An Informational Theory of Privacy*

Ole Jann
Nuffield College, University of Oxford

Christoph Schottmüller
University of Copenhagen and TILEC

July 17, 2017

Abstract

We develop a theory that explains how privacy can increase welfare. Removing privacy makes behavior observable but does not necessarily lead to an informational gain because individuals change their behavior when they are being observed. Privacy can also prevent rational but inefficient statistical discrimination. Our model allows us to weigh gains and losses from privacy and derive sufficient conditions for when it is efficient or even Pareto-optimal. We show how our theory can be applied to decide who should have which information, and we discuss applications to online privacy, credit decisions, and transparency in government.

JEL: D72, D82, J71, K40
Keywords: privacy, asymmetric information, information aggregation, statistical discrimination, law and economics

Privacy is one of the most pressing issues of the information age. Recent years have brought the revelation that many western governments routinely engage in comprehensive electronic surveillance of their own citizens (Greenwald, 2014). At the same time, some of the world's most valuable companies are built on the idea that people *voluntarily* give up personal information in exchange for free services or better prices. The collection and use of ”big data” to predict everything from consumer behavior to credit risk and life expectancy are widely talked about among experts as well as the general public.

Behind these developments lies the idea that removing (or voluntarily giving up) privacy will ultimately lead to gains in efficiency, as terrorists can be found and consumers

*Jann: Nuffield College and Department of Economics, University of Oxford; ole.jann@economics.ox.ac.uk. Schottmüller: Department of Economics, University of Copenhagen; christoph.schottmueller@econ.ku.dk. We are grateful for helpful comments by Sebastian Barfort, James Best, Ramon Caminal, Vincent Crawford, Drew Fudenberg, Paul Klemperer, Nenad Kos, Meg Meyer, Marco Ottaviani, Jens Prüfer, David Ronayne, Tomas Sjöström, Peter Norman Sørensen and Peyton Young, as well as audiences at the Universities of Copenhagen, Köln, Oxford and Tilburg, at NHH Bergen and the 2017 Barcelona GSE Summer Forum.

get more accurate prices. But how can we decide if and when that is the case? Consider, for example, the case for government surveillance. Proponents usually argue that the gains to public safety outweigh the inconveniences of privacy intrusions and unnecessary searches. Opponents respond that these costs outweigh the gains. Does their disagreement simply come down to differences in normative priorities that are outside the reach of positive theory?

In this paper, we develop a theory that allows us to analyze the welfare effects of strengthening or weakening privacy, without having to rely on any particular weighing of the competing objectives. Along the way, we show how privacy can lead to welfare- (and sometimes Pareto-) improvements, and when that is the case.

The general idea is to compare what is actually gained and what is lost when we remove an information asymmetry – i.e. when we remove someone's privacy.[1] If a previously hidden action by a person becomes observable, the outside world can learn something about that person. That is a gain to any observer, but could be a loss to the one who is being observed. To avoid this loss, the observed might change her behavior even if this change comes at a cost. This change in behavior also reduces the observer's information gain. We identify several forces that affect the magnitude of these different welfare effects. This allows us to say when privacy is (not) welfare optimal. Several of our results do not involve utility comparison between observed and observer as they identify scenarios where privacy is a Pareto-improvement.

We develop a general model that incorporates the different welfare effects; we also consider several extensions (sections 4 and 5) and we discuss the qualifications of (and exceptions to) our results (section 6). We discuss three applications of our model (beyond the examples that we use to illustrate our results) in section 7. In the remainder of this introduction, we go through an informal example that explains all of our results, and briefly comment on the generality of the results and the connection to other research.

An Example that Contains All of Our Results

Consider the following problem involving Alice and an employer. Alice would prefer if cannabis was legalized, and she wants to publish an overview of her arguments on an online social network to try to convince her friends. However, we assume that in Alice's world there is very little privacy: If she does something online, everyone can see it – not just her friends, but also potential employers, her parents, the police, and so on.

There is some statistical dependence between preferences on legalization and actual drug use: people who use drugs are more likely to support legalization. The correlation

[1] Throughout the paper, we think of privacy as an information asymmetry: The ability to take actions without being observed, and having interactions with others confined to the intended recipients. Classical economic theory has followed this same path to suggest that privacy is usually welfare-reducing – see, for example, Posner (1981). Of course, this is only one of many possible definitions and understandings of the term "privacy"; cf. Solove (2010) for an overview.

is of course far from perfect – many people might support legalization for philosophical or practical reasons without being drug users, and some users might even oppose it.

Employers do not want to hire drug users, but drug use is not observable. An employer will therefore use the observable characteristic (whether Alice did or did not publicly support legalization) to make a hiring decision: People who have supported legalization will not be hired. We can show that this happens in equilibrium if the correlation between types (i.e. drug use and preference for legalization) is sufficiently high. Being unable to observe the attribute that he is really interested in, the employer will statistically discriminate (as described by Arrow, 1973 and Phelps, 1972) based on observed preference.

Then, however, Alice has to make a choice: Stay quiet and get hired – or voice her preference, and go without the job. If she doesn't feel strongly about the subject (i.e. if she only has a weak preference for legalization), she will choose not to express her opinion. Lack of privacy therefore causes a "chilling effect". Despite Alice's preference being not only legal and legitimate, but also insubstantial for the job (recall that even the employer does not take issue with her preference for legalization itself), she decides not to express it for fear of the consequence.[2]

We can immediately see that Alice loses either way from not having privacy: Either she is forced to suppress her opinion, or she doesn't get the job. Furthermore, society as a whole loses, since the spectrum of opinions that are present in public debate is skewed: There is no reason for those who oppose legalization to hold back with their views. Since the optimal policy should be an unbiased aggregation of individual preferences, the policy that is implemented will systematically deviate from this optimum.

Yet where Alice loses and the information aggregation in society suffers, the employer gains: He can now distinguish between applicants whom he more or less likes to employ. But *how much* does he gain? Less than we might expect, because of the chilling effect: Since many people (drug-users as well as non-users) now misrepresent their preferences, observing what someone says about drug legalization becomes less informative about actual drug use.

To make statements about welfare, we must weigh Alice's loss against the employer's informational gain. One might think that such welfare considerations must depend on which weight we give to each of them when we aggregate welfare. But we can in fact derive three sufficient conditions under which privacy is welfare-optimal for any possible way of aggregating welfare. Under the first two of these sufficient conditions, privacy is in fact ex-ante Pareto-efficient (and Pareto-superior to the case without privacy).

First, consider population size. We do not assume Alice's motivation as reduced-form, but derive it from the influence she has with her actions. If Alice is part of a large community, her influence on whether drugs are legalized is small. The cost of speaking

[2]The term "chilling effect" has been used by legal scholars at least since 1952, when U.S. Supreme Court Justice Felix Frankfurter used it in a concurring opinion in Wieman v. Updegraff, 344 U.S. 183.

her mind, however, is independent of this. The chilling effect is therefore larger in large groups, where the cost of expressing one's preference easily outweighs an individual's influence. While it is still rational for the employer to base his decision on people's published opinions if they are available to him, they become less and less informative, up to a point where he gains no information at all.

Next, consider the costs of not being hired. If it is extremely costly to Alice to be thought of as a drug user, she would be willing to misrepresent her opinion almost regardless of how strongly she feels about it. The employer would then gain very little information from observing her choice. While it is still rational for him to make use of any information he can get, this distortionary equilibrium effect destroys any informational gain he could get. Hence welfare can be improved by privacy.[3]

Finally, even if neither of these sufficient conditions is fulfilled, any information the employer gains is about people's preference on drug legalization, and not on drug use directly (which is what he cares about). His informational gain thus depends on how statistically dependent drug use is on policy preference. Given that the chilling effect always reduces the informativeness of what the employer learns, and that the dependence between preference and drug use provides an upper limit on the information that the employer can gain, we can show that for any given parameter set, privacy is welfare-optimal unless the dependence between preference and drug use is too high.

We show in an extension that our results are robust to endogenizing the way in which information is aggregated, and to using different information aggregation mechanisms for the two cases of privacy and no privacy. We also consider alternative approaches to the information aggregation problem in society: What if Alice directly derives gain from speaking her mind, so that there is no information aggregation problem? What if the information aggregation problem is even more acute than in our lead example, so that there is an optimal policy that is the same for all individuals, and information aggregation is about trying to find that policy? We examine which of our results apply under these conditions in section 4, and we show that some of our results become even stronger with a more acute information aggregation problem.

In another extension, we ask: Can the optimal level of privacy be achieved if Alice can simply choose to keep her message private? That is not the case, since the act of choosing privacy becomes informative in itself. People who oppose drug legalization have no reason to hide this fact and might in fact want to broadcast it to potential employers, which means that there is no stable equilibrium in which everybody chooses privacy. To work well, privacy can therefore not always be left to the individual – sometimes it needs to be mandated.[4] We also consider when the introduction of a price for privacy can

[3]Note that there is also a first-order welfare effect of higher costs to Alice, but our general welfare result follows from the second-order effect on the employer's information gain.

[4]There are parallels to the obligatory secret ballot. Consider for example the point made by Schelling (1960): If ballot secrecy was optional, voters could be intimidated into making their ballot public. For-

guarantee optimal allocations, and show that taxes on information gathering can lead to Pareto-improvements (and generate revenue).

In a final extension, we look more closely at Alice's rational response to having no privacy. Imagine, for example, that Alice could bring a lawyer to every job interview – which slightly decreases, ceteris paribus, the probability of not being hired because of her expressed opinions, while making the interaction much more cumbersome for the employer and decreasing Alice's payoff if she gets hired. In this way, the statistical discrimination that results from lack of privacy can erode trust among the players, and can mean that by a chain of rational responses to each other's behavior, they end up in a Pareto-inferior equilibrium.

Our general model, which we introduce in section 1, considers a problem of information aggregation, in which a group of individuals have cardinal preferences over two options and express their preference by supporting one of the two options. While we restrict our main analysis to a specific mechanism, all of our main results hold for all mechanisms that fulfil a set of conditions. The example in this introduction already points to political information aggregation through public debate or voting as a possible application. However, the mechanism might just as well be a market in which two providers of goods or services compete for customers. Efficiency demands that the provider who is preferred by most customers also does more business. But if using one of the providers is in some way disreputable, lack of privacy and the chilling effect will systematically bias the result. We discuss examples of the mechanism in section 7.

What kind of privacy problem do we have in mind when we assume, as in our example, that some observable behavior is predictive of an unobservable type? Here, too, we keep our assumptions quite general, as we only assume that one unobservable type (in our example: drug use) is positively regression dependent (c.f. Lehmann, 1966) on another (policy preference). It is crucial to note that this does not require any sort of causal relationship – only dependency. We think that in the real world, almost any variable can be "predictive", in the sense of our model, of almost any other variable. Meehl (1990) calls this the "crud factor" and notes that "in social science, everything is somewhat correlated with everything." Even traits that seem unrelated are often dependent if we do not control for other variables – this is the idea of many businesses' use of "big data". For example, preferences for certain beverages or types of cars can be predictive enough of political leanings so that political parties exploit them (Hamburger and Wallsten, 2005). A large consulting firm advertises that it can reliably predict people's life expectancy from observing their buying decisions (Robinson et al., 2014, p. 6).

We would also like to point out that the assumptions that we make about statistically dependent observables and unobservables apply in almost all contexts, economic or otherwise. It is quite rare that banks, employers, law enforcement agencies or indeed anyone

bidding them to do so protects them from any such intimidation.

else can directly and unambiguously observe the variable that they are really interested in. One's future financial situation or intellectual ability, whether one is a terrorist, a criminal or a reliable friend are all essentially unobservable. Through years of everyday experience, we have gotten used to forming estimates through statistical discrimination by using (multiple) observables. But all observation is ultimately incomplete, and the correlation between what we conclude based on our observations and the truth is never 100 percent. A police officer could be trying to judge whether someone carries a gun based on what he sees in the suspect's hand, or whether someone is planning a terrorist attack based on their internet search history: the difference in correlation between the two situations is quantitative, not qualitative.

Relation to Other Research

In understanding privacy as the creation and maintenance of asymmetric information, our study takes a similar point of departure as the "Chicago school", exemplified by Stigler (1980) and Posner (1981). However, they go on to argue that since asymmetric information creates economic inefficiencies and reduces welfare, privacy must be welfare-reducing. This line of thought echoes the ubiquitous "nothing to hide"-argument, which Schneier (2006) has called "the most common retort against privacy advocates." According to Solove (2010), this argument usually takes the form: "If you aren't doing anything wrong, what do you have to hide? ... If you have nothing to hide, what do you have to fear?"

Our model allows us to argue that this argument, and hence the claim that privacy necessarily reduces welfare, is based on three faulty assumptions. Firstly, it assumes that all information is precise and unambiguous. But decisions that are made under uncertainty are routinely based on statistical discrimination. Secondly, it ignores the effect of rational behavior (the chilling effect) on the informativeness of observations. Thirdly, it ignores the secondary impact of the chilling effect on the informativeness of aggregate variables.

It is plausible that a first-best could be achieved in the total absence of any asymmetric information. But in the real world, asymmetric information is a fact of life, and questions of privacy are therefore about *how much* asymmetric information there should be, and how it should be structured. The Chicago argument and the "nothing to hide" argument therefore address an imaginary ideal case and have little to say about intermediate cases (and whether, for example, welfare is monotone in the amount of asymmetric information).[5]

Accepting our argument that privacy can be welfare-enhancing, and that sometimes privacy even needs to be mandated to work, also means refuting the argument that any regulation of privacy can at best be ineffective and at worst damaging.

Two recent papers have proposed rationales for privacy in public good settings where agents have an intrinsic motivation to contribute and also care about their image. That is,

[5] A similar argument against the Chicago school is made by Hermalin and Katz (2006).

each agent would like others to believe that he has a high intrinsic motivation. Daughety and Reinganum (2010) show that privacy can be optimal in this setting if a lack of privacy would lead to excessive contributions due to image concerns. Ali and Bénabou (2017) add a principal who has to decide on his own contribution in a setting where agents and principal have only noisy information about the usefulness of the public good. More privacy implies that the aggregate contribution by the agents is – as a signal of the usefulness of the public good – more informative and therefore allows the principal to better choose his contribution. The mechanism in our model differs in two important ways: First, we do not rely on image concerns but microfound the downside of taking a certain action (e.g. supporting drug legalization) through an interaction with another player (e.g. a future employer). Note that image concerns are not a reduced form for this because the (changing) utility of the interacting player is an integral and indispensable part of our welfare analysis.[6] Second, the inference is somewhat more subtle in our model as the interacting player is not interested in the preference for action (e.g. the preference for drug legalization) but only in unobservables that are correlated with this preference (e.g. drug use). In this sense, we link the literature on statistical discrimination (Arrow, 1973; Phelps, 1972) and the literature on privacy.

Apart from such general economic studies of privacy, there is a large literature in industrial organization and related fields that deals with demand for privacy and the meaning of privacy for issues like pricing. Acquisti (2010) and Acquisti et al. (2015) provide excellent overviews; here we want to point to some studies that are loosely related to ours.

