}$, $\xi^{sq} = \frac{(1-\alpha\beta)\gamma\psi\bar{\theta}v}{c}$, and $\xi'^{sq} = \frac{(1-\alpha\beta)\gamma\psi v}{c}[\bar{\theta} - (1-\epsilon)\alpha\beta\theta_H]$.

(4) if $u_{H|s}^{sq,in-out} = 0 < u_{L|s}^{sq,in-out}$, then low-type users prefer to join while high-type users are indifferent between joining and not joining, implying in equilibrium $s^{sq} \in [1-\alpha\beta, 1]$. Meanwhile, the condition $u_{H|s}^{sq,in-out} = 0$ gives $s^{sq} = \frac{v}{2(1-\epsilon)\lambda_H}$. So $\frac{v}{2(1-\epsilon)\lambda_H} \in [1-\alpha\beta, 1]$ leads to $2(1-\epsilon)(1-\alpha\beta)\lambda_H \le v \le 2(1-\epsilon)\lambda_H$, which is the sufficient and necessary condition for the existence of this equilibrium. In this equilibrium, a fraction, s_H^{sq}, of high-type non-committed users choose to join while the rest of high-type non-committed users choose not to join. And $s_H^{sq} = \frac{s^{sq}-(1-\alpha\beta)}{\alpha\beta} = \frac{v/(2(1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$. Accordingly, $Q^{sq} = \frac{\gamma(1+\psi)v^2}{2c(1-\epsilon)\lambda_H}$, $\Pi^{sq} = \frac{r(1+\psi)v^2}{2c}\left[1 - \frac{\bar{\lambda}}{(1-\epsilon)\lambda_H}\right]$, $\xi^{sq} = \frac{\gamma\psi\bar{\theta}v^2}{2c(1-\epsilon)\lambda_H}$, and $\xi'^{sq} = \frac{\gamma\psi v^2}{2c(1-\epsilon)\lambda_H}[\bar{\theta} - (1-\epsilon)(1 - s_H^{sq})\alpha\beta\theta_H]$.

(5) If $0 < u_{H|s}^{sq,in-out} < u_{L|s}^{sq,in-out}$, then all non-committed users prefer to join, implying in equilibrium $s^{sq} = 1$. Substituting $s = 1$ into $0 < u_{H|s}^{sq,in-out} < u_{L|s}^{sq,in-out}$, we obtain $v > 2(1-\alpha\beta)\lambda_H$, which is the sufficient and necessary condition for the existence of this equilibrium. Accordingly, $Q^{sq} = \frac{\gamma(1+\psi)v}{c}$,

$\Pi^{sq} = \frac{\gamma(1+\psi)v}{2c}(v - 2\bar{\lambda})$, $\xi^{sq} = \frac{\gamma\psi\bar{\theta}v}{c}$, and $\xi'^{sq} = \frac{\gamma\psi\bar{\theta}v}{c}$ (the same as ξ^{sq} since all users participate).

It is easy to check that $Q^{sq}(v)$, $\xi^{sq}(v)$, $\xi'^{sq}(v)$ are both continuous and increasing in v (to see $\xi'^{sq}(v)$ increases in v, one should note $s_L^{sq})(v)$ and $s_H^{sq}(v)$ both increase in v).

Proof of Lemma 2. First, it is easy to check that $\Pi^{sq}(v)$ is continuous in v.

(1) If $0 < v < 2(1-\epsilon)(1-\alpha)\lambda_L$: $\Pi^{sq} = \frac{(1-\alpha)\gamma(1+\psi)v}{2c}[v-2(1-\alpha+\epsilon\alpha)\bar{\lambda}] < 0$ because $v < 2(1-\epsilon)(1-\alpha)\lambda_L < 2(1-\alpha+\epsilon\alpha)\bar{\lambda}$. And Π^{sq} is decreasing in v when $0 < v < \min\left\{2(1-\epsilon)(1-\alpha)\lambda_L, (1-\alpha+\epsilon\alpha)\bar{\lambda}\right\}$ and increasing in v when $(1-\alpha+\epsilon\alpha)\bar{\lambda} < v < 2(1-\epsilon)(1-\alpha)\lambda_L$.

(2) If $2(1-\epsilon)(1-\alpha)\lambda_L \leq v \leq 2(1-\epsilon)(1-\alpha\beta)\lambda_L$: $\Pi^{sq} = \frac{r(1+\psi)v^2}{2c}\left[(1-\alpha) - \frac{(1-\alpha+\epsilon\alpha)\bar{\lambda}}{(1-\epsilon)\lambda_L}\right] < 0$ because $(1-\alpha) - \frac{(1-\alpha+\epsilon\alpha)\bar{\lambda}}{(1-\epsilon)\lambda_L} < 0$. And obviously Π^{sq} is decreasing in v.

(3) If $2(1-\epsilon)(1-\alpha\beta)\lambda_L < v < 2(1-\epsilon)(1-\alpha\beta)\lambda_H$: $\Pi^{sq} = \frac{(1-\alpha\beta)\gamma(1+\psi)v}{2c}\{v-2[\bar{\lambda}-(1-\epsilon)\alpha\beta\lambda_H]\}$. First note that $2[\bar{\lambda}-(1-\epsilon)\alpha\beta\lambda_H] > 2(1-\epsilon)(1-\alpha\beta)\lambda_L$. If $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, then $2[\bar{\lambda}-(1-\epsilon)\alpha\beta\lambda_H] \leq 2(1-\epsilon)(1-\alpha\beta)\lambda_H$, so $\Pi^{sq} < 0$ iff $2(1-\epsilon)(1-\alpha\beta)\lambda_L < v < 2[\bar{\lambda}-(1-\epsilon)\alpha\beta\lambda_H]$; If $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, then $2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H] > 2(1-\epsilon)(1-\alpha\beta)\lambda_H$, so $\Pi^{sq} < 0$ iff $2(1-\epsilon)(1-\alpha\beta)\lambda_L < v < 2(1-\epsilon)(1-\alpha\beta)\lambda_H$. And Π^{sq} is decreasing in v iff $2(1-\epsilon)(1-\alpha\beta)\lambda_L < v < \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H$ and increasing in v iff $\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H < v < 2(1-\epsilon)(1-\alpha\beta)\lambda_H$.

(4) If $2(1-\epsilon)(1-\alpha\beta)\lambda_H \leq v \leq 2(1-\epsilon)\lambda_H$: $\Pi^{sq} = \frac{r(1+\psi)v^2}{2c}\left[1 - \frac{\bar{\lambda}}{(1-\epsilon)\lambda_H}\right]$. If $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, then $1 - \frac{\bar{\lambda}}{(1-\epsilon)\lambda_H} > 0$, so $\Pi^{sq} > 0$ and is increasing in v; If $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, then $1 - \frac{\bar{\lambda}}{(1-\epsilon)\lambda_H} < 0$, so $\Pi^{sq} < 0$ and is decreasing in v.

(5) If $v > 2(1-\epsilon)\lambda_H$: $\Pi^{sq} = \frac{\gamma(1+\psi)v}{2c}(v - 2\bar{\lambda})$. If $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, then $2\bar{\lambda} \leq 2(1-\epsilon)\lambda_H$, so $\Pi^{sq} > 0$; If $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, then $2\bar{\lambda} > 2(1-\epsilon)\lambda_H$, so $\Pi^{sq} < 0$ iff $2(1-\epsilon)\lambda_H < v < 2\bar{\lambda}$. And Π^{sq} is decreasing in v iff $2(1-\epsilon)\lambda_H < v < \bar{\lambda}$ and increasing in v iff $v > \max\{\bar{\lambda}, 2(1-\epsilon)\lambda_H\}$.

Lemma 2 summarizes (1) through (5) above. If $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H}) \Leftrightarrow \beta \leq \frac{(1-\epsilon)\lambda_H - \lambda_L}{\lambda_H - \lambda_L}$, condition (1), (2), and (3) together yield $\Pi^{sq} < 0$ iff $0 < v < 2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H]$; if $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H}) \Leftrightarrow \beta > \frac{(1-\epsilon)\lambda_H - \lambda_L}{\lambda_H - \lambda_L}$, condition (1), (2), (3), (4), and (5) together yield $\Pi^{sq} < 0$ iff $0 < v < 2\bar{\lambda}$. The two conditions are equivalent to the conditions (i) and (ii) in Lemma 2.

It is also easy to check that Π^{sq} decreases in v only when $\Pi^{sq} < 0$.

Proof of Lemma 3. To obtain the equilibrium outcomes here, we follow the exact same approach in Proof of Lemma 1. Nothing that the impact of a nudge is basically a linear reduction in v, we can easily obtain Lemma 3.

Proof of Proposition 1. Note that the impact of a nudge relative to the status quo is decreasing v. By Lemma 1, the participation size, s, weakly increases with v in the status quo, so decreasing v would (weakly) reduce the participation rate in the community. Similarly, by Proof of Lemma 1, the total amount of information and the total privacy damage both increase with v in the status quo, so decreasing v would reduce the total information and total privacy damage.

