Supplementary material to An Informational Theory of Privacy

Ole Jann and Christoph Schottmüller

August 23, 2018

1. Welfare result with utilitarian welfare

Concentrating on utilitarian welfare we can derive a result similar to proposition 3 without imposing the assumption $\delta' > 0$.

Proposition 6. *(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.*

Proof of proposition 6: 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. □

2. 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.

2.1. A 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 1. *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 predictive algorithm that suggests other items based on what people usually buy together with the scale are almost all drug-related.*[1] *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.*

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

2.2. State matching

In this section, we consider a model where the private information of citizens in the information aggregation stage is not directly their personal payoff of policy $p = 1$. Instead citizens have all the same payoff of policy $p = 1$ but each citizen only receives a noisy signal of this payoff. This has a striking implication: Chilling makes every citizen worse off. The reason is that chilling inhibits information aggregation. In the main paper citizens have private preferences over outcomes and therefore some citizens (those with negative θ_i) gain from chilling. Since all citizens have the same interest – implementing the policy if and only if the common payoff consquence is positive, – everyone looses in this setup from chilling.

More precisely, the setting is as follows: The state of the world θ is distributed standard normally and this θ is the payoff consequence of policy $p = 1$ for each citizen. However, the realization of θ is unknown. Each citizen 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 of citizen i, τ_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 the model of the main paper. That is, without privacy a strategy for OP states which of the two actions A and M OP plays against a citizen who chose $p_i = 0$ and which against a citizen who chose $p_i = 1$. With privacy, OP only decides which of the two actions he chooses against all citizens. This means that – to keep the setting comparable to the main paper – we do not consider strategies (or beliefs) that are contingent upon the number of citizens 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 citizens' p_is. This is, for example, the case if the interaction is between i and an agent representing OP and p_i is only learned in the interaction. OP then has to instruct the agent in advance how to act.

The main change is, therefore, that citizen i's payoff is $\theta p - \mathbb{1}_{s(p_i)=A}\delta(\tau_i)$; that is, θ instead of θ_i enters the utility function. Again, we assume that the probability of $p = 1$ is $q(m/n) = m/n$.

We first replicate some intermediary results from the main text in this modified setting.

Lemma 6. *For citizens, only cutoff strategies $t(\tau_i)$ are rationalizable. In the privacy case, the optimal cutoff is $t^p(\tau_i) = 0$ for all τ_i.*

Proof. If citizen i receives signal θ_i, he updates his belief α about θ according to Bayes' rule yielding

$$\alpha(\theta'|\theta_i) = prob(\theta \leq \theta'|\theta_i) = \frac{\int_{-\infty}^{\theta'} \tilde{\phi}_\theta(\theta_i)\, d\Phi(\theta)}{\int_{\mathbb{R}} \tilde{\phi}_\theta(\theta_i)\, d\Phi(\theta)}.$$

From the normality assumptions, it follows that the pdf of the belief is single peaked with its peak between 0 (the mean of the prior) and θ_i. Furthermore, $\mathbb{E}[\theta|\theta_i] = \int_{\mathbb{R}} \theta \, d\alpha(\theta|\theta_i)$ is strictly increasing in θ_i with limits $\lim_{\theta_i \to \infty} = \infty$ and $\lim_{\theta_i \to -\infty} = -\infty$. To see this, note that