Hirshleifer (1971) argues that information revelation before trading can impair risk-sharing and therefore reduce welfare. This "Hirshleifer effect" means, for example, that providing health data about buyers of life insurance transfers risk from the seller to the more risk-averse buyers. The Hirshleifer effect deals with the presence or absence of information about the payoff relevant state of the world (or "type") and not with inference from behavior, statistical discrimination and chilling effects. Hermalin and Katz (2006) follow in a similar vein – considering the presence/absence of information about type – and show that privacy can be efficient in a model of price discrimination by a monopolist and a model of a competitive labor market. They also show that allocating property rights to control information does not affect equilibrium outcomes (and therefore the results) in their setup. Prat (2005) shows in a principal-agent model of career concerns that the principal benefits from not knowing the agent's action. The reason is that a less informed agent type might otherwise ignore his (somewhat informative) signal and simply take the action most likely taken by well-informed types in order to improve his

[6]Morris (2001), in a model of "political correctness," derives image concerns from model primitives in a way that has similarities to our model, but does not carry out a welfare analysis beyond identifying several countervailing effects.

reputation for being informed. Prat's simple model is well suited to make this point but less well suited to analyze the welfare consequences of privacy because the agent's expected utility (his expected reputation) does neither depend on information structure nor on equilibrium type and consequently welfare is simply equivalent to the amount of information the principal receives in equilibrium. Calzolari and Pavan (2006) consider information exchange between principals who contract with the same agent, and find that the principal moving first commits to not selling information to the second principal under certain conditions. We do not consider a setting where one of the players sets the information structure but view the presence/absence of privacy as a given regime – possibly set by an (unmodeled) legislator. Our welfare analysis corresponds to the decision problem of a welfare maximizing legislator. Cummings et al. (2015) analyze a model in which a consumer reveals information to an advertiser by his buying decision; they argue that – due to strategic responses – the *equilibrium effects* of privacy are different from what one might naively expect – this is similar to our description of the chilling effect.

Similar to the fifth extension of our model, Acquisti and Varian (2005) consider rational reactions by people who lack privacy – for example, that internet users employ anonymization tools. They argue that this can make it unprofitable (and hence inefficient) for the seller of goods to collect information.

1. Model

The model has two stages. First, an information aggregation stage in which each of n individuals has to decide between two options. Second an interaction stage in which each individual interacts with an opposing player (OP).

In the information aggregation stage, each of n individuals has to choose to support one of two options, which are either $p = 1$ or $p = 0$.[7] Individual i's choice is denoted by $p_i \in \{0, 1\}$. If m individuals choose $p_i = 1$, the probability that $p = 1$ is $q(m/n) = m/n$. That is, the decision rule is (i) monotone: more people supporting an option leads to a higher likelihood that the option is adopted, (ii) unbiased: the decision is not biased in favor of one option and (iii) anonymous: the influence of each individual is the same and does not depend on the choices of other individuals. (We will endogenize the decision rule q in an extension, see section 5.1, and we generalize our results to a class of aggregation mechanisms in 5.2)

The payoff of option $p \in \{0, 1\}$ for individual i is $\theta_i p$. That is, θ_i can be interpreted as the difference of i's valuations for $p = 1$ and $p = 0$. We assume that the θ_is are iid draws from a standard normal distribution Φ and that θ_i is private information of individual i.

Before describing the interaction stage let us connect the information aggregation stage to our example from the introduction.

[7] In the supplementary material to the paper, we show that our results are robust to giving individuals the possibility to "abstain".

Example 1. *There is a petition to liberalize drug laws. The more citizens sign the petition, the more likely it is that its demands will be implemented. Every citizen has to decide whether to sign the petition ($p_i = 1$) or not ($p_i = 0$). Every citizen has an expected payoff consequence of liberalization of θ_i.*

We now turn to the interaction stage. In this stage, each individual interacts with one opposing player (OP). We will describe this player as one central outside player, although nothing in the model rules out the alternative case where each individual interacts with a different player (possibly even one of the other individuals). OP has to choose how he interacts with individual i and he can choose from the actions A (aggressive) or M (mild). We normalize OP's payoff from playing M to 0 and assume that the payoff of playing A against a type τ_i is simply τ_i which is a private characteristic of individual i. The characteristics τ_i are drawn independently from a distribution Γ_{θ_i} with support in $[\underline{\tau}, \bar{\tau}]$. We assume that $\Gamma_{\theta_i'}$ first order stochastically dominates $\Gamma_{\theta_i''}$ if and only if $\theta_i' \geq \theta_i''$.[8] This implies that θ_i and τ_i are positively correlated as higher θ_i make higher τ_i more likely (and this positive correlation prevails if we only consider individuals with θ_i above a certain threshold). We also assume that $\Gamma_\infty = \lim_{\theta_i \to \infty} \Gamma_{\theta_i}$ is a non-degenerate distribution in the sense that $\Gamma_\infty(\tau_i) > 0$ for all $\tau_i > \underline{\tau}$ – a technical property that will be useful for some of our welfare results.

To make the problem interesting, we assume that A is OP's best response if $\tau_i = \bar{\tau}$ and M is the best response if $\tau_i = \underline{\tau}$. That is, $\bar{\tau} > 0$ and $\underline{\tau} < 0$. Furthermore, we assume that $\mathbb{E}[\tau_i] \leq 0$, so that M is an optimal response to any individual about whom nothing is known. OP does not observe τ_i when choosing his action and will only be able to form expectations about the individual's τ_i. We will distinguish two cases: In the *privacy* case, we consider OP's problem when he has no information on τ_i apart from the priors Γ_{θ_i} and Φ; in particular OP does not know p_i in this case. Our above assumption that $\mathbb{E}[\tau_i] \leq 0$ means that in this case, the OP's best response is to play M against all individuals since the expected payoff of playing A against any individual is simply $\mathbb{E}[\tau_i]$.

Most of the analysis, however, will deal with the case *without privacy* in which OP observes which opinion i voiced in the information aggregation stage, i.e. OP can observe p_i and can condition his expectation of τ_i on this information. The individual's payoff is normalized to 0 when OP plays M. If OP plays A against i, then i will have a payoff of $-\delta(\tau_i)$ where $\delta > 0$ is a differentiable function that is weakly increasing in τ_i. We assume that individual payoffs from the two stages are additive. All players are assumed to maximize their expected payoff.[9]

Figure 1 shows a graph of the model which illustrates the two types that each individual has, and how they are correlated. We will use and modify this figure in the following

[8]In the statistical literature, this property is called positive regression dependence (Lehmann, 1966).

[9]For the main part of the paper, we assume that q is the same across the two cases– privacy and no privacy. Section 5.1 endogenizes q and shows that our results are robust to relaxing this assumption.

sections to illustrate our main points. Table 1 shows the payoffs to the OP and an individual from a single interaction to illustrate the trade-off that OP faces between treating players who chose $p_i = 1$ aggressively or mildly.

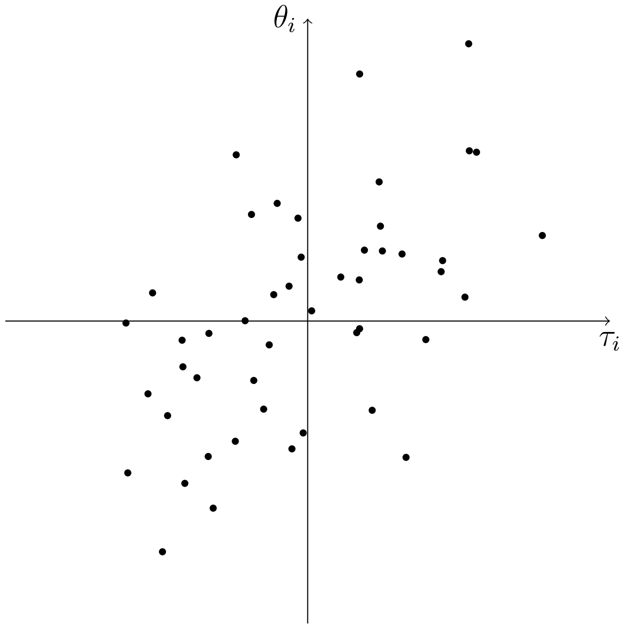

Figure 1: An illustration of our model. Each dot represents an individual. Individual i's type τ_i and θ_i are positively correlated. The OP wants to treat those with $\tau_i > 0$ aggressively and all others mildly, but he cannot observe τ_i. Individuals' preferences are given by θ_i; observing the choices of individuals will therefore provide the OP with information about τ_i. "Privacy" is the question whether the OP can or cannot observe an individual's choice before deciding how to treat her.

		Individual chooses	
		$p_i = 1$	$p_i = 0$
OP treats $p_i = 1$ ...	... aggressively (A)	$\tau_i, p\theta_i - \delta(\tau_i)$	$0, p\theta_i$
	... mildly (M)	$0, p\theta_i$	$0, p\theta_i$

Table 1: The payoffs of OP and a single individual under the assumption that OP treats all individuals who choose $p_i = 0$ mildly. OP interacts with all individuals. His payoff is the sum of the payoffs from all these interactions, but his choice of how to treat individuals applies to all interactions. Note that the payoff of the individual depends on p, which he influences with his choice of p_i.

Example 1 (Continued). *Continuing our example, OP might be a potential employer who has to decide whether to hire citizen i (action M) or not to hire i (action A). The employer would prefer to hire i if i is not a drug user and would prefer not to hire i if i is a drug user. The type τ_i would then be binary, i.e. $\tau_i \in \{\underline{\tau}, \bar{\tau}\}$, and would indicate whether i is a drug user or not. The first order stochastic dominance assumption on Γ_{θ_i} then simply means that the probability of being a drug user is increasing in θ_i. Hence, τ_i and θ_i are positively correlated which also means that citizens who support drug legalization are* *relatively more likely to be drug users than citizens opposing legalization. Citizen i prefers to be hired and the disutility of not being hired – denoted by δ – might be bigger for drug users because their outside options are generally worse.*

2. Preliminary Analysis – The Chilling Effect

2.1. OP's Beliefs

We start the analysis with some preliminary results on the individuals' and OP's beliefs and strategies. This will then allow us to establish the chilling effect and analyze its welfare implications.

The payoff of individual i from the information aggregation stage is $p\theta_i$. The higher θ_i, the higher is i's benefit from $p = 1$. Given this structure, it is not surprising that i will use a cutoff strategy: If θ_i is higher than some cutoff/threshold $t(\tau_i)$, i chooses $p_i = 1$ and otherwise he chooses $p_i = 0$. In the privacy case, payoffs of the interaction stage do not depend on actions chosen in the information aggregation stage and therefore i will choose $p_i = 1$ if and only if θ_i is positive. This pins down the equilibrium of the privacy case as we already established that OP plays M there by $\mathbb{E}[\tau_i] \leq 0$.

Lemma 1. *Only cutoff strategies are rationalizable for individuals, i.e. each individual will choose a cutoff $t(\tau_i)$ and play $p_i = 0$ if $\theta_i < t(\tau_i)$ and $p_i = 1$ if $\theta_i > t(\tau_i)$. In the privacy case, the optimal cutoff equals zero: $t^p(\tau_i) = 0$.*

Given a cutoff strategy $t(\tau_i)$, we can determine the beliefs of OP in the case without privacy using Bayes' rule as

$$\beta_1(\tau) \equiv prob(\tau_i \leq \tau | p_i = 1) = \frac{\int_{\mathbb{R}} \int_{\underline{\tau}}^{\tau} \mathbb{1}_{t(\tau_i) \leq \theta_i} \, d\Gamma_{\theta_i}(\tau_i) \, d\Phi(\theta_i)}{\int_{\mathbb{R}} \int_{\underline{\tau}}^{\bar{\tau}} \mathbb{1}_{t(\tau_i) \leq \theta_i} \, d\Gamma_{\theta_i}(\tau_i) \, d\Phi(\theta_i)} \tag{1}$$

$$\beta_0(\tau) \equiv prob(\tau_i \leq \tau | p_i = 0) = \frac{\int_{\mathbb{R}} \int_{\underline{\tau}}^{\tau} \mathbb{1}_{t(\tau_i) \geq \theta_i} \, d\Gamma_{\theta_i}(\tau_i) \, d\Phi(\theta_i)}{\int_{\mathbb{R}} \int_{\underline{\tau}}^{\bar{\tau}} \mathbb{1}_{t(\tau_i) \geq \theta_i} \, d\Gamma_{\theta_i}(\tau_i) \, d\Phi(\theta_i)}. \tag{2}$$

That is, $\beta_1(\tau)$ is the probability that τ_i is below τ given that i chose $p_i = 1$. These beliefs allow us to define OP's expected utility of playing A conditional on observing decision p_i and given cutoff strategy $t(\tau_i)$:

$$v_1 = \int_{\mathbb{R}} \tau \, d\beta_1(\tau) \tag{3}$$

$$v_0 = \int_{\mathbb{R}} \tau \, d\beta_0(\tau). \tag{4}$$

The best response of OP to a given cutoff strategy is to choose A against an individual who chose $p_i = j$ if $v_j > 0$ for $j \in \{0, 1\}$. Otherwise, it is a best response to choose M.[10]

[10]Note that OP's best response does not depend on the number of individuals choosing $p_i = 1$ in

2.2. The Chilling Effect

For the case without privacy, the following lemma states that OP is more likely to play A against individuals who have chosen $p_i = 1$ in the information aggregation stage than against those who have chosen $p_i = 0$. Intuitively, individuals with a high θ_i have more to gain from choosing $p_i = 1$ in the information aggregation stage. As θ_i and τ_i are positively correlated, OP is relatively more likely to play A against them.

Lemma 2. *In every perfect Bayesian equilibrium, $v_1 \geq v_0$.*

The previous lemma is the basis of the chilling effect. In equilibrium, OP is more likely to play A against individual i if i chose $p_i = 1$ in the information aggregation stage. For this reason, i is to some degree afraid of choosing $p_i = 1$. More technically, there are types (θ_i, τ_i) for which an individual would choose $p_i = 1$ in the privacy case but would choose $p_i = 0$ if OP learns p_i before taking his action. The decision in the information aggregation stage is therefore biased against $p = 1$ in the case without privacy. This effect is particularly pronounced if individuals have much to loose from being treated aggressively (high δ) or if n is high. In the latter case, individual i's impact on the choice of option p – and therefore his motivation for supporting his preferred option – is lower.

There is one minor caveat to this result: If OP's preferences are such that he always uses the same action, e.g. OP prefers to play M against individuals who have played $p_i = 0$ and individuals who have played $p_i = 1$, then no chilling occurs because information on p_i is not relevant for OP's decision and the equilibria with and without privacy are identical. Put differently, chilling occurs whenever information about p_i matters for OP's behavior. We denote by Δ the difference in the probability that OP plays A against individuals choosing $p_i = 1$ and $p_i = 0$ (in equilibrium). By lemma 2, $\Delta \geq 0$.

Proposition 1 (Chilling effect)**.** *Without privacy the optimal cutoff is*

$$t^{np}(\tau_i) = n\delta(\tau_i)\Delta. \tag{5}$$

The equilibrium cutoff for every type τ_i is weakly higher without privacy than in the privacy case: $t^{np} \geq 0$. The inequality is strict whenever the absence of privacy changes the equilibrium behavior of OP: $t^{np} > 0$ if $\Delta > 0$. The cutoff is increasing in τ_i.

Figure 2 illustrates the chilling effect. Individuals with a very high preference for $p = 1$ will choose $p_i = 1$ with and without privacy and individuals with a very low (that is, negative) preference will choose $p_i = 0$ in both cases. Those that are almost indifferent but choose $p_i = 1$ in the privacy case are the ones who change their behavior when OP uses information about p_i. In this sense, the individuals who change their

the first stage. Intuitively, this information does not contain any information about τ_i (given that p_i is known) because all θ_i and τ_i are independently drawn by assumption.

behavior do not lose a lot by their behavior change. However, individuals with strong preferences for $p = 1$ should be most worried about chilling: They do not change their own behavior but – because chilling changes the behavior of those with more moderate preferences – $p = 1$ will be less likely without privacy than it would have been with privacy. Furthermore, individuals with strong preferences suffer from being treated aggressively without privacy. In short, privacy changes the behavior of moderate people and protects people with extreme preferences.