Proof of Proposition 2. By the Proof of Lemma 2, the social welfare in the status quo may decrease with v when it is negative, but the social welfare always increases with v when it is positive. So when

the social welfare in the status quo is positive, imposing a positive nudge (i.e., decreasing v) would decrease the social welfare; when the social welfare in the status quo is negative, imposing a positive nudge (i.e., decreasing v) may increase the social welfare, but the maximum welfare is obviously zero, which is achieved by setting a sufficiently costly nudge $\tau = v$.

Proof of Proposition 3. When user i makes decisions about the four types of postings, we can define a Lagrange function as the following (ignoring the externality terms since they are not user i's decisions)

$$\begin{aligned}\Phi(x_{ij}, y_{ij}, x_{ik}, y_{ik}, \kappa) =& n\left(vx_{ij} - \frac{cx_{ij}^2}{2} + vy_{ij} - \frac{cy_{ij}^2}{2\psi}\right) + (1-n)\left(vx_{ik} - \frac{cx_{ik}^2}{2\delta} + vy_{ik} - \frac{cy_{ik}^2}{2\delta\psi}\right) \\ &+ \kappa[n(x_{ij} + y_{ij}) + (1-n)(x_{ik} + y_{ik}) - \Lambda],\end{aligned}$$

where κ is the Lagrange multiplier. The FOCs are:

$$\begin{aligned}&\frac{\partial}{\partial x_{ij}}\Phi = n(v - cx_{ij}) + n\kappa = 0, \\ &\frac{\partial}{\partial y_{ij}}\Phi = n(v - \frac{c}{\psi}y_{ij}) + n\kappa = 0, \\ &\frac{\partial}{\partial x_{ik}}\Phi = (1-n)(v - \frac{c}{\delta}x_{ik}) + (1-n)\kappa = 0, \\ &\frac{\partial}{\partial y_{ik}}\Phi = (1-n)(v - \frac{c}{\delta\psi}y_{ik}) + (1-n)\kappa = 0, \\ &\frac{\partial}{\partial \kappa}\Phi = n(x_{ij} + y_{ij}) + (1-n)(x_{ik} + y_{ik}) - \Lambda = 0.\end{aligned}$$

Then we can easily derive: $x_{ij}^q = \frac{\Lambda}{(1+\psi)\gamma}$, $y_{ij}^q = \frac{\psi\Lambda}{(1+\psi)\gamma}$, $x_{ik}^q = \frac{\delta\Lambda}{(1+\psi)\gamma}$ and $y_{ik}^q = \frac{\delta\psi\Lambda}{(1+\psi)\gamma}$, where $\gamma = n + (1-n)\delta$. Since the utility function is concave, these quantities are obviously globally optimal ones.

Substituting the above equilibrium quantities back into equation (2), we can obtain user i's utility from participating in the community conditional on the participation size, s:

$$u_{i|s}^{q,in} = v\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)} - \Lambda s\lambda_i. \tag{A.1}$$

And user i's utility from staying out conditional on s:

$$u_{i|s}^{q,out} = -\epsilon\Lambda s\lambda_i. \tag{A.2}$$

So user i's *net* utility from participating in the community is obtained by subtracting (A.2) from (A.1),

$$u_{i|s}^{q,in-out} = v\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)} - (1-\epsilon)\Lambda s\lambda_i = \Lambda\left[v - \frac{c\Lambda}{2\gamma(1+\psi)} - (1-\epsilon)s\lambda_i\right], i \in \{L, H\}. \tag{A.3}$$

Note that $u_{i|s}^{q,in-out}$ decreases in s and λ_i. Then any potential equilibrium can be characterized by one of the following 5 cases: (1) $u_{H|s}^{q,in-out} < u_{L|s}^{q,in-out} < 0$, (2) $u_{H|s}^{q,in-out} < u_{L|s}^{q,in-out} = 0$, (3) $u_{H|s}^{q,in-out} < 0 < u_{L|s}^{q,in-out}$, (4) $u_{H|s}^{q,in-out} = 0 < u_{L|s}^{q,in-out}$, or (5) $0 < u_{H|s}^{q,in-out} < u_{L|s}^{q,in-out}$. It is important to note that

the participation rate increases with the numbering of the five cases, i.e., (5) > (4) > (3) > (2) > (1) in participation rate. We next derive the conditions (more importantly, the appropriate ranges for Λ) under which the *best* participation rate is (1), (2), (3), (4), or (5) respectively.

(1) If $u_{H|s}^{q,in-out} < u_{L|s}^{q,in-out} < 0$, then no non-committed user will prefer to join, implying the equilibrium s is $s^q = 1-\alpha$ (i.e., only committed users join). Substituting $s = 1-\alpha$ into $u_{H|s}^{q,in-out} < u_{L|s}^{q,in-out} < 0$, we get $v < \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha)\lambda_L$, which is the sufficient and necessary condition for the existence of this equilibrium. Note that Λ is a decision variable and $\Lambda > 0$. To ensure this equilibrium is the best one, $v < \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha)\lambda_L$ must hold for any Λ, then we must have $v \leq (1-\epsilon)(1-\alpha)\lambda_L$. Therefore, when $v \leq (1-\epsilon)(1-\alpha)\lambda_L$, no Λ can attract non-committed users to join, so $s^q = 1-\alpha$. By Lemma 1, in this range of v, $s^{sq} = 1-\alpha$, so no Λ can improve the participation rate relative to the status quo.

(2) If $u_{H|s}^{q,in-out} < u_{L|s}^{q,in-out} = 0$, then low-type non-committed users are indifferent between joining and not joining while high-type non-committed users prefer not to join, implying in equilibrium $s^q \in [1-\alpha, 1-\alpha\beta]$. Meanwhile, the condition $u_{L|s}^{q,in-out} = 0$ gives $s^q = \frac{1}{(1-\epsilon)\lambda_L}\left[v - \frac{c\Lambda}{2\gamma(1+\psi)}\right]$. Note the range of s^q, we obtain $\frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha)\lambda_L < v < \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha\beta)\lambda_L$, which is the sufficient and necessary condition for the existence of this equilibrium. To ensure this equilibrium is the best one, $v < \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha\beta)\lambda_L$ must hold for any Λ, thus we must have $v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$. Meanwhile, $\frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha)\lambda_L < v$ implies $v > (1-\epsilon)(1-\alpha)\lambda_L$ and $\Lambda < \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha)\lambda_L]}{c}$. Moreover, since $s^q = \frac{1}{(1-\epsilon)\lambda_L}\left[v - \frac{c\Lambda}{2\gamma(1+\psi)}\right]$, the smaller Λ, the larger s^q. So to maximize s^q, Λ should be set infinitesimally close to zero, i.e., $\Lambda^\star = \iota$, so the maximum participation size $s^\star = \frac{v-\iota}{(1-\epsilon)\lambda_L}$. Therefore, when $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$, $\Lambda^\star = \iota$ can maximize the participation rate, and the fraction of non-committed low type users who participate is $f_L^\star = \frac{(v-\iota)/((1-\epsilon)\lambda_L)-(1-\alpha)}{\alpha(1-\beta)}$. By Lemma 1, in this range of v, $s^{sq} = \max\{1-\alpha, \frac{v}{2(1-\epsilon)\lambda_L}\} < s^\star$, so $\Lambda^\star$ improves the participation rate relative to the status quo.

(3) if $u_{H|s}^{q,in-out} < 0 < u_{L|s}^{q,in-out}$, then low-type non-committed users prefer to join while high-type non-committed users prefer not to join, implying in equilibrium $s^q = 1-\alpha\beta$. Substituting $s = 1-\alpha\beta$ into $u_{H|s}^{q,in-out} < 0 < u_{L|s}^{q,in-out}$, we obtain $\frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha\beta)\lambda_L < v < \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha\beta)\lambda_H$, which is the sufficient and necessary condition for the existence of this equilibrium. To ensure this equilibrium is the best one, $v < \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha\beta)\lambda_H$ must hold for any Λ, then we must have $v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H$. Meanwhile, $\frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha\beta)\lambda_L < v$ implies $v > (1-\epsilon)(1-\alpha\beta)\lambda_L$ and $\Lambda < \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_L]}{c}$. Therefore, when $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H$, $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_L]}{c}\right)$ can maximize the participation rate, and only all non-committed low type users participate ($s^\star = 1-\alpha\beta$). By Lemma 1, in this range of v, $s^{sq} = \min\{\frac{v}{2(1-\epsilon)\lambda_L}, 1-\alpha\beta\}$. So, when $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq \min\{(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)(1-\alpha\beta)\lambda_L\}$, $s^{sq} = \frac{v}{2(1-\epsilon)\lambda_L} < s^\star$, $\Lambda^\star$ improves the participation rate relative to the status quo; when $2(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H$, $s^{sq} = 1-\alpha\beta = s^\star$, $\Lambda^\star$ does not change the participation rate relative to the status quo.