$$\begin{aligned}
\mathbb{E}[\theta|\theta_i] &= \int_{\mathbb{R}} \theta \frac{\tilde{\phi}_\theta(\theta_i)\phi(\theta)}{\int_{\mathbb{R}} \tilde{\phi}_{\hat{\theta}}(\theta_i)\, d\Phi(\hat{\theta})}\, d\theta \\
&= \frac{\int_{\mathbb{R}} \theta e^{-(\theta_i-\theta)^2/(2\sigma^2)} e^{-\theta^2/2} d\theta}{\int_{\mathbb{R}} e^{-(\theta_i-\theta)^2/(2\sigma^2)} e^{-\theta^2/2} d\theta} \\
&= \frac{\theta e^{-(-2\theta_i\theta+\theta^2(1+\sigma^2))/(2\sigma^2)} d\theta}{\int_{\mathbb{R}} e^{-(-2\theta_i\theta+\theta^2(1+\sigma^2))/(2\sigma^2)} d\theta} \\
&= \frac{\frac{1}{\sqrt{2\pi}\sigma/(\sqrt{1+\sigma^2})} \int_{\mathbb{R}} \theta e^{-\frac{\theta_i^2/(1+\sigma^2)^2 - 2\theta_i\theta/(1+\sigma^2)+\theta^2}{2\sigma^2/(1+\sigma^2)}}\, d\theta}{\frac{1}{\sqrt{2\pi}\sigma/(\sqrt{1+\sigma^2})} \int_{\mathbb{R}} e^{-\frac{\theta_i^2/(1+\sigma^2)^2 - 2\theta_i\theta/(1+\sigma^2)+\theta^2}{2\sigma^2/(1+\sigma^2)}}\, d\theta} \\
&= \frac{\theta_i}{1+\sigma^2}
\end{aligned}$$

where the last equality holds as the numerator of the second but last line is the expected value of a random variable distributed $N(\theta_i/(1+\sigma^2), \sigma^2/(1+\sigma^2)^2)$ and the denominator of the second but last line is simply 1 (as it integrates over the density of this random variable).

Citizen i's expected payoff difference between choosing $p_i = 1$ and $p_i = 0$ is[2]

$$-\delta(\tau_i)\Delta + \mathbb{E}[\theta|\theta_i]/n = -\delta(\tau_i)\Delta + \frac{\theta_i}{(1+\sigma^2)n} \tag{1}$$

where Δ is again the difference between the probabilities that OP plays A against citizens with $p_i = 1$ and citizens with $p_i = 0$. Clearly, it is optimal to play $p_i = 0$ ($p_i = 1$) for sufficiently low (high) θ_i. (Note that $max_{\tau_i \in [\underline{\tau}, \bar{\tau}]} \delta(\tau_i)$ is bounded.) Furthermore, $\mathbb{E}[\theta|\theta_i]$ is strictly increasing in θ_i which implies that i's best response is a cutoff strategy. Consequently, only cutoff strategies are best responses. The optimal cutoff is given by $t(\tau_i) = (1+\sigma^2)n\delta(\tau_i)\Delta$.

In the privacy case, $\Delta = 0$ and therefore the optimal cutoff is $t^p(\tau_i) = 0$. □

OP's belief over τ_i given p_i is given by

$$\begin{aligned}
\beta_0(\tau') = prob(\tau \leq \tau' | p_i = 0) &= \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\underline{\tau}}^{\tau'} \mathbb{1}_{t(\tau_i) \geq \theta_i}\, d\Gamma_{\theta_i}(\tau_i)\, d\tilde{\Phi}_\theta(\theta_i)\, d\Phi(\theta)}{\int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\underline{\tau}}^{\bar{\tau}} \mathbb{1}_{t(\tau_i) \geq \theta_i}\, d\Gamma_{\theta_i}(\tau_i)\, d\tilde{\Phi}_\theta(\theta_i)\, d\Phi(\theta)} \\
\beta_1(\tau') = prob(\tau \leq \tau' | p_i = 1) &= \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\underline{\tau}}^{\tau'} \mathbb{1}_{t(\tau_i) \leq \theta_i}\, d\Gamma_{\theta_i}(\tau_i)\, d\tilde{\Phi}_\theta(\theta_i)\, d\Phi(\theta)}{\int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\underline{\tau}}^{\bar{\tau}} \mathbb{1}_{t(\tau_i) \leq \theta_i}\, d\Gamma_{\theta_i}(\tau_i)\, d\tilde{\Phi}_\theta(\theta_i)\, d\Phi(\theta)}.
\end{aligned}$$

[2]Recall that $q(m/n) = m/n$ which means that i's "influence" on the policy decision is $1/n$.