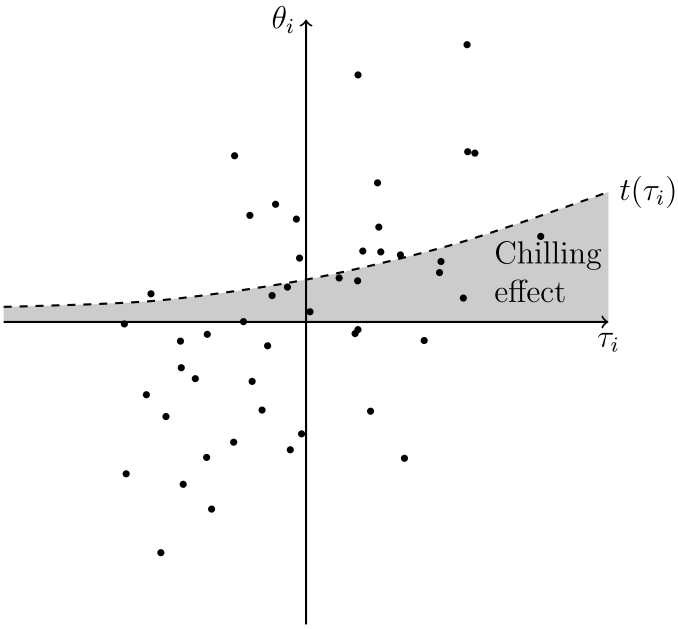

Figure 2: An illustration of proposition 1. If decisions are private, all individuals with positive θ_i will support $p = 1$ and all others support $p = 0$. If decisions are public, OP can use the individuals' decisions to predict their type τ_i. Therefore, some people with relatively low θ_i will misrepresent their preferences to avoid the statistical discrimination. Since the disutility from being treated aggressively rises in τ_i, we get the curve above. Individuals in the gray area are subject to the chilling effect and support $p = 0$ without privacy.

The cutoffs of higher τ_i are weakly higher. As a consequence, abolishing privacy becomes somewhat less profitable for OP compared to the case where individuals use the same cutoff: Increasing the cutoff reduces the statistical dependence between p_i and τ_i even if δ is only weakly increasing (and more so if it is strictly increasing). Hence, OP's benefits from statistical discrimination are reduced by the chilling effect. This means that an evaluation of whether privacy should be given up will be biased against privacy if it does not consider the behavior change of individuals, which also causes a reduction in the information gain for OP.

The following proposition makes this statement more formally. To do so, we have to add the technical condition that the distribution Γ_0 is symmetric around 0.[11] This ensures

[11]An alternative technical condition that is also sufficient for the result to hold is $\mathbb{E}[\tau_i|\theta_i = 0] \geq 0$.

that OP does not gain from the fact that all cutoffs increase (while the argument above shows that it is detrimental to OP that cutoffs of higher τ increase by a larger amount). Note that the following proposition does not compare OP's payoffs under privacy and no privacy. In line with the argument above, it compares OP's payoffs without privacy with his payoffs in a hypothetical situation where there is no privacy but individuals use their equilibrium strategies of the privacy case.

Proposition 2 (Reduced information gain)**.** *Assume that the distribution $\Gamma_0(\tau)$ is symmetric around $\tau = 0$. OP's payoff without privacy is lower if individuals use the cutoffs $t^{np}(\tau)$ than if they used the cutoffs $t^p(\tau) = 0$.*

As a side remark, note that the technical condition in proposition 2 is also sufficient to rule out somewhat uninteresting equilibria without privacy in which the OP plays M against everyone (i.e. equilibria in which OP does not use the information he has): Given that Γ_0 is symmetric around zero, OP would be indifferent between A and M if he knew that $\theta_i = 0$. By first order stochastic dominance, he will then prefer A to M when he knows that $\theta_i \geq 0$. But this is exactly the information $p_i = 1$ would give him because the cutoff in such a hypothetical equilibrium would clearly be the same as in the privacy case, namely zero. Hence, OP plays A against those choosing $p_i = 1$.

3. Welfare Analysis

What are the welfare consequences of the chilling effect? It is not hard to see that the chilling effect causes a welfare loss in the information aggregation stage. The bias against $p = 1$ means that information is no longer efficiently aggregated and decision 0 is more likely to be taken than optimal. The following lemma states formally that the privacy equilibrium yields a higher expected consumer surplus in the information aggregation stage than the equilibrium without privacy. (We define consumer surplus in the information aggregation stage as $p \sum_{i=1}^{n} \theta_i$.)

Lemma 3. *The cutoff strategy $t^p(\tau) = 0$, i.e. the equilibrium strategy in the privacy case, gives a higher expected consumer surplus in the information aggregation stage than any $t^{np}(\tau) > 0$.*

While the lemma shows that individuals are always better off under privacy, this does not allow us to say anything about overall welfare yet. Without privacy, OP can adjust his behavior according to people's choices p_i and thereby make use of the correlation between θ_i and τ_i to identify individuals with a relatively high τ_i. Hence, OP might be better off without privacy and his utility has to be part of a welfare analysis.

Our welfare analysis consists of two parts. First, we derive sufficient conditions for welfare optimality of privacy in a Pareto sense. Second, we study welfare in a utilitarian framework and consider how the information structure, in particular the correlation between θ_i and τ_i, affects the welfare comparison between privacy and no privacy.

3.1. Pareto Results

We will establish three different sufficient conditions for when privacy is welfare-optimal in an ex ante Pareto sense. First, if OP plays a mixed strategy in equilibrium (i.e. he mixes between treating people who choose $p_i = 1$ mildly or aggressively), privacy always provides higher welfare than no privacy. This simply follows from the fact that while individuals always lose from lack of privacy, OP is indifferent between privacy and no privacy if he plays a mixed strategy in the no privacy case.

Second, we show that for large n, i.e. if there are many individuals, there exist no equilibria in which OP plays a pure strategy, and privacy is therefore optimal.

Third, we show the same for large $\delta(\tau)$: If the cost of being treated aggressively is very high, then privacy is also optimal.

The second and the third part of the following proposition require the additional assumption that δ is strictly (and not just weakly) increasing in τ. This guarantees that we can make statements about how the correlation between p_i and τ_i develops in the limit. (We can derive the same results without this assumption if we consider welfare as being any convex combination of the welfare functions of individuals and OP – see proposition 4 below.)

Proposition 3 (Welfare comparison)**.** *1.) If OP uses a mixed strategy in the equilibrium without privacy, then privacy maximizes welfare.*
Assume $\delta' > 0$:
2.) For n sufficiently large, privacy welfare dominates no privacy in the following sense: Compared to the no privacy case, privacy leads to a higher expected consumer surplus and the same expected payoff for OP.
3.) Let the disutility of an individual facing action A by OP be $r\delta(\tau)$ *(instead of* $\delta(\tau)$*). For r sufficiently large, privacy welfare dominates no privacy.*

The intuition behind the second result is that the chilling effect is getting very large if the number of individuals grows. To be more specific, suppose for a moment that there is a pure strategy equilibrium with $\Delta = 1$. In this case, t^{np} becomes very high and very steep if n is large. This steepness reduces the correlation between p_i and τ_i because in particular individuals with a high τ_i are chilled (and choose $p_i = 0$). For sufficiently high n the effect is so strong that OP does not find it optimal to play A against those choosing $p_i = 1$. Consequently, no pure strategy equilibrium exists for large n. OP will, therefore, use a mixed strategy in equilibrium, which makes playing $p_i = 1$ less painful and therefore preserves some informativeness in the individuals' decisions. Hence, OP will be indifferent between his two actions, i.e. he would be equally well off by choosing M against everyone which would give him a payoff equal to his equilibrium payoff with privacy. This implies that OP is indifferent between privacy and no privacy. Since individuals are clearly worse off without privacy because of the biased information aggregation and the possibly increased probability of being treated aggressively in the interaction stage, the privacy case is welfare dominant.

The intuition for the third result is similar: If δ is high, an individual's benefit from the information aggregation stage is relatively small compared to the individual's potential losses in the interaction stage. Individuals will therefore be chilled a lot if OP plays A against individuals who chose $p_i = 1$. Playing A for sure against those who chose $p_i = 1$ is then no longer a best response. Consequently, OP uses a mixed strategy for r sufficiently high and privacy is welfare optimal.

Note that all the welfare results in proposition 3 are Pareto results *from an ex ante point of view*. That is, privacy makes individuals strictly better off in expectation (i.e. before knowing their type) while OP is indifferent. In the following, we will consider utilitarian welfare instead; in section 4.2 we describe conditions under which privacy can be *ex post* Pareto-optimal.

3.2. Results with Utilitarian Welfare

We now define welfare as the sum of individuals' and OP's payoff.[12] If, for example, OP uses a pure strategy without privacy ($\Delta = 1$) welfare will be

$$\sum_i p\theta_i + \mathbb{1}_{\theta_i \geq t^{np}(\tau_i)}(\tau_i - \delta(\tau_i)).$$

This allows us to derive results (2) and (3) of proposition 3 without imposing the additional assumption $\delta' > 0$. The intuition is that t^{np} will be arbitrarily high as n (or r) grows large. Consequently, the probability that OP benefits from treating an individual aggressively is low because the probability of a citizen having θ_i above the threshold converges to zero as n (or r) grows large. Privacy is then welfare optimal because consumer surplus is strictly lower without privacy.

Proposition 4. *(1) If n is sufficiently large, welfare is higher with privacy than without. (2) Let the disutility of an individual facing action A by OP be $r\delta(\tau)$ (instead of $\delta(\tau)$). For r sufficiently large, welfare is higher with privacy than without.*

When is welfare higher without privacy than with privacy? Intuitively, if the correlation between θ_i and τ_i is very high: then OP's gain from being able to distinguish individuals according to type is also large, while the individual's loss from not being able to choose their preferred p_i (or being treated aggressively if they do) only depends on $\delta(\tau_i)$ and not on the correlation. For a given δ, the correlation between θ_i and τ_i would therefore have to be sufficiently high to make no privacy welfare-optimal. Figure 3 illustrates this intuition for the case of $n = 1$.

[12] We could use weights to sum up payoffs. As this would be equivalent to a rescaling, this would not change our results qualitatively. Our results from this section therefore apply if welfare is any convex combination of the welfare of individuals and OP.

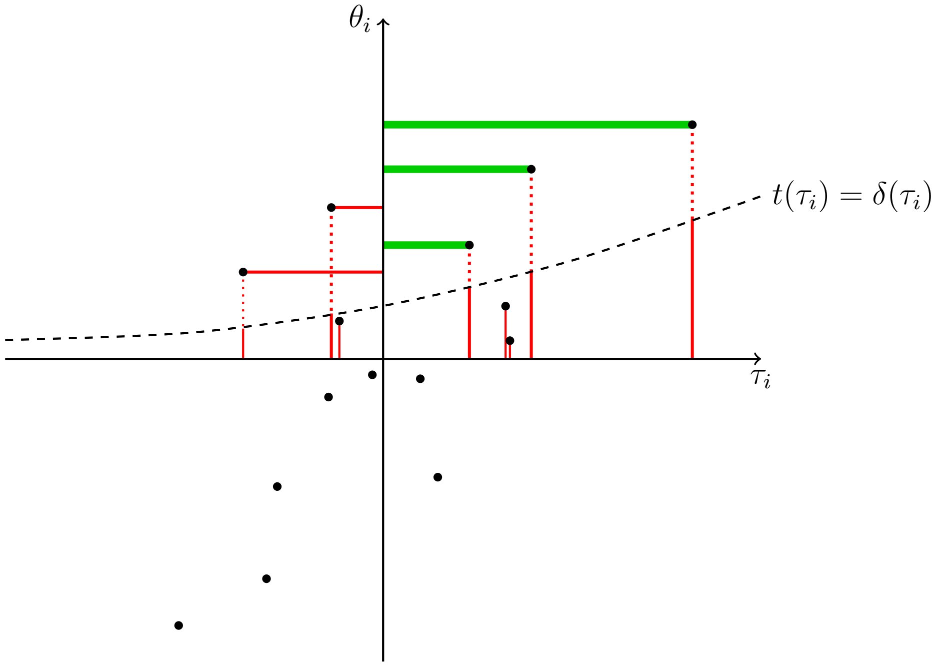

Figure 3: Gains (thick green lines) and losses (thin red lines) from lack of privacy (compared to privacy) for $n = 1$, for different types of individuals. Losses of the individual are vertical, losses and gains of OP are horizontal. The expected length of all green lines is the overall gain, the expected length of all (solid) red lines is the overall loss. An individual with $\theta_i > 0$ loses either θ_i (if she chooses $p_i = 0$) or $\delta(\tau_i)$ (if she chooses $p_i = 1$ and therefore gets treated aggressively). The OP gets τ_i for an individual who still chooses $p_i = 1$. Intuitively, if we increase correlation between θ_i and τ_i, an individual with $\theta_i > 0$ is likely to lie further to the right than before (as her expected τ_i increases), which increases the expected gain of OP.

If we want to analyze the connection between correlation and δ, we need to restrict the problem by imposing partial orderings of joint distributions, since the set of possible joint distributions is otherwise intractable. We will therefore make our argument in two ways that differ by how we order distributions. First, we restrict the joint distribution of θ_i and τ_i to a family of distributions which are convex combinations of a correlated and an uncorrelated distribution and show that – for a given δ – privacy is optimal unless the weight on the correlated distribution is sufficiently high.

In the remainder of this section, we will focus on the interesting case in which – given distributions $\Gamma_{\theta_i}(\tau_i)$ – there is a pure strategy equilibrium in which OP plays A (M) against those who choose $p_i = 1$ ($p_i = 0$). Now consider the marginal distribution of τ_i which we denote by $\bar{\Gamma}$:

$$\bar{\Gamma}(\tau_i) = \int_{\mathbb{R}} \Gamma_{\theta_i}(\tau_i)\, d\Phi(\theta_i).$$

$\bar{\Gamma}$ is the average distribution of τ_i (where the average is taken over θ_i). If for every given θ_i the distribution of τ_i was $\bar{\Gamma}$, then there would be no correlation between θ_i and τ_i and even knowing θ_i directly (instead of p_i) would not yield any benefit for OP. We will now consider convex combinations of the original distributions Γ_{θ_i} and the distribution $\bar{\Gamma}$. Denote these convex combinations by

$$\Gamma^{\lambda}_{\theta_i}(\tau_i) = \lambda \Gamma_{\theta_i}(\tau_i) + (1-\lambda)\bar{\Gamma}(\tau_i) \qquad \lambda \in [0,1].$$

For $\lambda = 1$ we are in the original problem. Decreasing λ, however, continuously decreases the correlation between θ_i and τ_i. For $\lambda = 0$, there is no correlation between these two variables left. If there is no correlation, then the equilibrium is the same as in the privacy case because OP does not get any information about τ_i from the choice of the individuals. Hence, the equilibrium is that OP plays M against everyone and individuals use the cutoff 0 if $\lambda = 0$. This is true regardless of whether there is privacy or not. By continuity, the same is true for low but positive λ. As λ increases OP finds it optimal to play A against $p_i = 1$. However, his benefit from doing so is at these intermediate values of λ not very large and therefore more than outweighed by the negative effects on the individuals (aggressive treatment and worse information aggregation). OP's gains are only sizeable when λ is sufficiently large. In this case, the welfare optimality of (no) privacy depends on parameter values. The proposition below establishes that privacy and no privacy are equivalent for very low values of λ and – more interestingly – privacy is welfare optimal for an intermediate range of λ.