(4) If $u_{H|s}^{q,in-out} = 0 < u_{L|s}^{q,in-out}$, then low-type users prefer to join while high-type users are indifferent between joining and not joining, implying in equilibrium $s^{sq} \in [1-\alpha\beta, 1]$. Meanwhile, the condition $u_{H|s}^{q,in-out} = 0$ gives $s^q = \frac{1}{(1-\epsilon)\lambda_H}\left[v - \frac{c\Lambda}{2\gamma(1+\psi)}\right]$. Note the range of s^q, we obtain $\frac{c\Lambda}{2\gamma(1+\psi)} +$

$(1-\epsilon)(1-\alpha\beta)\lambda_H < v < \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)\lambda_H$, which is the sufficient and necessary condition for the existence of this equilibrium. To ensure this equilibrium is the best one, $v < \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)\lambda_H$ must hold for any Λ, then we must have $v \leq (1-\epsilon)\lambda_H$. Meanwhile, $\frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)(1-\alpha\beta)\lambda_H < v$ implies $v > (1-\epsilon)(1-\alpha\beta)\lambda_H$ and $\Lambda < \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c}$. Moreover, since $s^q = \frac{1}{(1-\epsilon)\lambda_H}\left[v - \frac{c\Lambda}{2\gamma(1+\psi)}\right]$, the smaller Λ, the larger s^q. So to maximize s^q, Λ should be set infinitesimally close to zero, i.e., $\Lambda^\star = \iota$, so the maximum participation size $s^\star = \frac{v-\iota}{(1-\epsilon)\lambda_H}$. Therefore, when $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$, $\Lambda^\star = \iota$ can maximize the participation rate, and all non-committed low type users participate and the fraction of non-committed high type users who participate is $f_H^\star = \frac{(v-\iota)/((1-\epsilon)\lambda_H)-(1-\alpha\beta)}{\alpha\beta}$. By Lemma 1, in this range of v, $s^{sq} = \max\{1-\alpha\beta, \frac{v}{2(1-\epsilon)\lambda_H}\} < s^\star$, so $\Lambda^\star$ improves the participation rate relative to the status quo.

(5) If $0 < u_{H|s}^{q,in-out} < u_{L|s}^{q,in-out}$, then all non-committed users prefer to join, implying in equilibrium $s^q = 1$. Substituting $s = 1$ into $0 < u_{H|s}^{q,in-out} < u_{L|s}^{q,in-out}$, we obtain $\frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)\lambda_H < v$, which is the sufficient and necessary condition for the existence of this equilibrium. To ensure we are able to achieve this equilibrium, we must have $v > (1-\epsilon)\lambda_H$ and $\Lambda < \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}$. Therefore, when $v > (1-\epsilon)\lambda_H$, $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}\right)$ can maximize the participation rate, and all non-committed users participate ($s^\star = 1$). By Lemma 1, in this range of v, $s^{sq} = \min\{\frac{v}{2(1-\epsilon)\lambda_H}, 1\}$. So, when $(1-\epsilon)\lambda_H < v < 2(1-\epsilon)\lambda_H$, $s^{sq} = \frac{v}{2(1-\epsilon)\lambda_H} < s^\star$, $\Lambda^\star$ improves the participation rate relative to the status quo; when $v \geq 2(1-\epsilon)\lambda_H$, $s^{sq} = 1 = s^\star$, $\Lambda^\star$ does not change the participation rate relative to the status quo.

Proposition 3 summarizes all the above 5 cases and highlights the conditions under which quota can be used to improve participation rate relative to the status quo.

Proof of Proposition 4. With an effective quota $\Lambda \in (0, \frac{\gamma(1+\psi)v}{c})$, all users will post up to the quota. And relative to the status quo, each individual posts less. So if an effective quota increases the total information posted in the community relative to the status quo, it must have attracted more users to participate in the community. However, this would never happen. We next show why it is the case.

First, given the participation size s, the total amount of information posted in the community under a quota Λ is $Q^q(s) = s\Lambda$. So equation (A.3) (i.e., a representative user i's net utility from participating in the community under a quota Λ) can be written as

$$u_{i|s}^{q,in-out} = \underbrace{v\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)}}_{i} - \underbrace{(1-\epsilon)Q^q(s)\lambda_i}_{ii}, i \in \{L, H\}. \tag{A.4}$$

Part (i) in (A.4) is user i's net posting benefit under Λ and is less than $\frac{r(1+\psi)v^2}{2c}$ due to $\Lambda < \frac{\gamma(1+\psi)v}{c}$, meaning that any effective quota would make each user enjoy less net posting benefit relative to the status quo. Now if $Q^q(s)$ is increased relative to the status quo, then Part (ii) in (A.4) is larger than in the status quo, meaning that each user will suffer more net privacy cost. Then each user will unambiguously enjoy less net utility from participating in the community due to the effective quota Λ relative to the status quo (since Part (i) is smaller and Part (ii) is larger), which is in contradiction with the prerequisite

that more users should participate relative to the status quo in order for the total information to increase.

Proof of Proposition 5.

We now derive the optimal quota to maximize social welfare, Λ^*.

(1) When $0 < v \leq (1-\epsilon)(1-\alpha)\lambda_L$: by Proposition 3, no quota can change the participation rate in the status quo. So in any quota, no non-committed users will participate (labeled as Case A). In Case A, the aggregate user welfare is ($\Lambda \in (0, \frac{\gamma(1+\psi)v}{c}]$):

$$\Pi_A^q(\Lambda) = (1-\alpha)\left\{[v - (1-\alpha+\epsilon\alpha)\bar{\lambda}]\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)}\right\}. \tag{A.5}$$

Note that $v - (1-\alpha+\epsilon\alpha)\bar{\lambda} < 0$ since $(1-\alpha+\epsilon\alpha)\bar{\lambda} > (1-\epsilon)(1-\alpha)\lambda_L$, and thus $\Pi_A^q(\Lambda)$ decreases in Λ. So the welfare optimal quota is $\Lambda^* = \iota$ and accordingly $\Pi^{q*} = -\iota$. With Λ^* in place, no non-committed users will participate and $s^{q*} = 1-\alpha$.

(2) When $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$: By Proposition 3, the best a quota can do is to attract a fraction of non-committed low type users to participate (labeled as Case B). In Case B, the requirement on Λ is $\Lambda \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha)\lambda_L]}{c}]$. And in equilibrium, non-committed low type users are indifferent between participating and not participating. That is, $u_{L|s}^{q,in-out} = 0 \Leftrightarrow v = \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)s^q\lambda_L$, where s^q is the equilibrium participation rate and changes as Λ changes. Using this equality, the aggregate user welfare in Case B can be written as

$$\begin{aligned}\Pi_B^q(\Lambda) &= (1-\alpha\beta)\left\{\left\{v - \frac{s^q}{1-\alpha\beta}[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H]\right\}\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)}\right\} \\ &= (1-\alpha\beta)\left[1 - \frac{\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H}{(1-\epsilon)(1-\alpha\beta)\lambda_L}\right]\left[v\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)}\right].\end{aligned} \tag{A.6}$$

Note that in (A.6), $1 - \frac{\bar{\lambda}-(1-\epsilon)\alpha\beta\lambda_H}{(1-\epsilon)(1-\alpha\beta)\lambda_L} < 0$ since $\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H > (1-\epsilon)(1-\alpha\beta)\lambda_L$, and thus $\Pi_B^q(\Lambda)$ decreases in Λ. Case A is also achievable but it obviously cannot do better than Case B.

So the welfare optimal quota is also $\Lambda^* = \iota$ and accordingly $\Pi^{q*} = -\iota$. With Λ^* in place, only a fraction of non-committed low type users will participate and $s^{q*} = \frac{v-\iota}{(1-\epsilon)\lambda_L}$.

(3) When $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H$: By Proposition 3, the best a quota can do is to attract all the non-committed low type users (labeled as Case C). In Case C, the requirement on Λ is $\Lambda \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_L]}{c})$. The aggregate user welfare in Case C is

$$\Pi_C^q(\Lambda) = (1-\alpha\beta)\left\{[v - (\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H)]\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)}\right\}. \tag{A.7}$$

We can easily derive the welfare optimal quota according to (A.7) is

$$\Lambda_C^* = \begin{cases} \frac{\gamma(1+\psi)[v-\bar{\lambda}+(1-\epsilon)\alpha\beta\lambda_H]}{c}, & \text{if } v > \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H; \\ \iota, & \text{if } v \leq \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H. \end{cases}$$

Accordingly,

$$\Pi_C^{q*} = \begin{cases} \frac{(1-\alpha\beta)\gamma(1+\psi)[v-\bar{\lambda}+(1-\epsilon)\alpha\beta\lambda_H]^2}{2c}, & \text{if } v > \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H; \\ -\iota, & \text{if } v \leq \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H. \end{cases}$$

Case A or B is also achievable but they obviously cannot do better than Case C. Note that $\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H > (1-\epsilon)(1-\alpha\beta)\lambda_L$, so the optimal quota in this case can be expressed as:

$$\Lambda^* = \begin{cases} \iota, & \text{if } (1-\epsilon)(1-\alpha)\lambda_L < v \leq \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H. \\ \frac{\gamma(1+\psi)[v-\bar{\lambda}+(1-\epsilon)\alpha\beta\lambda_H]}{c}, & \text{if } \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H; \end{cases} \tag{A.8}$$

The second case exists only if $\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H < (1-\epsilon)(1-\alpha\beta)\lambda_H \Leftrightarrow \epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$. With Λ^* in place, only all non-committed low type users will participate and $s^{q*} = 1-\alpha\beta$.