OP's expected utility of playing A against a citizen choosing policy $p_i = 0$ or $p_i = 1$ are then

$$v_0 = \int_{\mathbb{R}} \tau \; d\beta_0(\tau)$$
$$v_1 = \int_{\mathbb{R}} \tau \; d\beta_1(\tau).$$

Lemma 7. *In every perfect Bayesian equilibrium (without privacy), $v_1 \geq v_0$.*

Proof. Suppose otherwise. Then $\Delta < 0$ which implies that $t(\tau_i)$ is decreasing. Denote the inverse of t by z. From OP's point of view θ_i is distributed according to the cdf

$$\hat{\Phi}(\theta_i) = \int_{\mathbb{R}} \tilde{\Phi}_\theta(\theta_i) \; d\Phi(\theta).$$

Using this distribution $\hat{\Phi}$ we can replicate the proof from the main paper one-to-one:

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

which contradicts our starting point $v_1 < v_0$. □

The previous result implies that $\Delta \geq 0$ and therefore $t^{np}(\tau_i) = (1 + \sigma^2) n \delta(\tau_i) \Delta \geq 0 = t^p(\tau_i)$. We therefore get chilling.

Proposition 7. *The equilibrium cutoff of a type τ_i is higher without privacy than with privacy. If the absence of privacy affects OP's behavior, this relation is strict. The difference of equilibrium cutoffs with and without privacy is increasing in τ_i.*

To establish that this chilling indeed hurts every citizen – as we claimed above – we have to show that the privacy cutoff zero leads to a higher expected consumer surplus than $t^{np}(\tau) > 0$.

Lemma 8. *The cutoff strategy $t^p(\tau) = 0$, i.e. the equilibrium strategy of the privacy case, gives a higher expected consumer surplus in the information aggregation stage than any other $t^{np}(\tau) > 0$.*

Proof. Let $t(\tau)$ be the strategy maximizing expected consumer welfare. Consider citizen i with type $(\theta_i, \tau_i) = (t(\tau'), \tau')$ for some $\tau' \in [\underline{\tau}, \bar{\tau}]$.

Optimality of t requires that expected welfare conditional on i being of type $(t(\tau'), \tau')$ is the same no matter whether i chooses $p_i = 0$ or $p_i = 1$: If this was not the case, say for concreteness $p_i = 1$ lead to a higher expected consumer welfare, then t could not be optimal: As the setup is continuous, it would then also be better for expected consumer surplus if i chose $p_i = 1$ if he was any type in an $\varepsilon > 0$ neighborhood of $(t(\tau'), \tau')$. But as expected welfare is just the expectation of expected welfare conditional on i's type over i's type we get that an alternative strategy t' which is slightly lower than t around τ' leads to higher expected consumer welfare than t. This contradicts the definition of t. Consequently, expected welfare conditional on i being of type $(t(\tau'), \tau')$ has to be the same no matter whether i chooses $p_i = 0$ or $p_i = 1$.

We are now going to show that the just stated (necessary) optimality condition cannot be satisfied for any $t > 0$. However, it is trivially satisfied for t^p by the symmetry of the setup. We focus on citizen i with type $\theta_i = t(\tau_i) > 0$. If citizen i chooses $p_i = 1$ instead of $p_i = 0$, he will increase the probability that $p = 1$ by $1/n$. This can be interpreted as follows: choosing $p_i = 1$ instead of $p_i = 0$ leads with probability $1/n$ to a payoff of θ instead of a payoff of zero (for each citizen). Hence, choosing $p_i = 1$ is best for expected consumer welfare (conditional on i's type) if $\mathbb{E}[\theta|\theta_i] > 0$.[3] As we showed above, $\mathbb{E}[\theta|\theta_i] = \theta_i/(1+\sigma^2)$ which is strictly positive for all $\theta_i > 0$. It follows that $p_i = 1$ leads to strictly higher expected consumer welfare than $p_i = 0$ as $\theta_i > 0$. This contradicts that $t > 0$ maximizes expected consumer surplus. □

The previous results can now be used to obtain a stronger version of our welfare result in the paper. While the paper argued that expected aggregated consumer surplus is higher under privacy if n is large (while OP's payoff is the same with and without privacy), we can now say that the expected utility of each citizen – regardless of his type (θ_i, τ_i) – is higher under privacy for n large. That is, privacy is an interim Pareto improvement here while it was only an ex ante Pareto improvement in the model of the paper.