Proposition 5 (Welfare optimality depending on type correlation)**.** *There exist* $0 < \underline{\lambda} < \bar{\lambda} \leq 1$ *such that*

1. *for* $\lambda \leq \underline{\lambda}$ *privacy and no privacy are welfare equivalent and*

2. *for* $\lambda \in (\underline{\lambda}, \bar{\lambda}]$ *privacy leads to strictly higher welfare than no privacy. The equilibrium for* $\lambda = \bar{\lambda}$ *is in pure strategies.*

In the remainder of this section we consider the special case $\delta(\tau) = \delta$, i.e. δ is constant. This allows us to use a more general ordering of distributions: Namely an ordering based on first order stochastic dominance of Γ_{θ_i}. Furthermore, it allows us to explore the effect of δ on welfare. We show that the welfare difference between no privacy and privacy is decreasing in δ and increasing in our measure of statistical dependence between θ_i and τ_i. This means the following: If the consequences of being treated aggressively by OP are severe (i.e. high δ), no privacy can only be optimal if the correlation between θ_i and τ_i is very high.

To introduce our more general ordering of distributions, recall that Γ_{θ_i} is the distribution of τ_i given θ_i; and that we have already assumed that $\Gamma_{\theta_i'}$ first-order stochastically dominates $\Gamma_{\theta_i''}$ if and only if $\theta_i' \geq \theta_i''$. Furthermore, we now assume that $\mathbb{E}\left[\tau_i | \theta_i = 0\right] \geq 0$ so that the expected τ_i is positive for $\theta_i > 0$ – this guarantees that OP wants to treat individuals aggressively if their θ_i is positive. We will now say that the correlation is higher in distribution Γ' than in distribution Γ'' if for every $\theta_i > 0$, Γ'_{θ_i} first-order stochastically dominates Γ''_{θ_i}. The following proposition shows that welfare is decreasing in δ and increasing in the correlation between θ_i and τ_i. This establishes that for a higher δ, the correlation between θ_i and τ_i needs to be higher to make no privacy welfare-optimal.

Proposition 6 (Monotone welfare difference)**.** *The welfare difference between no privacy and privacy is decreasing in δ and increasing in the correlation in Γ.*

Figure 4 summarizes our welfare results.

4. Alternative Utility Specifications

In this section, we discuss two alternatives to the information aggregation in the first stage modeled so far. First, we consider a setup where individual i's utility does not depend on choices of other individuals. That is, the first stage decision p_i is not about information aggregation but is simply a private decision without externalities. Second, we consider a setting in which there is again information aggregation but individual i's payoff from $p = 1$ is given by a common state θ (instead of a personal payoff parameter θ_i). This state, however, is unknown, and individuals obtain only noisy private signals of the true state θ. As we will see, similar results to the ones above hold in these setups and some additional insights can be obtained.

4.1. First Stage With Private Decisions Instead of Information Aggregation

We want to consider a setup where individual i's choice (p_i) directly influences his welfare. This is actually a special case of our model: If we set $n = 1$, we obtain a framework where by definition no externalities among players play a role. Note that in this case $p = p_i$ and the individuals payoff in the first stage can be written as $p_i\theta_i$. Clearly, all of our results continue to hold – with the obvious exception of the limit result for large n (proposition 3). In particular, there is still a chilling effect which leads to negative welfare consequences as described in the previous section.

Private decisions would be a reasonable assumption, for example, when considering first stage choices like listening to music, attending certain events or meeting certain people, which is also informative about some hidden type. In our example from the introduction, the question would be: If a preference for Reggae music is correlated with drug use, should the employer be able to observe, and base his decision on, the music that Alice listens to? We give another example below that emphasizes the result of proposition 2, i.e. the behavior change induced by abolishing privacy might render the additional information useless for OP.

Example 2. *Consider the case of data-based police work. The purchase of precision scales through the online retailer Amazon suggests that the buyer might be a drug dealer: the*

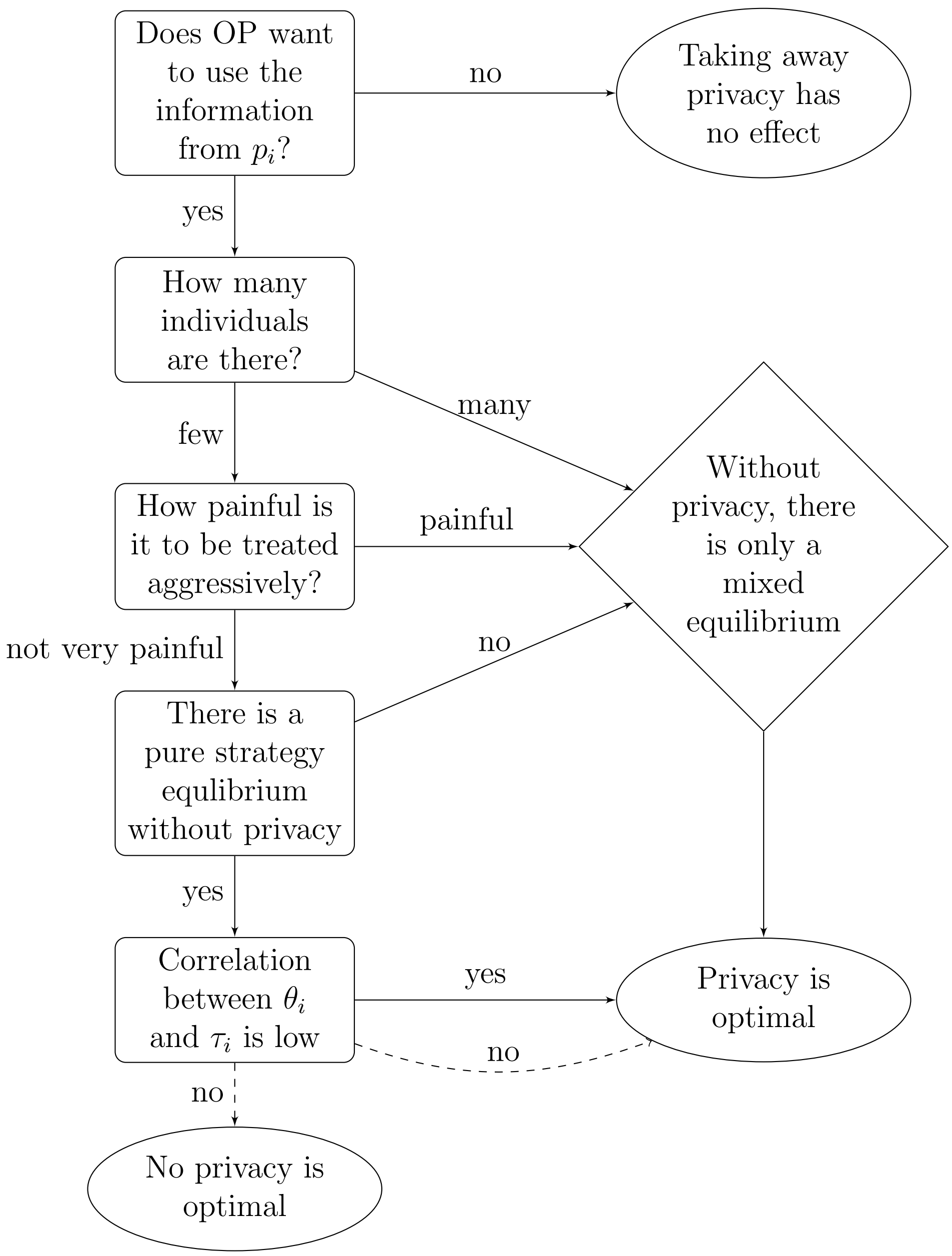

Figure 4: Sufficient conditions for when privacy is welfare-optimal, which follow from propositions 3 and 5. Note that the existence of a pure-strategy equilibrium without privacy and high correlation are only necessary, but not sufficient conditions for privacy to be suboptimal.

predictive algorithm that suggests other items based on what people usually buy together with the scale are almost all drug-related.[13] *Should the police (OP in our model) be allowed to access Amazon's purchase data? From the outset, it might seem that this could help to track down drug dealers. If, however, purchase data was used in this way, it is clear that drug dealers would be the first to procure their high precision scales in another way, and the police would be left with visiting a few enthusiastic coin collectors. The chilling effect would render the infringement of privacy useless.*

4.2. Individuals with Identical Preferences Under Uncertainty

This subsection considers an alternative model where the private information of individuals in the information aggregation stage is not directly their personal payoff of $p = 1$. Instead, individuals all have the same payoff of $p = 1$ but each one of them only receives a noisy signal of this payoff. That is, there is an unknown state of the world θ; each individual has a noisy signal about the state of the world and they try to "match the state", i.e. they prefer $p = 1$ if the state is positive and $p = 0$ if the state is negative.

This has a striking implication: Lack of privacy makes *every* individual worse off, since the chilling effect inhibits information aggregation. In our main model, individuals have private preferences over outcomes and therefore some of them (those with negative θ_i) gain from the chilling effect. Since all individuals now have the same interest – implementing the policy that matches the state – everyone loses from the chilling effect's impact on information aggregation. Hence, our welfare results in proposition 3 are somewhat stronger in this setting as privacy is now a Pareto improvement not only at the ex ante but even at the interim stage, i.e. after each individual has observed his signal.

The details of the setting are as follows: The state of the world θ is distributed standard normally and this θ is the payoff consequence of $p = 1$ for each individual. However, the realization of θ is unknown. Each individual obtains a private signal θ_i which is normally distributed around the true state θ, i.e. $\theta_i \sim N(\theta, \sigma^2)$ where we denote the cdf by $\tilde{\Phi}_\theta$ and the pdf by $\tilde{\phi}_\theta$. All θ_i are assumed to be independent draws from this distribution. The interaction type τ_i of individual i is drawn from Γ_{θ_i} where again $\Gamma_{\theta'_i}$ is assumed to first order stochastically dominate $\Gamma_{\theta''_i}$ if and only if $\theta'_i > \theta''_i$. This creates a positive correlation between θ_i and τ_i. The interaction stage is exactly the same as in our main model. That is, without privacy a strategy for OP states which of the two actions (A and M) OP plays against an individual who chose $p_i = 0$ or $p_i = 1$. With privacy, OP only decides which of the two actions he chooses against all individuals. This means that – to keep the setting comparable to the main model – we do not consider strategies (or beliefs) that are contingent upon the number of individuals choosing $p_i = 1$. This is a simplification. However, one can easily imagine settings where OP has to commit to a strategy before he gets to know the individuals' p_is. This is the case, for example, if the interaction is

[13] See https://www.cnet.com/uk/news/buy-a-scale-on-amazon-and-it-thinks-youre-a-drug-dealer/, retrieved September 29, 2016.

between i and an agent representing OP and p_i is only learned in the interaction. OP then has to instruct the agents in advance how to act.

In the supplementary material to this paper, we provide proofs that are mostly analogous to those of our main model. In particular, the absence of privacy causes a chilling effect and this chilling effect inhibits efficient information aggregation. For large n, there are still only equilibria where OP mixes, and in any equilibrium where OP mixes, OP is indifferent between privacy and no privacy. Now, however, we have Pareto dominance in the sense that every individual of every type is better off under privacy.

5. Extensions

This section contains five extensions to the main model. First, we relax the assumption that the information aggregation rule q is the same in the privacy and the no privacy case by allowing some hypothetical planner, who can be interpreted as a social norm or an institution, to ex ante choose a q in order to maximize the expected benefit from information aggregation. Second, we generalize the information aggregation mechanism q and show for which types of mechanisms our main results hold. Third, we show that privacy might have to be mandated, i.e. privacy as an opt in possibility will lead to the no privacy outcome. Fourth, we ask whether and when the introduction of a price for information gathering can improve welfare. Fifth, we consider the possibility of a defensive action against OP and use this setup to show that in some scenarios privacy can even make OP strictly better off.

5.1. An Endogenous Information Aggregation Process q

In this section we will endogenize the function q that assigns to each m/n a probability of implementing $p = 1$. In particular, we will assume that this function q is chosen by a planner in order to maximize the surplus in the information aggregation stage. The planner takes into account that individuals are chilled in the no privacy case and therefore the optimal q will differ in the privacy and no privacy case. The goal of this section is to show that our results from sections 2 and 3 remain valid in this setting.

We will assume that the planner has to choose an increasing function q and this function depends on the case – privacy and no privacy. For simplicity of exposition, we will assume that n is odd which ensures that a majority rule is clearly defined. Since we do not assume that q is *strictly* increasing, we will require individuals to choose a cutoff strategy $t(\tau)$ (after observing q) and we will concentrate on equilibria where each individual chooses the same cutoff strategy.[14] Consider the privacy case first. For any increasing q, it is a best response by the individuals to choose cutoff $t^p(\tau) = 0$. Given the

[14] This assumption rules out equilibria in which no individual can influence the outcome; e.g. if q is a majority rule and $n \geq 3$, an unreasonable equilibrium exists in which all individuals always choose $p_i = 0$ (regardless of type). If several equilibria exist that satisfy our assumption, we allow the planner to select the one that maximizes payoffs in the information aggregation stage.

independence of the θ_i, the planner's optimization problem is then

$$\max_q \sum_{m=0}^{n} q(m/n) \left(m\mathbb{E}[\theta_i|\theta_i \geq 0] + (n-m)\mathbb{E}[\theta_i|\theta_i < 0] \right).$$

As θ_i is normally distributed, $\mathbb{E}[\theta_i|\theta_i \geq 0] = -\mathbb{E}[\theta_i|\theta_i < 0]$ and it is easy to see that the optimal q is a majority rule, i.e. $p = 1$ if more than $n/2$ individuals choose $p_i = 1$ and $p = 0$ otherwise.

Now suppose for a moment that the planner could choose both t and q with the goal of maximizing expected surplus in the information aggregation stage. Given the symmetry of our setup, one can show that the planner would then choose $t = 0$ and majority rule; see the supplementary material for more details. That is, the privacy case delivers the maximal possible payoff in the information aggregation stage.

Next we consider the no privacy case. Obviously, a constant q is not optimal and therefore each individual's p_i will – with some probability – influence the decision on p. As in the main model, individuals will still be chilled to some extent if OP's behavior depends on p_i, i.e. $t_q^{np}(\tau) \geq 0$ with strict inequality if OP's behavior depends on p_i. Note that the threshold t_q^{np} depends on q. The planner's problem is now

$$\max_q \sum_{m=0}^{n} q(m/n) \left(m\mathbb{E}[\theta_i|\theta_i \geq t_q^{np}(\tau_i)] + (n-m)\mathbb{E}[\theta_i|\theta_i < t_q^{np}(\tau_i)] \right).$$

Given that $t_q^{np} \geq 0$ and that all θ_i are normally distributed, $q(1) = 1$ and $q(0) = 0$ are clearly optimal. All other values cannot be determined in general, that is, without specifying Γ_θ, although it is clear that there will be a cutoff such that $q(m/n)$ is 1 (0) for m above (below) the cutoff. Fortunately, our welfare result in proposition 3 can be derived without precise knowledge of q. The main argument is that the probability with which an individual expects to influence the decision on p is bounded from above by $1/n$. This follows directly from the assumption that all n individuals use the same strategy. Consequently, the same forces as in the original model are at work (regardless of the specific q): As n increases each individual is less likely to be pivotal and therefore the main motivation in the p_i choice is to avoid aggressive treatment by OP. The threshold t_q^{np} becomes arbitrarily high and steep and – due to the same reasoning as in the proof of proposition 3 – only mixed strategy equilibria exist. This implies that OP is indifferent between privacy and no privacy. As mentioned above, the privacy case maximizes the expected payoff from information aggregation and therefore privacy welfare dominates no privacy.