(4) $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$: By Proposition 3, the best a quota can do is to attract all the non-committed low type users and a fraction of non-committed high type users (labeled as Case D). In Case D, the requirement on Λ is $\Lambda \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c})$. And in equilibrium, non-committed high type users are indifferent between participating and not participating. That is, $u_{H|s}^{q,in-out} = 0 \Leftrightarrow v = \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)s^q\lambda_H$, where s^q is the equilibrium participation rate and changes as Λ changes. Using this equality, the aggregate user welfare in Case D can be written as

$$\Pi_D^q(\Lambda) = (v - s\bar{\lambda})\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)} = \left[1 - \frac{\bar{\lambda}}{(1-\epsilon)\lambda_H}\right]\left[v\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)}\right]. \tag{A.9}$$

Note that $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H}) \Leftrightarrow \bar{\lambda} > (1-\epsilon)\lambda_H \Leftrightarrow 1 - \frac{\bar{\lambda}}{(1-\epsilon)\lambda_H} < 0$, so when $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, $\Pi_D^q(\Lambda)$ decreases in Λ; when $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, $\Pi_D^q(\Lambda)$ increases in $\Lambda \in (0, \frac{\gamma(1+\psi)v}{c}]$. Therefore, we can derive the welfare optimal quota according to (A.9)

$$\Lambda_D^* = \begin{cases} \iota, & \text{if } \epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H}); \\ \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c}, & \text{if } \epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H}). \end{cases} \tag{A.10}$$

• When $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, Case C, B, or A obviously cannot do better than Case D. So the optimal quota is $\Lambda^* = \iota$, and all non-committed low type and a fraction of non-committed high type users will participate and $s^{q*} = \frac{v-\iota}{(1-\epsilon)\lambda_H}$.

• When $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, the optimal quota Λ_D^* in equation (A.10) degenerates to Case C wherein only all the non-committed low type users participate (because Λ_D^* is the corner solution). Moreover, we need to consider when $v > 2(1-\epsilon)(1-\alpha\beta)\lambda_H \Leftrightarrow \frac{\gamma(1+\psi)v}{c} < \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c}$, the corner solution is not achievable.

Under $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, consider:

(4.1) If $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq \min\{(1-\epsilon)\lambda_H, 2(1-\epsilon)(1-\alpha\beta)\lambda_H\}$, we need to compare the corner solution with the interior solution in Case C to see when the interior solution is achievable (if yes, the interior solution is globally optimal; if not the corner solution is globally optimal) (Case B or A is

obviously inferior). It is easy to obtain that

$$\Lambda^* = \begin{cases} \frac{\gamma(1+\psi)[v-\bar{\lambda}+(1-\epsilon)\alpha\beta\lambda_H]}{c}, & \text{if } v \leq (1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda} \\ \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c}, & \text{if } v > (1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda}. \end{cases} \tag{A.11}$$

With Λ^* in place, only all non-committed low type users will participate and $s^{q*} = 1 - \alpha\beta$.

(4.2) If $2(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$, Case C is not achievable now. Moreover, (A.9) is increasing in $\Lambda \in (0, \frac{\gamma(1+\psi)}{c}]$, so $\Lambda^* = \frac{\gamma(1+\psi)}{c}$ and is ineffective. With Λ^* in place, all non-committed low type users and a fraction of non-committed high type users will participate and $s^{q*} = \frac{v}{2(1-\epsilon)\lambda_H}$.

(5) When $v > (1-\epsilon)\lambda_H$: By Proposition 3, the best a quota can do is to attract all the non-committed users (labeled as Case E). In Case E, the requirement on Λ is $\Lambda \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}]$. The aggregate user welfare in Case E is

$$\Pi_E^q(\Lambda) = (v - \bar{\lambda})\Lambda - \frac{c\Lambda^2}{2\gamma(1+\psi)}. \tag{A.12}$$

Note that $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H}) \Leftrightarrow \bar{\lambda} > (1-\epsilon)\lambda_H$. We can derive the welfare optimal quota according to (A.12).

- When $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$,

$$\Lambda_E^* = \begin{cases} \iota, & \text{if } v \leq \bar{\lambda}; \\ \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}, & \text{if } v > \bar{\lambda}. \end{cases} \tag{A.13}$$

It is easy to check Λ_E^* is indeed globally optimal. So Λ^* is expressed as in (A.13). With Λ^*, all non-committed users will participate.

- When $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$,

$$\Lambda_E^* = \begin{cases} \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}, & \text{if } v \leq 2(1-\epsilon)\lambda_H - \bar{\lambda}; \\ \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}, & \text{if } v > 2(1-\epsilon)\lambda_H - \bar{\lambda}. \end{cases} \tag{A.14}$$

We need to see if Case E is globally optimal quota. Under $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$, consider:

(5.1) If $(1-\epsilon)\lambda_H < v < \min\{2(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)\lambda_H - \bar{\lambda}\}$: in this case Λ_E^* is the corner solution and can be easily proved to be inferior to Case C. Then similar analysis as in (4.1) would yield

$$\Lambda^* = \begin{cases} \frac{\gamma(1+\psi)[v-\bar{\lambda}+(1-\epsilon)\alpha\beta\lambda_H]}{c}, & \text{if } v \leq (1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda} \\ \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c}, & \text{if } v > (1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda}. \end{cases} \tag{A.15}$$

With Λ^* in place, only all non-committed low type users will participate and $s^{q*} = 1 - \alpha\beta$.

(5.2) If $\max\{(1-\epsilon)\lambda_H, 2(1-\epsilon)(1-\alpha\beta)\lambda_H\} \leq v \leq 2(1-\epsilon)\lambda_H - \bar{\lambda}$, in this case Λ_E^* is the corner solution and can be easily proved to be inferior to the status quo (Case C is not achievable now). So $\Lambda^* = \frac{\gamma(1+\psi)}{c}$ and is ineffective. With Λ^* in place, all non-committed low type users and a fraction of

non-committed high type users will participate and $s^{q*} = \frac{v}{2(1-\epsilon)\lambda_H}$.

(5.3) If $v > \max\{2(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)\lambda_H - \bar{\lambda}\}$, in this case Λ_E^* is the interior solution and is indeed globally optimal. That is, $\Lambda^* = \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}$. With Λ^* in place, all non-committed users will participate and $s^{q*} = 1$.

Appendix A in the main manuscript summarizes all the above (1) through (5). We present the optimal quotas separately according to the value of ϵ. We have also combined adjacent ranges of v under which the optimal quotas are common.

Lastly, it is easy to check that except for conditions (4.2) and (5.2), all the optimal quotas identified above are effective, i.e., $\Lambda^* < \frac{\gamma(1+\psi)v}{c}$. Also note that when the quota is ineffective (i.e., $\Lambda = \frac{\gamma(1+\psi)v}{c}$), the quota is equivalent the status quo. In the process of searching for the optimal quota above, the ineffective quota has also been considered. So when optimal quota is effective, then it will always improve the welfare relative to the status quo. In conditions (4.2) and (5.2), which can be combined to be $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$ and $2(1-\epsilon)(1-\alpha\beta)\lambda_H < v < 2(1-\epsilon)\lambda_H - \bar{\lambda}$, the optimal quota is ineffective. That is, $\Lambda^* = \frac{\gamma(1+\psi)v}{c}$ and the optimal quota retains the status quo.

It is easy to check that the participation rate in the community is weakly improved relative to the status quo. The aggregate privacy damage $\xi = \frac{\psi\bar{\theta}}{1+\psi}Q$. When the optimal quota is effective, by Proposition 4, it will always reduce the total amount of information in the community (i.e., Q), so it will always reduce the aggregate privacy damage.

Proof of Proposition 6. By Proposition 1, a positive nudge weakly reduces the participation rate relative to the status quo. Meanwhile, by Proposition 3, quota can be used to improve the participation rate relative to the status quo in some ranges. So obviously, a quota weakly dominates a nudge in increasing user participation.

Now consider welfare, we separate into two cases according to the value of ϵ:

• When $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$: By Lemma 2 and Proposition 2, $\Pi^{n*} = 0$ when $0 < v < 2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H]$ and $\Pi^{n*} = \Pi^{sq}$ when $v \geq 2[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H]$. Meanwhile, by Proposition 5 (and its proof) and Appendix A, $\Pi^{q*} = 0$ when $0 < v < \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H$ and $\Pi^{q*} \geq \Pi^{sq}$ when $v \geq \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H$. So $\Pi^{q*} \geq \Pi^{n*}$.

• When $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$: By Lemma 2 and Proposition 2, $\Pi^{n*} = 0$ when $0 < v < 2\bar{\lambda}$ and $\Pi^{n*} = \Pi^{sq}$ when $v \geq 2\bar{\lambda}$. Meanwhile, by Proposition 5 (and its proof) and Appendix A, $\Pi^{q*} = 0$ when $0 < v < \bar{\lambda}$ and $\Pi^{q*} > \Pi^{sq}$ when $v \geq \bar{\lambda}$. So $\Pi^{q*} \geq \Pi^{n*}$.

A technical issue worth clarifying here: we consider an extremely costly nudge (i.e., $\tau = v$ and thus $\Pi^n = 0$) equivalent to an extremely harsh quota (i.e., $\Lambda = \iota$ and thus $\Pi^q = -\iota$). This is purely because we have assumed $\Lambda \neq 0$ to avoid some trivial discussion.