Proposition 8. *Assume OP plays M in the privacy equilibrium.*
1.) If OP uses a mixed – that is not pure – strategy in the equilibrium without privacy, then changing to the privacy case increases expected welfare at the interim stage.
2.) Assume that (i) δ is differentiable and strictly increasing in τ, i.e. $\delta'(\tau) > 0$ for all $\tau \in [\underline{\tau}, \bar{\tau}]$ and (ii) $\Gamma_\infty = \lim_{\theta_i \to \infty} \Gamma_{\theta_i}$ is a non-degenerate distribution in the sense that $\Gamma_\infty(\tau_i) > 0$ for all $\tau_i > \underline{\tau}$. Then, privacy welfare dominates no privacy for large n in the following sense: Compared to the no privacy case, privacy leads to a higher expected consumer surplus for each consumer of every type and the same expected payoff for OP if n is sufficiently large.

[3] Note that conditioning on τ_i is immaterial as τ_i is – given θ_i – not correlated with θ.

In order to prove the proposition, we have to first restate the technical result on the limit of tails of Φ that we show in the appendix for $\hat{\Phi}(\theta_i)$.

Lemma 9. *Let $\hat{\Phi}(\theta_i) = \int_{\mathbb{R}} \tilde{\Phi}_\theta(\theta_i)\, d\Phi(\theta)$ be the distribution of θ_i from OP's perspective. Then, $\int_{ka-b}^{ka} d\hat{\Phi} / \int_{ka}^{\infty} d\hat{\Phi}$ diverges to infinity as $k \to \infty$ for $a, b > 0$.*

Proof. If we can show that $\hat{\phi}(x)/\hat{\phi}(x+b)$ diverges to infinity as $x \to \infty$ (where $\hat{\phi}$ is the density of $\hat{\Phi}$), then the same proof as in the paper applies. Note that

$$\hat{\phi}(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}} d\Phi(\theta) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}} e^{-\theta^2/2}\, d\theta$$

$$\begin{aligned}
\frac{\hat{\phi}(x)}{\hat{\phi}(x+b)} &= \frac{\int_{\mathbb{R}} e^{-(x-\theta)^2/(2\sigma^2)}\, d\Phi(\theta)}{\int_{\mathbb{R}} e^{-(x+b-\theta)^2/(2\sigma^2)}\, d\Phi(\theta)} \\
&= \frac{\int_{\mathbb{R}} e^{-(x-\theta)^2/(2\sigma^2) - \theta^2/2}\, d\theta}{\int_{\mathbb{R}} e^{-(x+b-\theta)^2/(2\sigma^2) - \theta^2/2}\, d\theta} \\
&= \frac{\int_{\mathbb{R}} e^{[-(1+\sigma^2)\theta^2 + 2x\theta]/(2\sigma^2)}\, d\theta}{\int_{\mathbb{R}} e^{[-2b(x-\theta) - b^2 - (1+\sigma^2)\theta^2 + 2x\theta]/(2\sigma^2)}\, d\theta} \\
&= \frac{\int_{\mathbb{R}} e^{-\frac{(\theta - x/(1+\sigma^2))^2}{2\sigma^2/(1+\sigma^2)}}\, d\theta}{\int_{\mathbb{R}} e^{(-2b(x-\theta)-b^2)/(2\sigma^2)} e^{-\frac{(\theta - x/(1+\sigma^2))^2}{2\sigma^2/(1+\sigma^2)}}\, d\theta} \\
&= \frac{\int_{\mathbb{R}} d\bar{\Phi}(\theta)}{\int_{\mathbb{R}} e^{(-2b(x-\theta)-b^2)/(2\sigma^2)}\, d\bar{\Phi}(\theta)}
\end{aligned}$$

where $\bar{\Phi}$ is the cdf of a normal distribution with mean $x/(1+\sigma^2)$ and variance $\sigma^2/(1+\sigma^2)$. As the numerator is 1, the previous expression can be written as

$$\frac{\hat{\phi}(x)}{\hat{\phi}(x+b)} = \frac{1}{\int_{\mathbb{R}} e^{-\frac{2b(x-\theta)+b^2}{2\sigma^2}}\, d\bar{\Phi}(\theta)}$$

which diverges to infinity as $x \to \infty$ (because the denominator converges to zero). Given this, the rest of the proof from the main paper goes through one-to-one which implies the lemma. □

Proof of proposition 8: Let M be optimal for OP in the privacy equilibrium.