Similarly, the argument of proposition 5 does not rely on the specific shape q. The crucial part for this result is that t_q^{np} does not depend on Γ_{θ_i} for a given OP strategy. This is generally true as each individual knows the realization of its type when acting in the information aggregation stage. It follows that the optimal (q, t_q^{np}) pair (without privacy) is the same for all λ in which OP plays A (M) against $p_i = 1$ ($p_i = 0$) in equilibrium. Consider the same scenario as in proposition 5 where $\Gamma_{\theta_i}^{\lambda}$ is given by a convex combination of a correlated and an uncorrelated distribution. Assume that under the correlated distribution Γ_{θ_i} there is a unique equilibrium without privacy in which OP plays A (M) against $p_i = 1$ ($p_i = 0$) while with privacy the unique equilibrium has OP playing M. The latter condition implies that for very small λ the equilibrium without privacy is the same as with privacy (OP playing M and $t_q^{np} = 0$). For λ very high OP plays A against $p_i = 1$ in the unique equilibrium. Denote the smallest λ where there is an equilibrium in which OP plays A against $p_i = 1$ for sure by λ^* (given the optimal q and t_q^{np} for this OP strategy which do not depend on λ). The individuals' equilibrium threshold is the same regardless of λ as long as OP uses the strategy of playing A (M) against $p_i = 1$ ($p_i = 0$). Therefore, the reason why such an equilibrium no longer exists for $\lambda < \lambda^*$ is that OP does not find it optimal to play A against $p_i = 1$ because the correlation between θ_i and τ_i is too low. Hence, OP is indifferent between his two actions when $\lambda = \lambda^*$ and $p_i = 1$. This implies that for λ slightly above λ^* privacy is welfare optimal: Since OP is almost indifferent between between A and M when facing $p_i = 1$, his welfare loss of privacy is very small while the welfare gain for the individuals is substantial. For $\lambda < \lambda^*$, the equilibrium without privacy is either mixed or equivalent to the equilibrium with privacy. Consequently, privacy is (weakly) welfare optimal also for these values of λ. This establishes the same result as in proposition 5.

5.2. General Information Aggregation Processes

In our main model, we have assumed that $q(m/n) = m/n$. We can show, however, that our results from propositions 1 to 5 qualitatively carry over to the more general case of information aggregation mechanisms that are (i) strictly increasing, (ii) unbiased, (iii) anonymous, and (iv) centrally pivotal. By "centrally pivotal", we mean that under our distributional assumptions, q is such that a given individual is more likely to be pivotal the more evenly preferences are distributed in the population.

Formally, our requirements mean that q must be monotone in its argument, point-symmetric around 0.5, and s-shaped, i.e. weakly convex in $[0, 0.5]$ and weakly concave in $[0.5, 1]$. (The linear function that we have assumed in our main model, and majority voting, which we discussed in the preceding section, can be seen as two extreme cases of s-shaped mechanisms.) Propositions 3 to 5 still apply, for the reasons given in section 5.1 above, if we can show that the reasoning in lemmas 1 and 3 is still valid – since in these lemmas a different q could conceivably produce different results. We provide these proofs in the supplementary material. In section 6, we comment on how exceptions to our assumptions can mean that privacy is not optimal.

5.3. Privacy as Opt In

Suppose that each individual has an additional decision to make in the information aggregation stage: They do not only have to choose p_i but also have to decide whether their choice should be private or public. OP can observe all public choices but not the private ones – in this case he can only observe that the individual chose privacy. To isolate the effect of the privacy choice, we will also assume that OP cannot make his behavior contingent on the outcome p (which might be realized only at a later point of time).

The possibility of hiding one's choice gives rise to multiple equilibria. To see this, consider first an equilibrium in which every individual always chooses "public" (no matter what θ_i, τ_i or p_i is). Then the equilibrium of the case without privacy results.[15] Second, consider an equilibrium in which every individual always chooses "private". This means that we are effectively in the case with privacy. OP's best response is to play M and consequently no individual has an incentive to deviate.

Naturally, the question arises which of the two equilibria is more robust. We will argue in two different ways that the "always private" equilibrium is not very robust. The reason is an unraveling logic. Individuals who choose $p_i = 0$ are not afraid of making this public as it suggests that their θ_i is low, which means that their expected τ_i is also relatively low because of the positive correlation between the two. Given that the expected τ_i is low, OP would therefore still play M against those who make a choice $p_i = 0$ public. If, however, everyone who chooses $p_i = 0$ makes this public, then making one's choice private is not different from publicly choosing $p_i = 1$.

The simplest way to formalize this intuition is to assume that making one's choice p_i private comes at a small cost $\varepsilon > 0$. In this case, the "all private" equilibrium would only be supported by off equilibrium beliefs such that both $\mathbb{E}[\tau|\text{"public"}, p_i = 0] \geq 0$ and $\mathbb{E}[\tau|\text{"public"}, p_i = 1] \geq 0$ as OP could then threaten to play A against anybody making his decision public (thereby saving the $\varepsilon > 0$ costs). Given that $\mathbb{E}[\tau] < 0$, these are straightforwardly unreasonable beliefs. In terms of equilibrium refinements, the equilibrium does not satisfy the well known D1 criterion of Banks and Sobel (1987). Roughly speaking, this refinement states the following for our game: Denote by $D(\theta_i, \tau_i)$ the set of OP mixed strategies that are (i) best responses for some OP belief and (ii) would make a deviation by an individual of type (θ_i, τ_i) profitable. D1 requires that OP's off path beliefs must be zero for type (θ_i', τ_i') if there is a type (θ_i'', τ_i'') such that $D(\theta_i', \tau_i')$ is a strict subset of $D(\theta_i'', \tau_i'')$. Put differently, when facing an off-path deviation, OP should believe that it is more likely to be committed by a type whose deviation could be justified by a bigger set of OP beliefs. It is straightforward to show that the "all private" equilibrium does not satisfy D1. The reason is that the off path beliefs supporting the "all private" equilibrium require that deviations to public stem from individuals with relatively high τ_i

[15] This equilibrium is supported by the following off equilibrium path belief: if a player chooses "private", OP believes that τ_i is sufficiently high so that A is a best response.

no matter whether p_i is zero or one. As δ is increasing in τ_i, there are mixed strategies by OP which would make the deviation profitable for individuals with low τ_i (who are less afraid of action A) but not for individuals with high τ_i. The "all public" equilibrium, on the other hand, satisfies D1.

The second way in which the "all private" equilibrium is not robust is the following. Assume that with probability $\varepsilon > 0$ OP has the alternative payoff $\tau_i + \varepsilon'$ from playing A. Assume that ε' is such that $\mathbb{E}[\tau] + \varepsilon' > 0$. That is, under the alternative preferences OP plays A given his prior beliefs. Suppose further that these alternative preferences are such that $\mathbb{E}[\tau|\theta_i \leq 0] + \varepsilon' < 0$, i.e. knowing that θ_i is negative OP still best responds by playing M. Again the "always private" equilibrium could then only be sustained by off path beliefs leading to $\mathbb{E}[\tau|\text{"public"}, p_i = 0] + \varepsilon' \geq 0$ and $\mathbb{E}[\tau|\text{"public"}, p_i = 1] + \varepsilon' \geq 0$. As pointed out above, such beliefs are unreasonable and violate the D1 refinement.

5.4. Can Prices Improve Welfare?

So far, we have considered privacy as a feature of the model that is externally imposed by a regulator (or by nature). In the previous section (5.3), we have already considered the case where individuals can choose their own privacy (and why this usually does not lead to optimal allocations). Our analysis also allows us to state a corollary result on whether a general price on information can improve welfare and lead to an optimal allocation of information.

Consider a world without privacy in which OP has to pay price P to observe all p_i. If he does not pay P, he cannot observe any p_i and has to treat everybody mildly. P could either be an actual cost, or a fee that is imposed by a regulator.

Timing could take one of two possible forms: Either OP has to choose whether to pay P first and this is observable to the individuals, or both choices are made simultaneously.[16] In the first case, OP effectively chooses between privacy and no privacy, and individuals adjust accordingly. In particular, OP chooses privacy as long as the cost P is at least as big as his expected gain from the interaction stage if there were no privacy. That means that for any positive cost $P > 0$, OP chooses privacy in any of the scenarios in which privacy is Pareto-optimal. That is not necessarily true, however, if privacy is efficient without being Pareto-optimal. As OP only considers his own gain, he could gather information even though privacy is efficient (if P is too low to reflect the individuals' loss) or could decide not to gather information (if P is high). P would have to be set exactly right to guarantee an optimal allocation.

The problem becomes somewhat more interesting if we consider the case in which individuals do not learn whether OP can observe p_i before they make their choice. In any pure-strategy equilibrium, individuals correctly anticipate being observed and the results are as in the sequential case. But if either n or δ are sufficiently large and $P > 0$, there

[16]The possibility that individuals choose p_i before OP chooses P is equivalent to simultaneity since it makes no sense to assume that OP can observe p_i when choosing whether to get information.

only exists a mixed equilibrium in which OP sometimes gathers information and treats everybody who has chosen $p_i = 1$ aggressively; individuals adjust by playing a threshold strategy $t(\tau_i) > 0$. (If P becomes very large, the privacy equilibrium in pure strategies is the unique equilibrium.)

In any such mixed equilibrium, OP mixes between gathering information and not gathering information. The latter gives zero payoff (since he has to treat everybody mildly), such that his expected equilibrium payoff is zero. That means that in equilibrium, individuals choose a threshold $t^*(\tau_i)$ such that OP's expected information gain is exactly counterbalanced by the price P that he pays for the information. OP mixes between gathering information and not gathering information such that $t^*(\tau_i)$ is optimal for the individuals.

A rise in P hence shifts the equilibrium in the following way: In the new equilibrium, individuals play a (weakly) lower threshold strategy, which increases OP's information gain to compensate him for the rise in P. OP gathers information with a lower probability. The following corollary results from applying lemma 3 to this comparative static:

Corollary 1. *If information collection costs are P and privacy would be Pareto-optimal, raising P leads to Pareto gains. If P is a newly introduced fee (or tax) on information gathering, it generates Pareto gains* and *raises revenue.*

5.5. Defensive Actions

Suppose that individuals have the opportunity to take a defensive action against being treated aggressively. More precisely, an individual can take an action D which increases his payoff if OP plays A but decreases his payoff if OP plays M. The defensive action reduces OP's payoff. In our example, Alice could hire a lawyer. Hiring the lawyer is costly but the lawyer will make it harder for the employer to discriminate against Alice. For the employer, dealing with a lawyer is a hassle (whether he discriminates or not) and reduces his payoffs.

What we want to illustrate is that the model can easily be extended in this way and that privacy could lead to (i) OP being *strictly* better off with privacy while (ii) individuals are in expectation strictly better off with privacy. Hence, privacy can be strictly Pareto superior from an ex ante point of view. To this end, it is sufficient to present an example with these features and we provide such an example in the supplementary material.

6. When is Privacy Bad?

So far, we have mostly concentrated on situations and sufficient conditions under which privacy is beneficial for society, since this is the main focus of our paper. But our model, and the assumptions under which we have derived our main results, also allow us to identify conditions under which privacy is not welfare-optimal – under which, in other words, intrusions into privacy can be efficient. In addition to the conditions that follow from our results in sections 3 and 4, we will briefly comment on some additional restrictions here.

Biased or non-centrally-pivotal information aggregation: In section 5.2, we have shown that our results apply for a wide class of information aggregation mechanisms. While our conclusions may hold for some mechanisms that do not fall into this general class, there are mechanisms for which our results do not apply.

One of the assumptions of our model has been that the information aggregation mechanism q is unbiased. If this assumption is not fulfilled, q systematically prefers either of the two options 0 and 1. This can be the case, for example, in situations where there is a bias for the status quo and it can only be changed with a supermajority. In this case, the information aggregation mechanism in itself is clearly not welfare-optimal from a narrow utilitarian point of view (even though it could of course be justified by other considerations, such as a desire to protect minorities). With such a biased information aggregation mechanism, the behavior of individuals under privacy is not welfare optimal and can potentially be improved by distorting it.

However, giving up privacy can only improve welfare in this case if it distorts behavior in the "right" direction. Assume, for example, that q is biased such that it chooses option 1 with probability (almost) one if $m/n > 0.1$, and chooses option 0 with probability (almost) one otherwise. Lack of privacy, by lowering the propensity of individuals to support 1, can improve the probability that option 0 is chosen in cases where, for example, 60% of individuals prefer 0. This does not work the other way around: If q has a similar bias towards 0, lack of privacy will exacerbate the situation by guaranteeing that 1 gets chosen in even fewer situations where it would be the welfare-maximizing choice.

We are not entirely sure how relevant these considerations are in most real-life examples, since they require that information aggregation and the actions of OP are biased in *opposite* directions. In our drug legalization example, it would require that drugs are more likely to be legalized than is optimal for the population, but that it is undesirable to be seen as a supporter of legalization. Then, and only then, can welfare be improved by removing privacy and thereby deterring some people from supporting legalization.

Another mechanism for which the chilling effect can improve welfare is a mechanism that is symmetric, but not centrally pivotal (not "s-shaped"). Such a mechanism would be more dependent on m/n if this fraction is very small or very large than if it is close to 0.5. Consider, for example, a mechanism that chooses 0 (1) if m/n is below 0.2 (above 0.8), and otherwise just flips a coin to determine p. If the θ_i are symmetrically distributed, this mechanism would make it quite unlikely that p has anything to do with people's preferences. If the chilling effect were to push the choices of individuals in either direction, they could (even if roughly evenly split in terms of preferences) get closer to the critical threshold at which m/n influences p.

Just like in the case of biased mechanisms that we discussed above, we do not think that such mechanisms are especially prevalent. However, such a mechanism could be a valid description of s situation in which decisions can only be made by a large majority, and if there is no such majority the decision is made according to other (less relevant and more or less random) criteria. Then a chilling effect (carefully calibrated so as not to be too large) could improve welfare.

Figure 5 illustrates some mechanisms for which our results hold (panel A) and for which they do not hold (panel B).

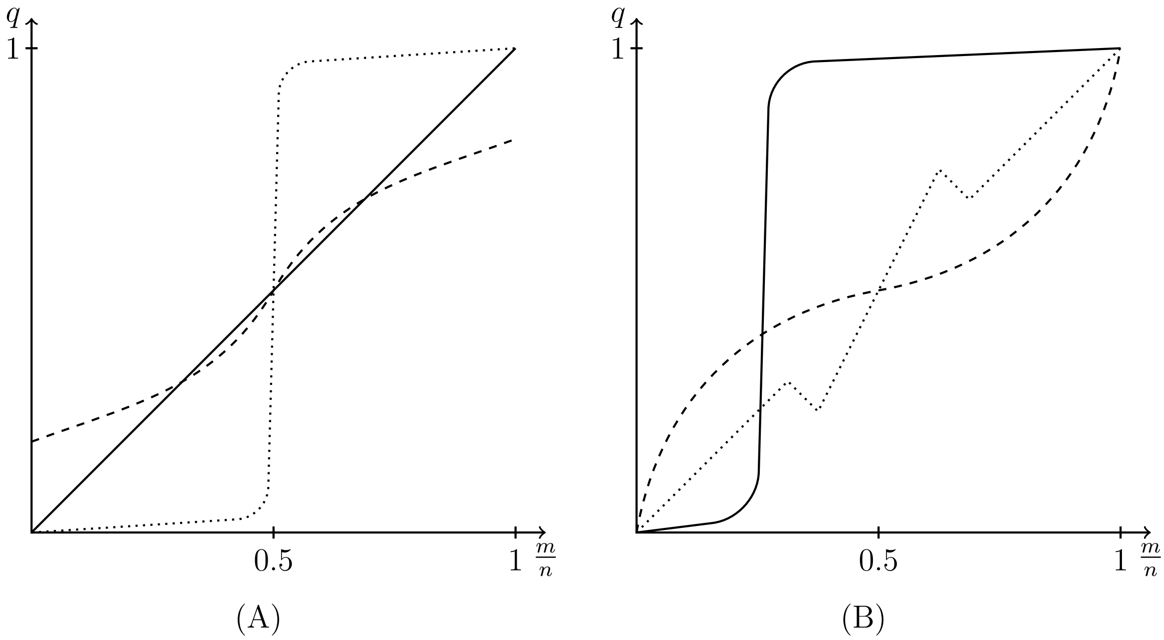

Figure 5: Left: An illustration of mechanisms for which our results from propositions 1 to 5 hold. Right: Mechanisms that are biased (solid line), not centrally pivotal (dashed) or non-monotonic (dotted).