Section 3.3.5. Formal results about overlapping of participation-optimal and welfare-optimal quotas. When the community owner's objective is to maximize participation, by Proposition 3, a quota can be used to achieve this goal. We now examine when the welfare-optimal quota (Λ^*) overlaps with the participation-optimal quota ($\Lambda^\star$). Referring to Proposition 3 and Appendix A in the

main manuscript,

(i) When $\epsilon \leq (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$:

(i.1) If $0 < v \leq (1-\epsilon)(1-\alpha)\lambda_L$, $\Lambda^* = \iota$, $\Lambda^\star \in (0, \frac{\gamma(1+\psi)v}{c}]$, and thus $\Lambda^* \in \Lambda^\star$.

(i.2) If $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$, $\Lambda^* = \iota$, $\Lambda^\star = \iota$, and thus $\Lambda^* = \Lambda^\star$.

(i.3) If $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq \bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H$, $\Lambda^* = \iota$, $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_L]}{c}\right)$, and thus $\Lambda^* \in \Lambda^\star$.

(i.4) If $\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H < v \leq (1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)[v-\bar{\lambda}+(1-\epsilon)\alpha\beta\lambda_H]}{c}$. $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_L]}{c}\right)$ when $v < \min\{(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)(1-\alpha\beta)\lambda_L\}$, so $\Lambda^* \in \Lambda^\star$; $\Lambda^\star \in \left(0, \frac{\gamma(1+\psi)v}{c}\right)$ when $2(1-\epsilon)(1-\alpha\beta)\lambda_L < v < (1-\epsilon)(1-\alpha\beta)\lambda_H$, so $\Lambda^* \in \Lambda^\star$; $\Lambda^\star = \iota$ when $(1-\epsilon)(1-\alpha\beta)\lambda_H < v < (1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda}$, so $\Lambda^* \neq \Lambda^\star$.

(i.5) If $(1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda} < v \leq 2(1-\epsilon)(1-\alpha\beta)\lambda_H$, $\Lambda^* = \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_H]}{c}$. $\Lambda^\star = \iota$ when $(1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda} < v < (1-\epsilon)\lambda_H$, so $\Lambda^* \neq \Lambda^\star$; $\Lambda^\star \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}]$ when $\max\{(1-\epsilon)(2-\alpha\beta)\lambda_H - \bar{\lambda}, (1-\epsilon)\lambda_H\} < v \leq 2(1-\epsilon)(1-\alpha\beta)\lambda_H$, so $\Lambda^* \neq \Lambda^\star$.

(i.6) If $2(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq 2(1-\epsilon)\lambda_H - \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)v}{c}$. $\Lambda^\star = \iota$ when $2(1-\epsilon)(1-\alpha\beta)\lambda_H < v < (1-\epsilon)\lambda_H$, so $\Lambda^* \neq \Lambda^\star$; $\Lambda^\star \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}]$ when $\max\{2(1-\epsilon)(1-\alpha\beta)\lambda_H, (1-\epsilon)\lambda_H\} < v \leq 2(1-\epsilon)\lambda_H - \bar{\lambda}$, so $\Lambda^* \neq \Lambda^\star$.

(i.7) If $v > \max\{2(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)\lambda_H - \bar{\lambda}\}$, $\Lambda^* = \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}$. $\Lambda^\star \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}]$ when $\max\{2(1-\epsilon)(1-\alpha\beta)\lambda_H, 2(1-\epsilon)\lambda_H - \bar{\lambda}\} < v < 2(1-\epsilon)\lambda_H$, so $\Lambda^* \in \Lambda^\star$; $\Lambda^\star \in (0, \frac{\gamma(1+\psi)v}{c}]$ when $v > 2(1-\epsilon)\lambda_H$, so $\Lambda^* \in \Lambda^\star$.

(ii) When $\epsilon > (1-\beta)(1-\frac{\lambda_L}{\lambda_H})$:

(ii.1) If $0 < v \leq (1-\epsilon)(1-\alpha)\lambda_L$, $\Lambda^* = \iota$, $\Lambda^\star \in (0, \frac{\gamma(1+\psi)v}{c}]$, and thus $\Lambda^* \in \Lambda^\star$.

(ii.2) If $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$, $\Lambda^* = \iota$, $\Lambda^\star = \iota$, and thus $\Lambda^* = \Lambda^\star$.

(ii.3) If $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H$, $\Lambda^* = \iota$, $\Lambda^\star \in \left(0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)(1-\alpha\beta)\lambda_L]}{c}\right)$, and thus $\Lambda^* \in \Lambda^\star$.

(ii.4) If $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$, $\Lambda^* = \iota$, $\Lambda^\star = \iota$, and thus $\Lambda^* = \Lambda^\star$.

(ii.5) If $(1-\epsilon)\lambda_H < v \leq \bar{\lambda}$, $\Lambda^* = \iota$, $\Lambda^\star \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}]$, and thus $\Lambda^* \in \Lambda^\star$.

(ii.6) If $v > \bar{\lambda}$, $\Lambda^* = \frac{\gamma(1+\psi)(v-\bar{\lambda})}{c}$. $\Lambda^\star \in (0, \frac{2\gamma(1+\psi)[v-(1-\epsilon)\lambda_H]}{c}]$ when $\bar{\lambda} < v < 2(1-\epsilon)\lambda$, and thus $\Lambda^* \in \Lambda^\star$; $\Lambda^\star \in (0, \frac{\gamma(1+\psi)v}{c}]$ when $v > 2(1-\epsilon)\lambda_H$, so $\Lambda^* \in \Lambda^\star$.

Proof of Proposition 7. Given a free allowance $\tilde{\Lambda} \in (0, \frac{\gamma(1+\psi)v}{c}]$, if a user i doesn't post in excess of this $\tilde{\Lambda}$, then the equilibrium quantities will be just the same as those in a pure quota. So by Proof of Proposition 3, we have $x_{ij} = \frac{\Lambda}{(1+\psi)\gamma}$, $y_{ij} = \frac{\psi\Lambda}{(1+\psi)\gamma}$, $x_{ik} = \frac{\delta\Lambda}{(1+\psi)\gamma}$ and $y_{ik} = \frac{\delta\psi\Lambda}{(1+\psi)\gamma}$. If user i

posts in excess of the free allowance, then she will face additional nudging cost for the excessive part of information.

Denote the amount of non-sensitive information user i would post about user j as x_{ij}^c. If user i decides to post in excess of $\tilde{\Lambda}$, she makes the following decision: $x_{ij}^c = \arg\max_{x_{ij} > \frac{\Lambda}{(1+\psi)\gamma}} \; vx_{ij} - \frac{cx_{ij}^2}{2} - \tilde{\tau}[x_{ij} - \frac{\Lambda}{(1+\psi)\gamma}] = \frac{v-\tilde{\tau}}{c}$. Similarly, we derive the optimal quantities of the other three types of information: $y_{ij}^c = \frac{\psi(v-\tilde{\tau})}{c}$, $x_{ik}^c = \frac{\delta(v-\tilde{\tau})}{c}$, and $y_{ik}^c = \frac{\psi\delta(v-\tilde{\tau})}{c}$. So the total amount of information each user would post is $Q^c = n(x_{ij}^c + y_{ij}^c) + (1-n)(x_{ik}^c + y_{ik}^c) = \frac{\gamma(1+\psi)(v-\tilde{\tau})}{c}$. So if $Q^c = \frac{\gamma(1+\psi)(v-\tilde{\tau})}{c} > \tilde{\Lambda}$, then users will post in excess of the free allowance, otherwise they just use up the free allowance.

We now see, when users post in excess of the free allowance (i.e., $\frac{\gamma(1+\psi)(v-\tilde{\tau})}{c} > \tilde{\Lambda} \Leftrightarrow \tilde{\tau} < v - \frac{c\tilde{\Lambda}}{\gamma(1+\psi)}$), whether the composite policy will outperform the pure quota in terms of participation and welfare.

Substituting the equilibrium quantities of information derived above back into equation (2) and (3) in the main manuscript, we can derive a representative user i's net utility from participating in the community condition on participation size s

$$u_{i|s}^{c,in-out} = \frac{\gamma(1+\psi)(v-\tilde{\tau})}{2c}[v - \tilde{\tau} - 2(1-\epsilon)s\lambda_i] + \tilde{\tau}\tilde{\Lambda}. \tag{A.16}$$

We calculate $\partial u_{i|s}^{c,in-out}/\partial\tilde{\tau} = \frac{\gamma(1+\psi)}{c}[(1-\epsilon)s\lambda_i + \tilde{\tau} - v] + \tilde{\Lambda}$ for later use in the proof. We next first prove that the composite policy cannot improve the participation rate better than a pure quota.