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. Citizens are strictly better off with privacy as (a) there is no chilling effect which means by lemma 8 that expected welfare of

every consumer (no matter which type) 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).

Now recall that $t^{np}(\tau_i) = (1+\sigma^2)n\delta(\tau_i)\Delta$. Consequently, the threshold values become arbitrarily large as n gets large (assuming $\Delta = 1$). Note also that t is increasing in τ and the slope also becomes arbitrarily large as n increases. From here, the proof of the main paper applies with $\hat{\Phi}$ in place of Φ. □

Proposition 6 and its proof go through without change. Proposition 4 of the paper holds true with slightly changed proof and is therefore restated here. Note that the definition of "increasing in correlation" is as in the paper and also we concentrate on the interesting case where there is a pure strategy equilibrium in the case without privacy.

Proposition 9 (Monotone welfare difference)**.** *Assume $\delta(\tau_i)$ is constant. The welfare difference between no privacy and privacy is decreasing in δ and increasing in the correlation in Γ.*

Proof. With δ being constant, $t^{np} = n(1+\sigma^2)\delta$ in a pure strategy equilibrium with $\Delta = 1$. OP's payoff is

$$n \int_{n(1+\sigma^2)\delta}^{\infty} \int_{\underline{\tau}}^{\bar{\tau}} \tau_i \, d\Gamma_{\theta_i} \, d\hat{\Phi}(\theta_i)$$

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 $\theta_i \geq n(1+\sigma^2)\delta$.

We now turn to expected consumer surplus. Note that consumer surplus in the privacy case depend neither on δ nor on the correlation between θ_i and τ_i. We can therefore concentrate on the case without privacy (and will again focus on the interesting case of a pure strategy equilibrium with $\Delta = 1$). Consumer surplus can then be written as

$$\begin{aligned} CS^{np} &= \int_{\mathbb{R}} \sum_{l=0}^{n} \left[(1 - \tilde{\Phi}_\theta(n(1+\sigma^2)\delta))^l \tilde{\Phi}_\theta(n(1+\sigma^2)\delta)^{n-l} \left(\frac{l}{n} n\theta - l\delta \right) \right] d\Phi(\theta) \\ &= \int_{\mathbb{R}} (\theta - \delta) \left(1 - \tilde{\Phi}_\theta(n(1+\sigma^2)\delta) \right) n \, d\Phi(\theta). \end{aligned}$$

This shows that CS^{np} does not depend on the correlation between θ_i and τ_i. Taking the

derivative with respect to δ (and using $t = n(1+\sigma^2)\delta$ to save space) gives

$$\begin{aligned}
\frac{d\,CS^{np}}{d\,\delta} &= n\int_{\mathbb{R}} -(\theta-\delta)\tilde{\phi}_\theta(n(1+\sigma^2)\delta)n(1+\sigma^2) - \left(1-\tilde{\Phi}_\theta(n(1+\sigma^2)\delta)\right)\, d\Phi(\theta) \\
&= K\int_{\mathbb{R}} -(\theta-\delta)e^{-\frac{(t-\theta)^2}{2\sigma^2}}e^{-\frac{\theta^2}{2}}\, d\theta - n\int_{\mathbb{R}}(1-\tilde{\Phi}_\theta(t))\, d\Phi(\theta) \\
&= Ke^{-\frac{t^2}{2(1+\sigma^2)}}\int_{\mathbb{R}} -(\theta-\delta)e^{-\frac{(\theta-t/(1+\sigma^2))^2}{2\sigma^2/(1+\sigma^2)}}\, d\theta - n\int_{\mathbb{R}}(1-\tilde{\Phi}_\theta(t))\, d\Phi(\theta) \\
&= Ke^{-\frac{t^2}{2(1+\sigma^2)}}\left(-\frac{t}{1+\sigma^2}+\delta\right) - n\int_{\mathbb{R}}(1-\tilde{\Phi}_\theta(t))\, d\Phi(\theta) \\
&= Ke^{-\frac{t^2}{2(1+\sigma^2)}}\left(-(n-1)\delta\right) - n\int_{\mathbb{R}}(1-\tilde{\Phi}_\theta(t))\, d\Phi(\theta) < 0
\end{aligned}$$