Negative externalities from choices: We could think of situations where one of the choices that individuals can take is inherently more desirable from a welfare perspective. For example, if OP is trying to distinguish between criminals and non-criminals, and criminals are also more likely to enjoy engaging in small-scale vandalism, then introducing video surveillance (to detect vandalism) will have the benefit of identifying some potential criminals (those who are not subject to the chilling effect) *and* of deterring vandalism (through the chilling effect). While such externalities add an additional complication to our model, we can accommodate them by assuming that they subtract a certain length from all vertical loss lines in graph 3 that end below the curve of $t(\tau_i)$ (potentially making them negative). Our main welfare results change accordingly. No privacy can now be optimal even in cases of mixed equilibria if the gain that results from the chilling effect is large enough. No privacy is also optimal for a larger interval of correlation parameters.

7. Examples

7.1. Information Aggregation and Sorting Among Individuals: Opinion Polls and the Secret Ballot

The main informational tension that underlies questions of information aggregation and privacy is between inducing individuals to reveal valuable information by promising them influence on an outcome, and the countervailing threat of using such information to discriminate among individuals.

Perhaps the most prominent example, and one where almost everyone's intuition will come down on the side of privacy at least some of the time, is voting. Democratic societies use elections to collect information about their citizens' values, opinions, beliefs and preferences. It seems intuitively clear (similar to our lemma 3) that the information aggregation in such elections can usually not be improved by making public how individuals have voted, as this would allow partisans of a candidate (OP of our model) to intimidate or reward voters who would otherwise express themselves freely. In our main model, we derive sufficient conditions for when the information aggregation of a secret ballot cannot be improved by removing secrecy, and we further develop this result in sections 5.1, 5.2 and 6.

The effect of privacy can also be observed in the problem of predicting election outcomes with opinion polls. Polls are usually conducted by interview and therefore offer less privacy than actual elections. Respondents may therefore adjust their answers to what they think the pollster wants to hear. This can be motivated by a fear of actual reprisals, or simply of being viewed unfavorably by the pollster conducting the interview – both are equivalent to the δ of our model.[17] At the same time, opinion polls offer only very limited influence to anyone who answers them, so that the δ can easily outweigh the benefit of answering honestly. The resulting bias in polls towards more socially acceptable options has become known under different names, such as the "Bradley effect" or the "Shy Tory factor".[18]

For a classic example of how the chilling effect can lead to a systematic error in opinion polls, consider the 1990 presidential election in Nicaragua. Opinion polls for this election varied widely and mispredicted the result substantially. Bischoping and Schuman (1992), in a well-known experiment, deployed researchers who conducted opinion polls while subtly looking as if they supported one of the candidates. This was achieved by having the pollsters use pens showing the symbol of either of the candidates to fill out the poll. Polls who were thus "associated" with different parties produced different results. In

[17]"Being viewed unfavorably" might seem like a reduced-form reputation effect, but if it causes direct disutility there is no difference in how our model would cover it in terms of informational impact.

[18]Newer terms like "Brexit effect" or "silent Trump vote" suggest that the phenomenon persists.

particular, polls who were either neutral or visually associated with the incumbent were quite different from the election result, while polls that seemed to be associated with the challenger were much closer to it. This suggests that, without the privacy of the voting booth, respondents feared the potential costs of revealing their opinion to an "OP", and that this fear substantially reduced the informativeness of their answers.

Of course, a crucial point in this example is that the elections in Nicaragua were widely expected to be fair and secret, so that the chilling effect in the polls was not likely to be reproduced in the actual election. If a chilling effect will occur in the election itself, a poll is more likely to predict the result if it induces an effect of a similar magnitude in the same direction.

7.2. Which Discrimination Should be Permitted: Credit Scores

Consider the problem of a bank deciding to whom to lend. Ideally, it would like to base its decision on the probability that a debtor will repay the loan, but this variable is not directly observable. Instead, the bank can rely on measures that indirectly predict default probability. There are several socioeconomic variables that are easily observed and correlated with default risk, such as national origin, race, gender, age, or place of residence. But using such variables to make credit decisions, and hence treat native-borns, whites or women differently solely because of their identity, is illegal in many countries. In the United States, for example, such "redlining" practices are explicitly outlawed by the Equal Credit Opportunity Act (ECOA) of 1974.

Imagine, however, that the bank starts looking for other pieces of data that can inform its decision and allow it to statistically discriminate among loan applicants. Two such pieces of information are the education level (which can easily be documented by the applicant) and the taste in music (which many millions of people reveal on various websites and in buying decisions). While usage of the former information is common practice, the latter is more speculative but not implausible: Facebook owns a patent on aggregating credit scores from the data it collects about its users, and there are many firms that claim to make use of big data to develop more accurate credit scores.[19]

We would expect that a preference for some genres of hip hop, since it is correlated with socioeconomic status, can be highly predictive of default risk. The expressed music preference would then be the variable p_i that the bank uses to discriminate between people who do and those who don't get loans, and our model would consequentially predict a chilling effect in which some hip hop fans are held back in their freedom of expression, since they want to improve their credit rating. The individual loss (of not being able to express their personality) is probably more substantial than the loss in information aggregation here, but it is a welfare loss nonetheless (cf. the results in section 4.1). On top of that, since loan decisions can be of huge importance to an individual, we would

[19]One of them, Zest Finance, advertises with the slogan: "All Data is Credit Data." (https://www.zestfinance.com/how-we-do-it.html, retrieved May 2, 2016.)

expect the equilibrium informativeness of revealed music preferences to be quite low (in line with proposition 3), so that the considerable welfare loss among consumers is not counterbalanced by a massive gain in information for the bank.

But it should also be noted here that fans of gangsta rap music tend to be similar to each other in many ways, so that the use of innocuous (and predictive) music preference data allows the bank to discriminate based on ethnicity, age and geography without explicitly saying so. (Possibly even unknowingly: If decisions are made or supported by a machine-learning algorithm, the bank would not necessarily understand what they are based on.) This points to a larger question to which our research contributes, but to which we have no definitive answer: What should banks, employers, governments be allowed to discriminate upon? Most people would probably agree that to treat someone better or worse purely because of race or gender is not acceptable (and that contrary to the arguments made by Friedman, 1962, such discrimination will not automatically disappear as it can be rational statistical discrimination). But demanding that job applicants have a diploma, or giving loans based on past income, is also statistical discrimination: these factors are predictive of whether the employee will be up to the task or the loan will be repaid, but the correlation is less than 1.

Our first extension suggests that an equilibrium where everyone keeps their music preferences secret is not stable (especially if there is some payoff to sharing them). Regulation which prohibits the use of some data for credit decisions, beyond existing laws like the ECOA, could therefore be welfare-enhancing.

Since our model only requires that variables are statistically dependent without necessarily being causally related, many other variables might be informative about creditworthiness. Clever bankers, or even mindless machine learning algorithms, could pick up on those relationships and use them to improve credit decisions. A regulator would be forced to keep up by continuously evaluating which new sources of information could give rise to unwanted discrimination. Our results would therefore support the regulatory use of "whitelists", which specify which data can be legally used in credit decisions (as opposed to "blacklists", which only specify which data cannot be used).

7.3. "The Tape Has Had Some Chilling Effect": Decision-Making and Transparency

The last decades have seen a move towards transparency in many public bodies – governments, authorities, central banks. But to the extent that the quality of decisions in these institutions depends on aggregating the information of their employees and members, our results suggest that transparency does not necessarily improve welfare – regardless of how highly you weigh the public's interest in being informed. Transparency itself can destroy the very information that it was supposed to reveal.

Consider, for example, the board of a central bank that has to decide on monetary policy. If the deliberations are private and no minutes are made public, board members express their opinion quite freely.[20] If minutes are later published, however, members will worry about the reputation effect of what they say.

A proponent of public meetings could argue that openness can discipline board members who might otherwise be beholden to special interests. But even if the public can observe the board's deliberations, it is still unobservable *why* someone makes or rejects a suggestion. If we see that a board member supported low interest rates ($p_i = 1$), we know that this is the policy she prefers ($\theta_i > 0$) – but does she prefer it because she thinks it the right strategy, or because it benefits her friends in the financial industry? Regardless of this uncertainty, it may be rational for the public to discriminate and accuse all those who supported $p = 1$ of being corrupt. But then, of course, people who support low interest rates will hold back, and the board may struggle to aggregate its members' opinions. If the board's size is large enough compared to how worried members are about their reputation, and corruption is not endemic (so that correlation between θ_i and τ_i is low), the public would gain no information from being able to follows the meetings – without having gained any improvement in the quality of decisions.

This is in line with the effects of a reform introduced in 1993, which mandated that minutes from meetings of the Federal Open Markets Committee (FOMC) of the U.S. Federal Reserve should be published after a short delay. Meade and Stasavage (2008) found that the reform significantly increased conformity and decreased the number of people who criticized the chairman's proposed interest rate adjustment. Thomas Hoenig, president of the Federal Reserve Bank of Kansas City, remarked in a meeting in 1995 that "the tape has had some chilling effect on our discussions. I see a lot more people reading their statements" (Meade and Stasavage, p. 13).

Our model therefore suggests that if board members, government ministers or civil servants are worried about how they are being perceived by the outside world, secret meetings can substantially increase the quality of decision making without depriving the public of any meaningful information.[21]

But what is more, our model allows us to weigh the disciplining motive of publicity against the loss in information aggregation – taking into account the fact that the disciplining can in itself be ineffective at finding those who need to be disciplined if the committee has many members, if members are very concerned about their reputation or if the correlation between preferences and corruption is sufficiently small. Privacy is not a panacea, but neither is transparency.

[20] This is under our standard assumption that arguing one's viewpoint increases the probability that one's preferred policy will be implemented.

[21] Consider also the literature on reputation concern and advice, such as Ottaviani and Sørensen (2006), which would also suggest that advisers are more helpful if they are unconcerned about their reputation.

8. Conclusion

Why should an individual care about his or her privacy, why should a society care about the privacy of its members? We have argued that since asymmetric information is a fact of life, questions of privacy are never about whether there should be private information or not, but only how much there should be and how it should be structured. That allows us to answer: Individuals can worry that information about them could be used "against them", i.e. expose them to discriminative treatment. This result does not require ill will among the discriminator – the discrimination can be perfectly rational, as in the case of the employer trying to distinguish applicants. But it will make it harder for people to choose according to their preferences, and the rational reaction of individuals to having no privacy can impair the ability of a society to efficiently aggregate information while providing at most minor gains. Privacy is not only individually optimal, but also welfare-enhancing.

Our examples show, however, that privacy is not a silver bullet. The solution to problems of "redlining" and new forms of discrimination in lending is not to prohibit borrowers from revealing any information about themselves; and not all governments would be improved by being able to work in total secrecy. Our analysis allows us to distinguish when privacy can improve welfare, and when it cannot.

Apart from the welfare effects, privacy often has a distributive effect: In our main model, there are always people whose preferred policy becomes less likely to be implemented under privacy. (In section 4.2, however, we argue that there can be situations where privacy improves everybody's outcome.) Others gain: Those who would be subject to the chilling effect without privacy are more likely to get their preferred option with privacy. Moreover, those with strong preferences gain twice from privacy: They are no longer statistically discriminated against, and their preferred option is more likely to be implemented. How should such distributive effects influence whether privacy is implemented? We have no definitive answer, but would like to point out that similar distributive effects arise with free speech: On any single issue, many would prefer if those with opposing viewpoints were prohibited from expressing it. Yet in the abstract, most of us would agree that freedom of expression should be universal.

We started this paper by criticizing the "Chicago view", that perceives privacy as inefficient and economically undesirable. But as we have argued that privacy can be fundamental to allowing individuals to freely express themselves, we are returning to another "Chicago argument": In his discussion of "rules instead of authorities", Friedman (1962, p. 52) considers the question of whether free speech issues should be decided from case to case, or in the abstract. He concludes that:

> When a vote is taken on whether Mr. Jones can speak on the corner, it cannot allow [...] for the fact that a society in which people are not free to speak on the

> corner without special legislation will be a society in which the development of new ideas, experimentation, change, and the like will all be hampered in a great variety of ways that are obvious to all.

Our analysis suggests that a similar argument can be made about privacy.[22]

[22] It has been pointed out to us that the whistleblower Edward Snowden drew a similar comparison between privacy and free speech in an online debate: "Arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say." (https://www.reddit.com/r/IAmA/comments/36ru89/just_days_left_to_kill_mass_surveillance_under/crglgh2, retrieved on July 1, 2016.)

Appendix

Technical Results

Lemma 4. *Let Φ be the standard normal distribution. Then $\int_{ka-b}^{ka} d\Phi / \int_{ka}^{\infty} d\Phi$ diverges to infinity as $k \to \infty$ for a, $b > 0$.*

Proof of lemma 4: We concentrate on the right tail of the standard normal distribution. If for all $x \in [ka - b, ka]$ and some constant c we have that $\frac{\phi(x)}{\phi(x+b)} \geq c$, then is is also true that

$$\frac{\int_{ka-b}^{ka} d\Phi}{\int_{ka}^{ka+b} d\Phi} \geq c.$$

(This can be seen by noting that the first inequality holds for the range of the integrals of the second inequality.) The pdf of the standard normal distribution is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

and the quotient of $\phi(x)$ and $\phi(x + b)$ is therefore $e^{-\frac{1}{2}(x^2-(x+b)^2)} = e^{xb+\frac{1}{2}b^2}$. For $x \to \infty$, this quotient diverges, and hence $\frac{\int_{ka-b}^{ka} d\Phi}{\int_{ka}^{ka+b} d\Phi}$ diverges for $k \to \infty$. Now note that $\int_{ka}^{\infty} d\Phi = \int_{ka}^{ka+b} d\Phi + \int_{ka+b}^{ka+2b} d\Phi + \ldots$ and that for large k, the quotient between any summand on the RHS and the following summand diverges. This means that the overall sum is smaller than $2\int_{ka}^{ka+b} d\Phi$ as – for k sufficiently high – $\int_{ka}^{ka+b} d\Phi + \int_{ka+b}^{ka+2b} d\Phi + \ldots \leq \int_{ka}^{ka+b} d\Phi \sum_{i=0}^{\infty}(1/2)^i = 2\int_{ka}^{ka+b} d\Phi$. Since we have established above that $\frac{\int_{ka-b}^{ka} d\Phi}{\int_{ka}^{ka+b} d\Phi}$ diverges for large k, that means that $\frac{\int_{ka-b}^{ka} d\Phi}{\int_{ka}^{\infty} d\Phi}$ diverges as well. □

Proofs

Proof of lemma 1: Write the expected utility difference of playing $p_i = 1$ and playing $p_i = 0$ as[23]

$$-\delta(\tau_i)\Delta + \theta_i/n \tag{6}$$

where $\Delta \in [-1, 1]$ is the difference between the (believed) probability that OP plays A when facing an individual who has played $p_i = 1$ and an individual who has played $p_i = 0$. Clearly, (6) is strictly increasing and continuous in θ_i. As it is optimal to play $p_i = 1$

[23] In principle Δ could depend on the number of individuals choosing $p_i = 1$ in the information aggregation stage. In this case, the expected utilty difference is

$$\sum_{k=1}^{n} \{[-\delta(\tau_i)\Delta(k, k-1) + \theta_i] * prob(k-1)/n\}$$

where $\Delta(k, k-1)$ is the difference between the believed probability that OP plays A when facing an individual who played $p_i = 1$ and k individuals chose 1 and the probability that OP plays A when facing an individual who played $p_i = 0$ and $k - 1$ individuals chose 1. The same argument as below holds: this expression is strictly increasing in θ_i. As will become apparent from (1)– (4), OP's best response strategy will not depend on the number of individuals choosing 1; see the comment in footnote 10.