(i) When $0 < v \leq (1-\epsilon)(1-\alpha)\lambda_L$: Obviously, $\partial u_{i|s}^{c,in-out}/\partial\tilde{\tau} > 0$. Note that $\tilde{\tau} < v - \frac{c\tilde{\Lambda}}{\gamma(1+\psi)}$, so $u_{i|s}^{c,in-out} < u_{i|s,\tau=v-\frac{c\tilde{\Lambda}}{\gamma(1+\psi)}}^{c,in-out} = \tilde{\Lambda}\left[v - \frac{c\tilde{\Lambda}}{2\gamma(1+\psi)} - (1-\epsilon)s\lambda_i\right]$. The last term is user i's net utility in the pure quota. So each user would obtain less net utility from participating in the community in this composite policy than in the pure quota and thus this composite policy cannot improve the participation rate better than a pure quota.

(ii) When $(1-\epsilon)(1-\alpha)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_L$: By Proposition 3, the best a quota can do is to attract a fraction of non-committed low type users to participate and $u_{L|s}^{q,in-out} = 0 \Leftrightarrow v = \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)s\lambda_L$, which implies $\partial u_{L|s}^{c,in-out}/\partial\tilde{\tau} = \frac{\gamma(1+\psi)\tilde{\tau}}{c} + \frac{\tilde{\Lambda}}{2} > 0$. Similar logic in (i) yields that low type users obtain less net utility from participating in the community in this composite policy than in the pure quota. But for the composite policy to improve the participation rate better than a pure quota, it should make low type users obtain at least the same net utilities. A contradiction. So the composite policy cannot improve the participation rate better than a pure quota.

(iii) When $(1-\epsilon)(1-\alpha\beta)\lambda_L < v \leq (1-\epsilon)(1-\alpha\beta)\lambda_H$: By Proposition 3, the best a quota can do is to attract all the non-committed low type users. Obviously, $\partial u_{H|s}^{c,in-out}/\partial\tilde{\tau} > 0$. Similar logic in (i) yields that high type users obtain less net utility from participating in the community in this composite policy than in the pure quota. But for the composite policy to improve the participation rate better than a pure quota, it should make high type users obtain more net utilities. A contradiction. So the composite policy cannot improve the participation rate better than a pure quota.

(iv) When $(1-\epsilon)(1-\alpha\beta)\lambda_H < v \leq (1-\epsilon)\lambda_H$: By Proposition 3, the best a quota can do is to attract all the non-committed low type users and a fraction of non-committed high type users. Moreover,

$u_{H|s}^{q,in-out} = 0 \Leftrightarrow v = \frac{c\Lambda}{2\gamma(1+\psi)} + (1-\epsilon)s\lambda_H$, which implies $\partial u_{H|s}^{c,in-out}/\partial\tilde{\tau} = \frac{\gamma(1+\psi)\tilde{\tau}}{c} + \frac{\tilde{\Lambda}}{2} > 0$. Similar logic in (i) yields that high type users obtain less net utility from participating in the community in this composite policy than in the pure quota. But for the composite policy to improve the participation rate better than a pure quota, it should at least make high type users obtain the same net utilities. A contradiction. So the composite policy cannot improve the participation rate better than a pure quota.

(v) When $v > (1-\epsilon)\lambda_H$: By Proposition 3, the best a quota can do is to attract all the non-committed users. So no way for the composite policy to improve the participation rate better than the pure quota.

In summary, the composite policy cannot improve the participation rate better than the pure quota.

Now consider the welfare, now that the composite policy cannot improve the participation rate better than the pure quota, we just need to see if the composite policy can improve the welfare better than the pure quota when it produces a lower participation rate than the pure quota. But this is impossible because the welfare-optimal quota identified in Proposition 5, by construction, considers all the possible participation rates and then select the right participation rate that can produce the optimal welfare. In other words, if there indeed exists a lower participation rate that can allows the composite policy to outperform the pure quota, the pure quota can also achieve it without the additional nudging and thus deliver a larger welfare.

We can also prove such composite policy will not increase the total quantity of information posted in the community when compared with the status quo. To do so, we only need to see, when users post in excess of the free allowance (i.e., $\frac{\gamma(1+\psi)(v-\tilde{\tau})}{c} > \tilde{\Lambda}$), whether the composite policy will increase total information relative to the status quo.

User i's net utility in (A.16) can be written as

$$u_{i|s}^{c,in-out} = \underbrace{\frac{\gamma(1+\psi)(v-\tilde{\tau})^2}{2c} + \tilde{\tau}\tilde{\Lambda}}_{i} - \underbrace{(1-\epsilon)Q^c(s)\lambda_i}_{ii}. \tag{A.17}$$

Part (i) in (A.17) is less than $\frac{\gamma(1+\psi)v^2}{c}$ due to $\tilde{\Lambda} < \frac{\gamma(1+\psi)(v-\tilde{\tau})}{c}$, implying each user still enjoy less net utility in this composite policy compared with in the status quo. Part (ii) in (A.17) is the net privacy cost user i suffers ($Q^c(s)$ is the total amount of information posted in the community in this composite policy). So the same contradiction as in the Proof of Proposition 4 will also occur here. Therefore, this composite policy cannot increase total information relative to the status quo either.

Proof of Proposition 8. Proof of Proposition 8 is evident from the discussion that accompanies Proposition 8 in the main manuscript.

II. Detailed Results in Numerical Example and Extensions

A. Numerical Example

Here we explain how we derive users' posting quantities in the targeted nudge and quota mechanisms when "false positive" and "false negative" are allowed. Denote the probability of "false positive" as p_x and the probability of "false negative" as p_y. Note that only detected "sensitive information" would be regulated by nudge or quota.

In targeted nudging, user i's net utility is

$$u_{i|s}^{tn,in-out} = n\left[vx_{ij} - \frac{cx_{ij}^2}{2} + vy_{ij} - \frac{cy_{ij}^2}{2\psi} - \tau(p_x x_{ij} + (1-p_y)y_{ij})\right] + (1-n)\left[vx_{ik} - \frac{cx_{ik}^2}{2\delta} + vy_{ik} - \frac{cy_{ik}^2}{2\delta\psi} - \tau(p_x x_{ik} + (1-p_y)y_{ik})\right] + (1-\epsilon)(eQ_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i}).$$

FOCs w.r.t this equation would give user i's optimal posting quantities in targeted nudging,

$$x_{ij}^{tn} = \frac{v - p_x\tau}{c},\ y_{ij}^{tn} = \frac{\psi[v-(1-p_y)\tau]}{c},\ x_{ik}^{tn} = \frac{\delta(v-p_x\tau)}{c},\ y_{ik}^{tn} = \frac{\psi\delta[v-(1-p_y)\tau]}{c}.$$

In targeted quota, user i's net utility is

$$u_{i|s}^{tq,in-out} = n\left[vx_{ij} - \frac{cx_{ij}^2}{2} + vy_{ij} - \frac{cy_{ij}^2}{2\psi}\right] + (1-n)\left[vx_{ik} - \frac{cx_{ik}^2}{2\delta} + vy_{ik} - \frac{cy_{ik}^2}{2\delta\psi}\right] + (1-\epsilon)(eQ_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i}),$$

$$\text{with the constraint } n[p_x x_{ij} + (1-p_y)y_{ij}] + (1-n)[p_x x_{ik} + (1-p_y)y_{ik}] = \Lambda$$

As in the Proof of Proposition 3, we can define a Lagrange function to derive user i's optimal posting quantities in targeted quota,

$$x_{ij}^{tq} = \left[1 - \frac{p_x + \psi(1-p_y)}{p_x + \psi(1-p_y)^2/p_x}\right]\frac{v}{c} + \frac{\Lambda}{[p_x + \psi(1-p_y)^2/p_x]\gamma},$$

$$y_{ij}^{tq} = \left[1 - \frac{p_x + \psi(1-p_y)}{p_x^2/(1-p_y) + \psi(1-p_y)}\right]\frac{\psi v}{c} + \frac{\psi\Lambda}{[p_x^2/(1-p_y) + \psi(1-p_y)]\gamma},$$

$$x_{ik}^{tq} = \left[1 - \frac{p_x + \psi(1-p_y)}{p_x + \psi(1-p_y)^2/p_x}\right]\frac{\delta v}{c} + \frac{\delta\Lambda}{[p_x + \psi(1-p_y)^2/p_x]\gamma},$$

$$y_{ik}^{tq} = \left[1 - \frac{p_x + \psi(1-p_y)}{p_x^2/(1-p_y) + \psi(1-p_y)}\right]\frac{\delta\psi v}{c} + \frac{\delta\psi\Lambda}{[p_x^2/(1-p_y) + \psi(1-p_y)]\gamma}.$$

In the numerical example, we set $p_x = p_y = 0.1$ and fix all other exogenous parameters, including the values for τ in targeted nudge and Λ in targeted quota (see specific values in the main text). We choose six different values of v. For each v, using all the fixed parameters, we can compute each user's optimal posting quantities according to the equations derived above. Then we substitute the optimal posting quantities back into each user's net utility function to determine their participation decisions. Finally, all the aggregate measures (i.e., total quantity of information posted, social welfare, and total

privacy damage) can be calculated once participation rate is determined.