where we used the short hand notation $K = n/\sqrt{4\sigma^2\pi^2} > 0$. Hence, CS^{np} is decreasing in δ. Taking the results together gives the proposition. □

3. Extension: Abstention

In this section we show that the same chilling effects as in the paper also occur if we consider a referendum like setting in the first stage in which citizens can either vote for a policy, against it *or abstain*. We assume that the policy "1" is implemented with probability $q(m_0, m_1, n) = 0.5 + (m_1 - m_0)/(2m_0 + 2m_1)$. Instead of interpreting the setup as the question whether a certain policy should be implemented one can also interpret it as a probabilistic election between two candidates. For simplicity, we make the technical assumption that was used in the main text at some points that $\Gamma_0(\tau_i)$ is symmetric around zero which implies that $\mathbb{E}[\tau_i|\theta_i = 0]$. Everything else is as in the main paper.

We will replicate the results in section 2 of the paper with emphasis on the things that are different. For proofs that are identical to those in the appendix of the main paper we will simply refer to the paper.

Lemma 10. *Only cutoff strategies are rationalizable for citizens, i.e. each citizen will choose two cutoffs $t_0(\tau_i)$ and $t_1(\tau_i)$ and play $p_i = 0$ if $\theta_i < t_0(\tau_i)$ and $p_i = 1$ if $\theta_i \geq t_1(\tau_i)$. In the privacy case, the optimal cutoffs are $t_0^p(\tau_i) = t_1^p(\tau_i) = 0$.*

Proof of lemma 10. As shown in the paper $p_i = 1$ ($p_i = 0$) is dominant for high (low) θ_i. If OP plays A against citizens abstaining with higher probability than against the 0 and 1 voters, then it is clear that no citizen will abstain and the analysis in the paper applies. A citizen prefers $p_i = 1$ to $p_i = a$ if and only if

$$-\delta(\tau_i)\Delta_{1a} + \theta_i/n \tag{2}$$

is positive where $\Delta_{1a} \in [-1, 1]$ is the difference between the (believed) probability that OP

plays A when facing a citizen who has played $p_i = 1$ and a citizen who has played $p_i = a$. Clearly, (2) is strictly increasing and continuous in θ_i. Which means that for a given τ_i citizen i prefers $p_i = 1$ over $p_i = a$ if and only if θ_i is above a certain threshold $\tilde{t}_1(\tau_i)$. Similarly, citizen i prefers $p_i = a$ over $p_i = 0$ if and only if θ_i above a certain threshold $\tilde{t}_0(\tau_i)$. If $\tilde{t}_0(\tau_i) > \tilde{t}_1(\tau_i)$, then citizen i will not abstain for any θ_i and for the purpose of lemma 10 we can then take $t_0(\tau_i) = t_1(\tau_i)$ which will then be the θ_i for which citizen i is indifferent between $p_i = 0$ and $p_i = 1$ (similar to the paper). If $\tilde{t}_0(\tau_i) \leq \tilde{t}_1(\tau_i)$, then we can use $t_0(\tau_i) = \tilde{t}_0(\tau_i)$ and $t_1(\tau_i) = \tilde{t}_1(\tau_i)$. In the case of privacy, $\Delta_{1a} = \Delta_{a0} = \Delta_{10} = 0$ and it is straightforward that $t_0^p(\tau_i) = t_1^p(\tau_i) = 0$. □

Lemma 11. *Consider the no privacy case. In every perfect Bayesian equilibrium, OP plays A with weakly higher probability against citizens choosing $p_i = 1$ than against citizens choosing $p_i = a$ and OP plays A with weakly higher probability against citizens choosing $p_i = a$ than against citizens choosing $p_i = 0$.*

Proof of lemma 11. If OP plays A against citizens abstaining with higher probability than against the 0 and 1 voters, then it is clear that no citizen will abstain and the analysis in the paper applies.