($p_i = 0$) if (6) is positive (negative), the best response to any given belief is a cutoff strategy where the cutoff is given by the θ_i for which the utility difference above is 0. (Note that the cutoff is necessarily interior as $p_i = 1$ ($p_i = 0$) is dominant for sufficiently high (low) θ_i.) Since all best responses are cutoff strategies, all rationalizable actions are cutoff strategies.

In the privacy case, $\Delta = 0$ by definition and therefore (6) is zero if and only if $\theta_i = 0$. Consequently, $t^p(\tau_i) = 0$. $\square$

Proof of lemma 2: Suppose $v_1 < v_0$ in equilibrium. In this case, (6) is strictly increasing in τ_i as $\Delta < 0$ and therefore $t(\tau_i)$ is strictly decreasing in τ_i.

This implies that we can partition $\mathbb{R}$ in three intervals $(-\infty, t(\bar{\tau})]$, $(t(\bar{\tau}), t(\underline{\tau})]$, $(t(\underline{\tau}), \infty)$. Denoting the inverse of the equilibrium cutoff t by s, we get

$$\begin{aligned}
v_1 &= \frac{\int_{t(\bar{\tau})}^{t(\underline{\tau})} \int_{s(\theta_i)}^{\bar{\tau}} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i) + \int_{t(\underline{\tau})}^{\infty} \int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)}{\int_{t(\bar{\tau})}^{t(\underline{\tau})} \int_{s(\theta_i)}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i) + \int_{t(\underline{\tau})}^{\infty} \int_{\underline{\tau}}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)} \\
&\geq \frac{\int_{t(\bar{\tau})}^{\infty} \int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)}{\int_{t(\bar{\tau})}^{\infty} \int_{\underline{\tau}}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)} \\
&> \frac{\int_{-\infty}^{t(\underline{\tau})} \int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)}{\int_{-\infty}^{t(\underline{\tau})} \int_{\underline{\tau}}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)} \\
&\geq \frac{\int_{-\infty}^{t(\bar{\tau})} \int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i) + \int_{t(\bar{\tau})}^{t(\underline{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)}{\int_{-\infty}^{t(\bar{\tau})} \int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i) + \int_{t(\bar{\tau})}^{t(\underline{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)} \\
&= v_0
\end{aligned}$$

where the inequalities use the assumption that $\Gamma_{\theta_i'}$ first order stochastically dominates $\Gamma_{\theta_i''}$ if $\theta_i' > \theta_i''$ and therefore θ_i and τ_i are positively correlated.[24] The result that $v_0 < v_1$ contradicts our initial supposition and therefore $v_1 \geq v_0$ in all equilibria. $\square$

Proof of proposition 1: Consider (6) which has to be zero if θ_i equals the equilibrium cutoff level. Hence, $t^{np}(\tau_i) = n\Delta\delta(\tau_i)$. By lemma 2, $\Delta \geq 0$ and therefore $t^{np} \geq 0$ with strict inequality if $\Delta > 0$. In an equilibrium of the privacy case $\Delta = 0$ by assumption and $t^p(\tau_i) = 0$, see lemma 1, and therefore $t^{np} \geq t^p$. Furthermore, t^{np} is increasing in τ_i

[24]To be clear, take the first of the inequalities and denote the inverse of t by s:

$$\begin{aligned}
\mathbb{E}[\tau|\theta_i > t(\bar{\tau})] = \frac{\int_{t(\bar{\tau})}^{\infty} \mathbb{E}[\tau|\theta_i] \, d\Phi(\theta_i)}{\int_{t(\bar{\tau})}^{\infty} \int_{\underline{\tau}}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)} &\leq \frac{\int_{t(\bar{\tau})}^{t(\underline{\tau})} \mathbb{E}[\tau|\theta_i] \int_{s(\theta_i)}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\theta_i + \int_{t(\underline{\tau})}^{\infty} \mathbb{E}[\tau|\theta_i] \, d\Phi(\theta_i)}{\int_{t(\bar{\tau})}^{t(\underline{\tau})} \int_{s(\theta_i)}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i) + \int_{t(\underline{\tau})}^{\infty} d\Phi(\theta_i)} \\
&\leq \frac{\int_{t(\bar{\tau})}^{t(\underline{\tau})} \mathbb{E}[\tau|\theta_i, \tau \geq s(\theta_i)] \int_{s(\theta_i)}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\theta_i + \int_{t(\underline{\tau})}^{\infty} \mathbb{E}[\tau|\theta_i] \, d\Phi(\theta_i)}{\int_{t(\bar{\tau})}^{t(\underline{\tau})} \int_{s(\theta_i)}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i) + \int_{t(\underline{\tau})}^{\infty} d\Phi(\theta_i)} = v_1
\end{aligned}$$

where the first inequality holds as $\mathbb{E}[\tau|\theta_i]$ is strictly increasing in θ_i (by the first order stochastic dominance assumption on Γ_{θ_i}) and therefore putting less weight on lower θ_i increases the expectation. The third inequality follows a similar logic and the second one uses that $\mathbb{E}[\tau|\theta_i]$ is strictly increasing in θ_i directly.

as $\delta' \geq 0$ by assumption.

Finally, we show that $\Delta > 0$ whenever the equilibrium strategy of OP is influenced by the presence of privacy. By lemma 2, $\Delta \geq 0$. By assumption, OP plays M in the privacy case. If OP behavior was influenced by the presence of privacy and $\Delta = 0$ then the probability of A has to change in both groups (individuals choosing $p_i = 0$ and individuals choosing $p_i = 1$) by the same amount compared to the privacy case. That is, OP would have to play A with the same positive probability against $p_i = 0$ and $p_i = 1$ in the no privacy case. This can only be optimal if $v_1 \geq 0$ and $v_0 \geq 0$. Furthermore, $\Delta = 0$ implies $t^{np} = 0$ and therefore $v_1 > v_0$ (as θ_i and τ_i are positively correlated by the stochastic dominance assumption on Γ_{θ_i}). Hence, $v_1 > 0$ and $v_0 \geq 0$. But this is incompatible with Bayesian updating and the assumption $\mathbb{E}[\tau_i] \leq 0$. Hence, $\Delta > 0$ whenever the presence of privacy influences OP behavior. $\square$

Proof of proposition 2: We start with the case where OP finds it optimal to play A against all individuals choosing $p_i = 1$ and M against all individuals choosing $p_i = 0$ under both strategies t^{np} and t^p. Recall that OP's payoff is the expected value of τ of all those individuals against which he plays A. Hence, the payoff difference of OP's payoff between the two scenarios is the expected value of τ in the area between the horizontal axis and t^{np} in figure 6 below.

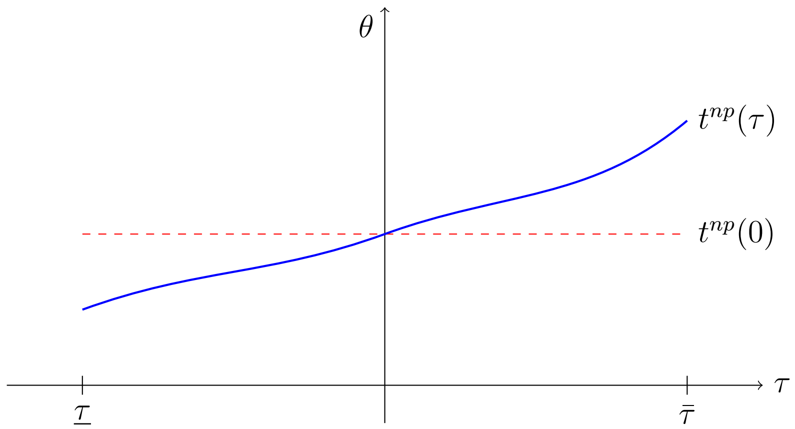

Figure 6: Integration range for difference in OP payoff

Denote the inverse function of $t^{np}(\tau)$ as $s(\theta)$. The difference of OP's payoffs between individuals using t^{np} and t^p is

$$\int_0^{t^{np}(\underline{\tau})} \int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_\theta(\tau) d\Phi(\theta) + \int_{t^{np}(\underline{\tau})}^{t^{np}(\bar{\tau})} \int_{s(\theta)}^{\bar{\tau}} \tau \, d\Gamma_\theta(\tau) d\Phi(\theta)$$
$$= \int_0^{t^{np}(0)} \int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_\theta(\tau) d\Phi(\theta) - \int_{t^{np}(\underline{\tau})}^{t^{np}(0)} \int_{\underline{\tau}}^{s(\theta)} \tau \, d\Gamma_\theta(\tau) d\Phi(\theta) + \int_{t^{np}(0)}^{t^{np}(\bar{\tau})} \int_{s(\theta)}^{\bar{\tau}} \tau \, d\Gamma_\theta(\tau) d\Phi(\theta)$$

where the equality simply splits up the integration range which can be easily visualized in figure 6. The first of the three double integrals is positive by the following argument: As –

by assumption – Γ_0 is symmetric around 0, $\int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_0(\tau) = 0$. It follows that $\int_{\underline{\tau}}^{\bar{\tau}} \tau \, d\Gamma_\theta(\tau) > 0$ for all $\theta > 0$ because Γ_θ first order stochastically dominates Γ_0 for all $\theta > 0$. This implies that the first double integral is positive as $t^{np}(0) \geq 0$ by proposition 1. The second double integral is negative as it integrates only over $\tau \leq 0$ and with the minus sign this second term becomes positive as well. The third double integral is positive as it integrates only over positive τ. Consequently, OP would like to play A against individuals with (τ_i, θ_i) in the area between the horizontal axis and t^{np} which means that OP is better off (given the strategy of playing A if and only if $p_i = 1$) under $t^p(\tau) = 0$ than under t^{np}.

We established that playing A against individuals who play $p_i = 1$ is relatively more attractive if individuals use strategy $t^p(\tau) = 0$ than if they use strategy t^{np}. This implies that whenever OP prefers to play A against individuals who play $p_i = 1$ under t^{np}, the same is true under t^p. Hence, we do not have to consider a case where OP plays M against individuals choosing $p_i = 1$ if they use t^p but A if they use t^{np}. In all other cases, OP uses the same action against individuals choosing $p_i = 0$ and against individuals choosing $p_i = 1$. Hence, $t^p = t^{np}$ and OP's payoffs are the same under both strategies (t^p and t^{np}). □

Proof of lemma 3: As the type draws are independent across individuals and as τ is not payoff relevant in the information aggregation stage, it is clear that the consumer surplus optimal cutoff will be independent of τ.

Write consumer surplus given cutoff t as

$$\begin{aligned}
CS &= \sum_{l=0}^{n} \binom{n}{l} (1-\Phi(t))^l \Phi(t)^{n-l} \left[\frac{l}{n} \left(l\mathbb{E}[\theta|\theta > t] + (n-l)\mathbb{E}[\theta|\theta < t] \right) \right] \\
&= \sum_{l=0}^{n} \binom{n}{l} (1-\Phi(t))^l \Phi(t)^{n-l} \left[\frac{l}{n} \left(\mathbb{E}[\theta|\theta > t](l - (n-l)(1-\Phi(t))/\Phi(t) \right) \right] \\
&= \frac{1}{\Phi(t)} \sum_{l=0}^{n} \binom{n}{l} (1-\Phi(t))^l \Phi(t)^{n-l} \frac{l}{n} \left[(\mathbb{E}[\theta|\theta > t](l - n(1-\Phi(t))) \right] \\
&= \frac{1}{\Phi(t)} \sum_{l=1}^{n-1} \binom{n-1}{l-1} (1-\Phi(t))^l \Phi(t)^{n-l} \left[(\mathbb{E}[\theta|\theta > t](l - n(1-\Phi(t))) \right] \\
&= \frac{1-\Phi(t)}{\Phi(t)} \sum_{k=0}^{n-1} \binom{n-1}{k} (1-\Phi(t))^k \Phi(t)^{n-1-k} \left[(\mathbb{E}[\theta|\theta > t](k+1 - n(1-\Phi(t))) \right] \\
&= \frac{1-\Phi(t)}{\Phi(t)} \left[(\mathbb{E}[\theta|\theta > t]((n-1)(1-\Phi(t)) + 1 - n(1-\Phi(t))) \right] \\
&= \frac{1-\Phi(t)}{\Phi(t)} \left[\mathbb{E}[\theta|\theta > t]\Phi(t) \right] \\
&= \int_t^\infty \theta \, d\Phi(\theta)
\end{aligned}$$

where we use $(1-\Phi(t))\mathbb{E}[\theta|\theta > t] + \Phi(t)\mathbb{E}[\theta|\theta < t] = 0$ – which holds by the law of iterated expectation as $\mathbb{E}[\theta] = 0$ – in the second line. When going to the third but last line, we exploit commonly known properties of the binomial distribution: Its probability mass sums to 1 and the expected value of $n-1$ independent draws from $0, 1$ where 1 has probability $1 - \Phi(t)$ equals $(n-1)(1-\Phi(t))$. From the expression in the last line, it is clear that consumer surplus is maximized by $t = 0$. □

Proof of proposition 3:

1.) Suppose there is a mixed strategy equilibrium in the case without privacy. Then, OP has to play M against both groups with positive probability. If he played A against those who chose $p_i = 1$ for sure and mixed for those who chose $p_i = 0$, then M could not be optimal in the privacy case. Hence, OP can in the case without privacy achieve a payoff equal to his equilibrium payoff by playing M against both groups. Consequently, OP's payoff with and without privacy is the same. Individuals are strictly better off with privacy as (a) there is no chilling effect which means by lemma 3 that expected welfare in the information aggregation stage is maximized and (b) M will be played with probability 1 against them in the interaction stage.

2.) Now assume that $\delta'(\tau) > 0$. We will show that for n sufficiently high the privacy equilibrium welfare dominates the equilibrium in the case without privacy (or the two are identical).

Note that $t^{np\prime}(\tau_i) = n\Delta\delta'(\tau_i)$ and therefore t^{np} is strictly increasing in τ_i and the slope also becomes arbitrarily large as n increases. To economize on notation we will denote t^{np} simply by t in the remainder of the proof.

Denoting the inverse of t by s, we can write

$$v_1 = \frac{\int_{t(\underline{\tau})}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)}{1 - \Phi(t(\bar{\tau})) + \int_{t(\underline{\tau})}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)} + \frac{\mathbb{E}[\tau | \theta_i > t(\bar{\tau})]}{1 + \frac{\int_{t(\underline{\tau})}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)}{1 - \Phi(t(\bar{\tau}))}}.$$

As s becomes arbitrarily flat for n sufficiently high, we can choose – for n high enough – an $\varepsilon > 0$ such that $\int_{t(\bar{\tau})-\varepsilon}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)/(1-\Phi(t(\bar{\tau}))) > 0.5 \int_{t(\bar{\tau})-\varepsilon}^{t(\bar{\tau})} \int_{\underline{\tau}}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)/(1-\Phi(t(\bar{\tau})))$. It follows that the second term in v_1 goes to zero as $n \to \infty$ because $\int_{t(\bar{\tau})-\varepsilon}^{t(\bar{\tau})} \int_{\underline{\tau}}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau) \, d\Phi(\theta_i)/(1 - \Phi(t(\bar{\tau})))$ and therefore its denominator diverges to infinity by lemma 4.