B. Extensions

(i) Heterogeneity in "friendship". We assume $n_i \in U[0,1]$ and $\theta_i = 1 - n_i$, then $\lambda_i = \frac{\psi\theta_i - e(1+\psi) - \omega}{1+\psi}$, $\gamma_i = n_i + (1-n_i)\delta = 1 - (1-\delta)\theta_i$. User i's net utility from participating in the community in the status quo (conditional on the total information posted in the community Q^{sq}) is given by

$$u_i^{sq,in-out} = \frac{\gamma_i(1+\psi)v^2}{2c} - (1-\epsilon)Q^{sq}\lambda_i \tag{A.18}$$

Note that $u_i^{sq,in-out}$ in Equation (A.18) decreases in θ_i because γ_i decreases in θ_i and λ_i increases with θ_i. So less privacy sensitive (more connected) users are more likely to participate. That means we can characterize equilibrium participation rate by a cutoff user type (θ^o) such that users with privacy sensitivity below θ^o will participate in the community and the others will not. So θ^o indicates the equilibrium participation rate in the community.

Note also that the utility function in Equation (A.18) increases with v. As v becomes sufficiently large, every user will get positive net utility and thus participate in the community, which implies $\theta^o = 1$. Now focus on $\theta^o < 1$, then the condition to solve for θ^o is that the cutoff type has a net utility of zero:

$$u_i^{sq,in-out}(\theta^o) = \frac{\gamma_i(\theta^o)(1+\psi)v^2}{2c} - (1-\epsilon)Q^{sq}(\theta^o)\lambda_i(\theta^o) = 0 \tag{A.19}$$

The total quantity of information posted in the community in equilibrium conditional on θ^o is given by

$$Q^{sq}(\theta^o) = \int_0^{\theta^o} \frac{\gamma_i(\theta_i)(1+\psi)v}{c} d\theta_i = \theta^o[1 - \frac{1-\delta}{2}\theta^o]\frac{(1+\psi)v}{c} \tag{A.20}$$

Combining Equation (A.19) and (A.20), we can derive the condition to solve for θ^o:

$$v = \frac{(1-\epsilon)[\psi\theta^o - e(1+\psi) - \omega]\theta^o[2-(1-\delta)\theta^o]}{(1+\psi)[1-(1-\delta)\theta^o]} \tag{A.21}$$

RHS of Equation (A.21) is a strictly increasing function of θ^o because $\psi\theta^o - e(1+\psi) - \omega > 0$ and increases with θ^o, $\theta^o[2-(1-\delta)\theta^o] > 0$ and increases with θ^o, and $1-(1-\delta)\theta^o > 0$ and decreases with θ^o. So as v increases, θ^o will increase. But note the condition of $\theta^o < 1$, which requires $v < (1-\epsilon)(1+\frac{1}{\delta})\frac{\psi - e(1+\psi) - \omega}{1+\psi}$. When $v \geq (1-\epsilon)(1+\frac{1}{\delta})\frac{\psi - e(1+\psi) - \omega}{1+\psi}$, then all users will get positive net utility and participate in the community. So we obtain Lemma xx.

Nudge

Note that the impact of a nudge relative to the status quo is decreasing v. So according to Lemma xx, implementing a nudge will (weakly) decrease participate rate relative to the status quo. According to Equation (A.20), it is easy to verify that $Q^{sq}(\theta^o)$ increases with v and θ^o. That is, the total quantity of information posted in the community will decrease as posting benefit decreases and participation rate decreases, which are the impacts from a nudge. To examine the impact of a nudge on social welfare, we first derive the social welfare in the status quo:

$$\begin{aligned}\Pi^{sq} &= \int_0^{\theta^o} \left[\frac{\gamma_i(\theta_i)(1+\psi)v^2}{2c} - Q^{sq}(\theta^o)\lambda_i(\theta_i)\right] d\theta_i + \int_{\theta^o}^1 [-\epsilon Q^{sq}(\theta^o)\lambda_i(\theta_i)] d\theta_i \\ &= \frac{(1+\psi)v}{c}\theta^o[1 - \frac{1-\delta}{2}\theta^o]\left\{\left[\frac{v}{2} - \frac{\psi(\theta^o)^2}{2(1+\psi)} + (e + \frac{\omega}{1+\psi})\theta^o\right] - \right. \\ &\left. \epsilon(1-\theta^o)\left[\frac{\psi(1+\theta^o)}{2(1+\psi)} - (e + \frac{\omega}{1+\psi})\right]\right\}\end{aligned} \tag{A.22}$$

Note the equality between θ^o and v in Equation (A.21). So Π^{sq} can be viewed as a function of θ^o- the equilibrium participation rate. The function in Equation (A.22) is obviously a highly non-monotonic function. Figure A.1 illustrates this via an example with $\psi = 0.5$, $\epsilon = 0.01$, $e = 0.01$, $\omega = 0.01$, $\delta = 0.5$ and $c = 1$ (similar to Figure 3 in the main text). As participation rate increases, the social welfare may decrease. Therefore, we have shown numerically that, when a nudge is implemented, the participation rate decreases, which may result in increase in social welfare. This result shows that the effects of a nudge is similar to that in the main model.

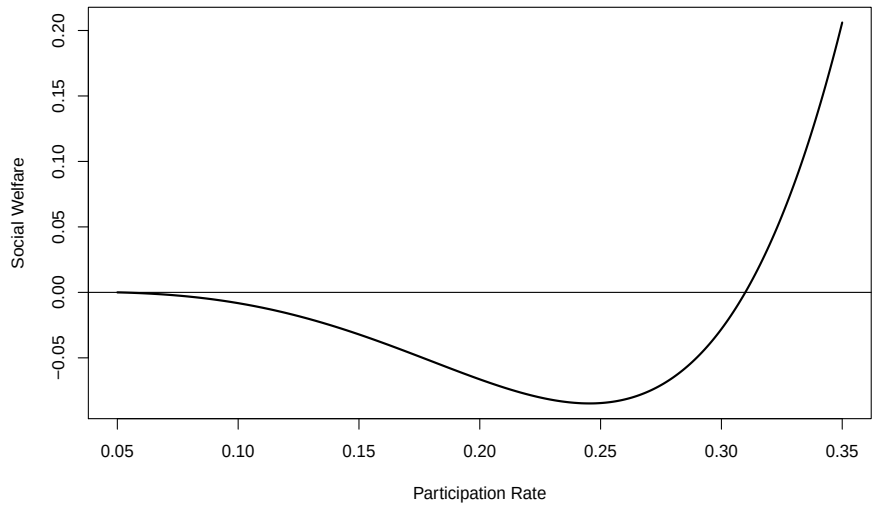

Figure A.1: Social welfare as a function of participation rate in status quo

Quota

We consider a quota of the following form: $\Lambda_i = f \cdot \frac{\gamma_i(1+\psi)v}{c}$, where $f \in [0,1]$ and $\frac{\gamma_i(1+\psi)v}{c}$ is the total amount of information user i will post by her own choice in the status quo. User i's net utility from participating in the community under the quota (conditional on the total information posted in the community Q^q) is given by

$$u_i^{q,in-out} = v\Lambda_i - \frac{c\Lambda_i^2}{2\gamma_i(1+\psi)} - (1-\epsilon)Q^q\lambda_i \tag{A.23}$$

$u_i^{q,in-out}$ in Equation (A.23) decreases in θ_i because $v\Lambda_i - \frac{c\Lambda_i^2}{2\gamma_i(1+\psi)}$ decreases in θ_i (this in turn is because $v\Lambda_i - \frac{c\Lambda_i^2}{2\gamma_i(1+\psi)}$ increases with Λ_i and Λ_i decreases with θ_i) and λ_i increases with θ_i. So we can

again characterize equilibrium participation rate by a cutoff user type (θ^o) such that users with privacy sensitivity below θ^o will participate in the community and the others will not.

Focus on $\theta^o < 1$, then the condition to solve for θ^o is that the cutoff type has a net utility of zero:

$$u_i^{q,in-out}(\theta^o) = v\Lambda_i(\theta^o) - \frac{c\Lambda_i(\theta^o)^2}{2\gamma_i(\theta^o)(1+\psi)} - (1-\epsilon)Q^q(\theta^o)\lambda_i(\theta^o) = 0 \tag{A.24}$$

And the total quantity of information posted in the community in equilibrium conditional on θ^o is given by

$$Q^q(\theta^o) = \int_0^{\theta^o} f\frac{\gamma_i(\theta_i)(1+\psi)v}{c}d\theta_i = f\theta^o[1 - \frac{1-\delta}{2}\theta^o]\frac{(1+\psi)v}{c} \tag{A.25}$$

Combining Equation (A.24) and (A.25), we can derive the condition to solve for θ^o:

$$(2-f)v = \frac{(1-\epsilon)[\psi\theta^o - e(1+\psi) - \omega]\theta^o[2-(1-\delta)\theta^o]}{(1+\psi)[1-(1-\delta)\theta^o]} \tag{A.26}$$

Now comparing Equation (A.26) with Equation (A.21), the RHSs of both equations are the same and increase with θ^o. Note that $(2-f)v \geq v$ (LHSs of the two equations), so the equilibrium θ^o under a quota will be larger than that in the status quo. In other words, implementing a quota to the status quo will increase the participation rate. This holds as long as the participation rate in the status quo is not 1 and there is room for improvement (i.e., $v < (1-\epsilon)(1+\frac{1}{\delta})\frac{\psi-e(1+\psi)-\omega}{1+\psi}$). The impact of a quota on the total amount of information posted in the community can still be seen by the dilemma highlighted in Proposition xx. As in the main model, a quota will decrease the total amount of information and hence the total privacy damage relative to the status quo.

The social welfare under a quota is given by

$$\begin{aligned}\Pi^q &= \int_0^{\theta^o}\left[(2f-f^2)\frac{\gamma_i(\theta_i)(1+\psi)v^2}{2c} - Q^q(\theta^o)\lambda_i(\theta_i)\right]d\theta_i + \int_{\theta^o}^1[-\epsilon Q^q(\theta^o)\lambda_i(\theta_i)]d\theta_i \\ &= f\cdot\frac{(1+\psi)v}{c}\cdot\theta^o[1-\frac{1-\delta}{2}\theta^o]\left\{\left[(1-\frac{f}{2})v - \frac{\psi(\theta^o)^2}{2(1+\psi)} + (e+\frac{\omega}{1+\psi})\theta^o\right] - \right. \\ &\left.\epsilon(1-\theta^o)\left[\frac{\psi(1+\theta^o)}{2(1+\psi)} - (e+\frac{\omega}{1+\psi})\right]\right\}\end{aligned} \tag{A.27}$$

Due to the complexity of the model, it is challenging to derive the closed-form optimal quota by comparing Π^q with Π^{sq}. We thus resort to numerical analysis to examine the impact of a quota. We set $f = 0.7$ and use the same set of parameter values as in Figure A.1. Figure A.2 shows that there exists a quota that increase the social welfare relative to the status quo, which is consistent with Proposition 5 and Figure 4 in the main text. Note that, we also vary the values of the key parameters - ψ and δ from 0.1 to 0.9, and solve the numerical example under the different values. Under all these cases, we discover consistent result that, with an appropriate quota, the social welfare can be increased relative to the status quo.

(ii) Nonlinear externality. Following the same approach in the main model, we can derive the equilibrium outcomes in the status quo as follows.

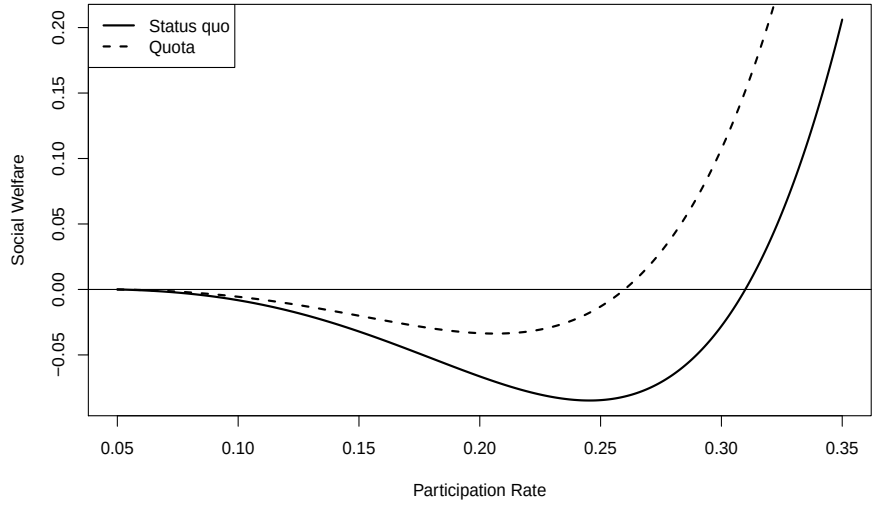

Figure A.2: Social welfare as a function of participation rate in status quo and quota

(1) When $0 < v < 2(1-\epsilon)(1-\alpha)^2\lambda_L$: no non-committed users will participate. $s^{sq} = 1-\alpha$. $Q^{sq} = \frac{(1-\alpha)\gamma(1+\psi)v}{c}$. $\Pi^{sq} = \frac{(1-\alpha)\gamma(1+\psi)v}{2c}[v - 2(1-\alpha)(1-\alpha+\epsilon\alpha)\bar{\lambda}]$. $\xi^{sq} = \frac{(1-\alpha)^2\gamma\psi\bar{\theta}v}{c}$.

(2) When $2(1-\epsilon)(1-\alpha)^2\lambda_L \leq v \leq 2(1-\epsilon)(1-\alpha\beta)^2\lambda_L$: only a fraction of non-committed low type users will participate. $s^{sq} = \sqrt{\frac{v}{2(1-\epsilon)\lambda_L}}$. $Q^{sq} = \frac{\gamma(1+\psi)v}{c}\sqrt{\frac{v}{2(1-\epsilon)\lambda_L}}$. $\Pi^{sq} = \frac{\gamma(1+\psi)v^2}{2c}\left\{(1-\alpha) - \frac{(1-\alpha+\epsilon\alpha)\bar{\lambda}}{(1-\epsilon)\lambda_L}\right\}$. $\xi^{sq} = \frac{\gamma\psi\bar{\theta}v^2}{2c(1-\epsilon)\lambda_L}$

(3) When $2(1-\epsilon)(1-\alpha\beta)^2\lambda_L < v < 2(1-\epsilon)(1-\alpha\beta)^2\lambda_H$: only all non-committed low type users will participate. $s^{sq} = 1-\alpha\beta$. $Q^{sq} = \frac{(1-\alpha\beta)\gamma(1+\psi)v}{c}$. $\Pi^{sq} = \frac{(1-\alpha\beta)\gamma(1+\psi)v}{2c}\left\{v - 2(1-\alpha\beta)[\bar{\lambda} - (1-\epsilon)\alpha\beta\lambda_H]\right\}$. $\xi^{sq} = \frac{(1-\alpha\beta)^2\gamma\psi\bar{\theta}v}{c}$.

(4) When $2(1-\epsilon)(1-\alpha\beta)^2\lambda_H \leq v \leq 2(1-\epsilon)\lambda_H$: all non-committed low type users and a fraction of non-committed high type users will participate. $s^{sq} = \sqrt{\frac{v}{2(1-\epsilon)\lambda_H}}$. $Q^{sq} = \frac{\gamma(1+\psi)v}{c}\sqrt{\frac{v}{2(1-\epsilon)\lambda_H}}$. $\Pi^{sq} = \frac{\gamma(1+\psi)v^2}{2c}\left[1 - \frac{\bar{\lambda}}{(1-\epsilon)\lambda_H}\right]$. $\xi_{sq} = \frac{\gamma\psi\bar{\theta}v^2}{2c(1-\epsilon)\lambda_H}$.

(5) When $v > 2(1-\epsilon)\lambda_H$: all non-committed users will participate. $s^{sq} = 1$. $Q^{sq} = \frac{\gamma(1+\psi)v}{c}$. $\Pi^{sq} = \frac{\gamma(1+\psi)v}{2c}(v - 2\bar{\lambda})$. $\xi^{sq} = \frac{\gamma\psi\bar{\theta}v}{c}$.

Comparing the cutoff value of v when non-committed low type users start to participate here with that in the main model, $2(1-\alpha)^2(1-\epsilon)\lambda_L < 2(1-\alpha)(1-\epsilon)\lambda_L$. Comparing the cutoff value of v when non-committed high type users start to participate here with that in the main model, $2(1-\alpha\beta)^2(1-\epsilon)\lambda_H < 2(1-\alpha\beta)(1-\epsilon)\lambda_H$. So we obtain Lemma 5.

As we can see, the corresponding impact is that the cutoff values of v which characterize different equilibrium outcomes change quantitatively. But all the results obtained in the main model can be verified to hold qualitatively.

(iii) Unintentional Disclosure of Sensitive Information. Users may not know the information they post is sensitive. We can modify the setup in line with this consideration. Assume users only make decisions about the total amount of information to be posted about friends and non-friends, denoted as

x_{ij} and x_{ik} respectively. There is a fraction, $\phi \in (0,1)$, of the information posted by a user will cause privacy harm on others. Users may not know which information belongs to this ϕ fraction. All other settings and notations remain the same as in the main model. Then user i's utility from participating in the community conditional on participation size, s, is

$$u_{i|s}^{in} = n(vx_{ij} - \frac{cx_{ij}}{2}) + (1-n)(vx_{ik} - \frac{cx_{ik}^2}{2\delta}) + eQ_{-i} + \omega X_{\cdot i} - \theta_i Y_{\cdot i}. \tag{A.28}$$

(A.28) is the counterpart of equation (2) in the main model. Solving the FOCs yields $x_{ij}^{sq} = \frac{v}{c}$ and $x_{ik}^{sq} = \frac{\delta v}{c}$. Using these equilibrium quantities, we can derive user i's net utility from participating in the community is

$$u_{i|s}^{in-out} = \frac{\gamma v}{2c}[v - 2(1-\epsilon)s\lambda_i],$$

where $\lambda_i = \phi\theta_i - e - (1-\phi)\omega$ has the same interpretation as in the main model. Then all the results in the main model can be easily replicated here. As we can see, the only difference is that $\frac{\psi}{1+\psi}$ in the main model is replaced by ϕ here. That is because we conveniently capture the differences in the quantities of non-sensitive and sensitive information posted by a user through the cost ratio parameter, ψ, which is also exogenously given. As a result, an exogenous and fixed fraction of the total amount of information posted by a user is privacy-infringing. Therefore, $\frac{\psi}{1+\psi}$ is equivalent to ϕ.