The first term in v_1 converges to something below the unconditional mean of τ which we denote by $\tau^E = \mathbb{E}[\tau]$: For n large, the previous step implies that,

$$\begin{aligned}
v_1 \quad &\approx \quad \frac{\frac{\int_{t(\underline{\tau})}^{t(\tau^E)} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}{\int_{t(\underline{\tau})}^{t(\tau^E)} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)} + \frac{\int_{t(\tau^E)}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}{\int_{t(\underline{\tau})}^{t(\tau^E)} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}}{1 + \frac{\int_{t(\tau^E)}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i) + 1 - \Phi(t(\bar{\tau}))}{\int_{t(\underline{\tau})}^{t(\tau^E)} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}} \\
&\leq \quad \frac{\int_{t(\underline{\tau})}^{t(\tau^E)} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}{\int_{t(\underline{\tau})}^{t(\tau^E)} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)} + \frac{\int_{t(\tau^E)}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}{\int_{t(\underline{\tau})}^{t(\tau^E)} \int_{\underline{\tau}}^{s(\theta_i)} d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}
\end{aligned}$$

Note that the first term equals $\mathbb{E}[\tau_i | t(\underline{\tau}) \leq \theta_i \leq t(\tau^E) \ \wedge \ \tau_i \leq s(\theta_i)]$. Clearly, this is below the unconditional mean τ^E. It follows that for a sufficiently small $\varepsilon' > 0$ (and large n)

$$v_1 \leq \tau^E + \frac{\int_{t(\tau^E)}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}{\int_{t(\tau^E)-\varepsilon'}^{t(\tau^E)} \int_{\underline{\tau}}^{\bar{\tau}} d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}.$$

Note that the same ε' appropriately chosen for some n will also work for higher n (as the density of ϕ thins out for higher θ_i and $t(\tau^E) - t(\underline{\tau})$ is increasing in n). This implies that we can conclude for the limit $n \to \infty$ that

$$v_1 \leq \tau^E + \frac{\int_{t(\tau^E)}^{t(\bar{\tau})} \int_{\underline{\tau}}^{s(\theta_i)} \tau \, d\Gamma_{\theta_i}(\tau)\, d\Phi(\theta_i)}{\int_{t(\tau^E)-\varepsilon'}^{t(\tau^E)} \int_{\underline{\tau}}^{\bar{\tau}} d\Gamma_{\infty}(\tau)\, d\Phi(\theta_i)} \leq \tau^E + \frac{\bar{\tau} \int_{t(\tau^E)}^{t(\bar{\tau})} d\Phi(\theta_i)}{\int_{t(\tau^E)-\varepsilon'}^{t(\tau^E)} d\Phi(\theta_i)} \xrightarrow{n \to \infty} \tau^E$$

where the limit follows from lemma 4 and the above established fact that t goes to infinity as $n \to \infty$. By assumption, OP's best response when facing the unconditional mean τ^E (or a lower τ_i) is M which contradicts the supposition $\Delta = 1$. Hence, $\Delta < 1$ which implies that OP uses a mixed strategy. By the first part of the proposition, privacy then welfare dominates no privacy.

3.) We will show that OP either plays M (independent of p_i) or uses a mixed strategy in the no privacy equilibrium if r is sufficiently high. Result (1) above then implies result (3).

Suppose OP plays a pure strategy in equilibrium. If OP plays M against $p_i = 1$, then – by the assumption that OP plays M in the privacy case – privacy and no privacy case lead to the same equilibrium and the result holds trivially. OP cannot play A against $p_i = 0$: By lemma 2, OP would then also play A against $p_i = 1$. But this is incompatible with Bayesian updating and the assumption that OP plays M in the privacy case. Hence, we only need to consider the case where OP plays M against $p_i = 0$ and A against $p_i = 1$. In this case, $t^{np}(\tau_i) = nr\delta(\tau_i)$ and t^{np} diverges to ∞ as $r \to \infty$. Furthermore, the slope of t^{np} is linearly growing in r. Hence, the derivative of $t^{np}(\tau)$ also diverges to ∞ as r grows. But then the same steps as in the proof of result (2) above imply that $v_1 \leq \tau^E$, i.e. playing A against $p_i = 1$ is not a best response which contradicts that OP uses the pure strategy corresponding to $\Delta = 1$ in the equilibrium without privacy for r sufficiently large. As – for r sufficiently large – OP uses either mixed strategy in the no privacy equilibrium or plays M regardless of p_i, (1) implies that privacy dominates no privacy. $\square$

Proof of proposition 4: First, consider the result for large n. If the equilibrium without privacy is mixed (for large n), then the result is implied by proposition 3. If there is a pure strategy equilibrium with $\Delta = 0$, privacy and no privacy do not differ and the result holds trivially (in a weak sense). We will therefore concentrate on the case where there are arbitrarily high n for which $\Delta = 1$ in equilibrium. Recall that $t_n^{np}(\tau_i) = n\delta(\tau_i)$. Consequently, OP's payoff is bounded from above by $n \int_{n\delta(\underline{\tau})}^{\infty} \bar{\tau}\, d\Phi(\theta) = \bar{\tau}n(1 - \Phi(n\delta(\underline{\tau})))$ as $\delta' \geq 0$. By L'Hopital's rule, this upper bound converges to zero as $n \to \infty$. That is, OP payoffs in equilibrium are arbitrarily close to OP payoffs with privacy (which are zero) for n sufficiently high. Consumer surplus from information aggregation was derived – for a constant cutoff t – in the proof of lemma 3 and equals $\int_t^{\infty} \theta\, d\Phi(\theta)$. Consequently, an upper bound on consumer surplus in the information aggregation stage without privacy is $\int_{n\delta(\underline{\tau})}^{\infty} \theta\, d\Phi(\theta)$. This converges to zero as well as $n \to \infty$. Hence, consumer surplus from information aggregation is strictly higher with privacy than without for n sufficiently large (as the privacy consumer surplus is $\int_0^{\infty} \theta\, d\Phi(\theta) > 0$). Since expected consumer surplus from interaction is 0 in the privacy case but strictly negative without privacy (given $\Delta = 1$), welfare is higher with privacy than without for n sufficiently large.

Concerning large r,notice that $t^{np} = nr\delta(\tau_i)$ (given that $\Delta = 1$) also diverges to infinity as $r \to \infty$. The same arguments as in the previous paragraph establish the welfare optimality of privacy. $\square$

Proof of proposition 5: First consider $\lambda = 0$. Note that the distribution of τ_i under $\bar{\tau}$ is the same as the distribution of τ_i that OP faces in the privacy case of the original model (with distribution Γ_{θ_i}). As we assumed that OP plays M in the privacy equilibrium, it is clear that the privacy equilibrium is also an equilibrium for $\lambda = 0$. In fact, it is the unique equilibrium: Since M is the best response against the distribution $\bar{\Gamma}$ by assumption, OP has to play M for sure against at least one group of individuals (either those choosing $p_i = 0$ or those choosing $p_i = 1$) by Bayesian updating. Suppose OP played A with positive probability against those who chose $p_i = 1$. Then some individuals with low θ_i would be chilled and play $p_i = 0$. As δ is increasing in τ_i, the best response cutoff would be increasing in τ_i, see (5). But then the average τ_i among those choosing $p_i = 1$ is lower than the average τ_i under $\bar{\Gamma}$. Consequently, M is a strict best response by OP because M is a best response against $\bar{\Gamma}$. This contradicts that OP plays A with positive probability.

Note that $\mathbb{E}[\tau_i|\theta_i \geq 0]$ is continuous in λ. Since M is a best response against $\bar{\Gamma}$, that is $\mathbb{E}[\tau_i|\theta_i \geq 0] < 0$ for $\lambda = 0$, the same is true for sufficiently small $\lambda > 0$. Hence, a $\underline{\lambda} > 0$ exists such that for all $\lambda \leq \underline{\lambda}$ the unique equilibrium without privacy is that OP plays M and all individuals use a cutoff of zero. This is equivalent to the privacy equilibrium and therefore privacy and no privacy are welfare equivalent for all $\lambda \leq \underline{\lambda}$. For the result in the proposition, let $\underline{\lambda}$ be the highest λ such that the equilibrium in the no privacy is that OP plays M against individuals choosing $p_i = 1$. Note that $\underline{\lambda} < 1$ as by assumption OP plays A against individuals choosing $p_i = 1$ for $\lambda = 1$.

For $\lambda = 1$, the equilibrium of the no privacy case was assumed to be that OP plays A (M) against $p_i = 0$ ($p_i = 1$) in the no privacy case. Denote by λ^* the infimum of all λ for which such an equilibrium exists. Clearly, $\lambda^* \in (\underline{\lambda}, 1)$. Since such an equilibrium no longer exists for $\lambda < \lambda^*$, it has to hold true that at $\lambda = \lambda^*$ OP is indifferent between playing A and playing M against those playing $p_i = 1$ (given that individuals use $t^{np} = n\delta(\tau_i)$). (For lower λ OP will then prefer to play M as the correlation is too weak and that is why the equilibrium breaks down.) Note that the best response cutoffs of the individuals do not depend on λ but only on OP's strategy. It follows that $\mathbb{E}[\tau_i|\theta_i \geq t^{np}(\tau_i)]$ is continuous in λ for $\lambda \geq \lambda^*$. As OP is indifferent at λ^*, we have $\mathbb{E}[\tau_i|\theta_i \geq t^{np}(\tau_i)] = 0$ at λ^*. Continuity, implies that $\mathbb{E}[\tau_i|\theta_i \geq t^{np}(\tau_i)]$ is arbitrarily small for λ close but strictly above λ^*. That is, for any $\varepsilon > 0$ there is a $\varepsilon' > 0$ such that imposing privacy leads only to less than ε losses for OP if $\lambda < \lambda^* + \varepsilon'$. Imposing privacy leads (for $\lambda \in [\lambda^*, \lambda^* + \varepsilon']$) to a discrete increase in citizen welfare for several reasons: First, those choosing $p_i = 1$ no longer face the aggressive response which increases their payoff by $\delta(\tau_i)$. Second, in the privacy case individuals use the cutoff zero instead of $t^{np} > 0$ which leads to a higher surplus in the information aggregation stage. This implies that for $\varepsilon' > 0$ small enough, privacy welfare dominates no privacy for $\lambda \in (\lambda^*, \lambda^* + \varepsilon']$. Let $\bar{\lambda} = \lambda^* + \varepsilon'$. Note that for $\lambda \in (\underline{\lambda}, \lambda^*)$ the equilibrium in the no privacy case is necessarily mixed which means implies that privacy is Pareto and therefore utilitarian welfare dominant also for these λ, see proposition 3. This establishes the claim. □

Proof of proposition 6: With δ being constant, $t^{np} = n\delta$ in a pure strategy equilibrium with $\Delta = 1$. OP's payoff is

$$n \int_{n\delta}^{\infty} \int_{\underline{\tau}}^{\bar{\tau}} \tau_i \, d\Gamma_{\theta_i} \, d\Phi(\theta_i)$$

which is clearly decreasing in δ. Furthermore, this payoff is higher if correlation is higher as then $\int_{\underline{\tau}}^{\bar{\tau}} \tau_i \, d\Gamma_{\theta_i}$ is higher for every $\theta_i \geq n\delta$.

We now turn to the expected payoff of the individuals without privacy for constant threshold (denoted by t^{np} for brevity). Using the same steps as in the proof of 3 (but adding the expected disutility of being treated aggressively), we get

$$\begin{aligned} CS^{np} &= \sum_{l=0}^{n} \binom{n}{l} (1 - \Phi(t))^l \Phi(t)^{n-l} \left[\frac{l}{n} \left(l\mathbb{E}[\theta|\theta > t^{np}] + (n-l)\mathbb{E}[\theta|\theta < t^{np}] \right) - l\delta \right] \\ &= \int_{t^{np}}^{\infty} \theta \, d\Phi(\theta) - (1 - \Phi(t^{np}))n\delta = \int_{n\delta}^{\infty} \theta \, d\Phi(\theta) - (1 - \Phi(n\delta))n\delta. \end{aligned}$$

Therefore CS^{np} is decreasing in δ:

$$\frac{d\,CS^{np}}{\partial\delta} = -n^2\delta\phi(n\delta) - n(1-\Phi(n\delta)) + n^2\delta\phi(n\delta) = -n(1-\Phi(n\delta)) < 0.$$

Note that the distribution Γ_{θ_i} does not play a role for consumer surplus (given that δ is constant). Furthermore, neither δ nor Γ plays a role in the privacy equilibrium. Taking the effects of OP payoff and consumer surplus together yields the result in the proposition. $\square$

References

Acquisti, A. (2010). The economics of personal data and the economics of privacy. Background paper 3, OECD WPISP-WPIE Roundtable.

Acquisti, A., C. R. Taylor, and L. Wagman (2015). The economics of privacy. *Journal of Economic Literature 54*(2), 442–492.

Acquisti, A. and H. R. Varian (2005). Conditioning prices on purchase history. *Marketing Science 24*(3), 367–381.

Ali, S. N. and R. Bénabou (2017). Image versus information. mimeo.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter and A. Rees (Eds.), *Discrimination in Labor Markets.* Princeton, NJ: Princeton University Press.

Banks, J. S. and J. Sobel (1987). Equilibrium selection in signaling games. *Econometrica 55*(3), 647–661.

Bischoping, K. and H. Schuman (1992). Pens and polls in Nicaragua: An analysis of the 1990 preelection surveys. *American Journal of Political Science 36*(2), 331–350.

Calzolari, G. and A. Pavan (2006). On the optimality of privacy in sequential contracting. *Journal of Economic Theory 130*(1), 168–204.

Cummings, R., K. Ligett, M. M. Pai, and A. Roth (2015). The strange case of privacy in equilibrium models. *arXiv preprint abs/1508.03080*, 1–21.

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

Friedman, M. (1962). *Capitalism and Freedom.* Chicago, IL: University of Chicago Press.

Greenwald, G. (2014). *No place to hide: Edward Snowden, the NSA, and the US surveillance state.* London, UK: Macmillan.

Hamburger, T. and P. Wallsten (2005, July 24). Parties are tracking your habits. *Los Angeles Times*.

Hermalin, B. E. and M. L. Katz (2006). Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics 4*(3), 209–239.

Hirshleifer, J. (1971). The private and social value of information and the reward to incentive activity. *American Economic Review 61*(4), 561–574.

Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics 37*(5), 1137–1153.

Meade, E. E. and D. Stasavage (2008). Publicity of debate and the incentive to dissent: Evidence from the US federal reserve. *Economic Journal 118*(528), 695–717.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry 1*(2), 108–141.

Morris, S. (2001). Political correctness. *Journal of Political Economy 109*(2), 231–265.

Ottaviani, M. and P. N. Sørensen (2006). Reputational cheap talk. *RAND Journal of Economics 37*(1), 155–175.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review 62*(4), 659–661.

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

Prat, A. (2005). The wrong kind of transparency. *American Economic Review 95*(3), 862–877.

Robinson, D., H. Yu, and A. Rieke (2014). Civil rights, big data, and our algorithmic future: A September 2014 report on social justice and technology. Technical report, Upturn.

Schelling, T. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schneier, B. (2006, May 18). The eternal value of privacy. https://www.schneier.com/essays/archives/2006/05/the_eternal_value_of.html.

Solove, D. J. (2010). *Understanding Privacy*. Cambridge, MA: Harvard University Press.

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