First, we show that $v_a \geq v_0$ (assuming that citizens abstain in equilibrium with strictly positive probability). Suppose otherwise. Then – given that OP plays best response – $\Delta_{a0} \leq 0$. By (2) (with Δ_{a0} instead of Δ_{1a}), we have then $t_0 < 0$ and – by the implicit function theorem – $t_0' \leq 0$ as $\delta' \geq 0$. Note that by the assumption $\mathbb{E}[\tau_i|\theta_i = 0] = 0$ and the assumption on first order stochastic dominance of Γ_{θ_i}, we have $\int_{\underline{\tau}}^{\bar{\tau}} \tau_i \, d\Gamma_{\theta_i} < 0$ for all $\theta_i < 0$. This implies that $v_0 < 0$ as $t_0 < 0$ and $t_0' \leq 0$. Hence, playing M against $p_i = 0$ is the best response of OP but this implies that $\Delta_{a0} \leq 0$ can only hold with equality, i.e. OP plays M against both those that abstain and those that choose $p_i = 0$ which is in line with the lemma. If $v_a > v_0$ the lemma holds by OP playing best response. Note that this paragraph implies that $\Delta_{a0} \geq 0$ which implies $t_0 \geq 0$.

Second, we show that OP plays A with at least as high probability against $p_i = 1$ as against $p_i = a$ (assuming that citizens abstain in equilibrium with strictly positive probability). Suppose otherwise. Then $\Delta_{1a} \leq 0$. Given that (2) has to be 0 at t_1, this implies $t_1 \leq 0$ and $t_1' \leq 0$. But then $t_1 \leq t_0$ which contradicts that citizens of some types choose $p_i = a$. □

This implies that in equilibrium we have $0 \leq t_0^{np} \leq t_1^{np}$ and both t_0^{np} and t_1^{np} are increasing (as δ is increasing). This implies chilling.

Proposition 10 (Chilling effect)**.** *The equilibrium cutoffs t_1 and t_0 are for every type τ_i weakly higher without privacy than in the privacy case. At least one of the inequalities is strict whenever the absence of privacy changes the equilibrium behavior of OP. The difference of either equilibrium cutoff without and with privacy is increasing in τ_i.*

Proof of proposition 10: We already established $t_1^{np} \geq t_0^{np} \geq 0 = t_0^p = t_1^p$. The second result obviously holds if $t_0^{np}(\tau_i) > 0$. Therefore, assume $t_0(\tau_i) = 0$. Note that by (2) (with Δ_{a0} instead of Δ_{1a}) this is only possible if $\Delta_{a0} = 0$. But this implies that OP treats citizens abstaining and those voting 1 exactly the same. In this case, the equilibrium is the same as in the paper and the proof of proposition 1 shows the result. The proof that the difference between no privacy and privacy cutoff is increasing in τ_i is the same as the proof in proposition 1. □

Proposition 11. *OP's payoff without privacy is lower if citizens use the cutoffs $t_0^{np}(\tau)$ and $t_1^{np}(\tau)$ than if they used the cutoffs $t_0^p(\tau) = t_1^p(\tau) = 0$.*

Proof of proposition 11. Start from the case without privacy. If OP plays A with the same probability against two of the three groups (voters of 1, a, 0), then the model boils down to the one in the paper and the result is shown there as proposition 2. Therefore assume now that OP chooses different probabilities of playing A against the three groups which by 11 implies that OP plays A with an interior probability against citizens choosing $p_i = a$. Hence, OP is indifferent between A and M when facing a citizen who played $p_i = a$. Hence, OP could achieve his equilibrium payoff by playing A against $p_i = a$ with the same probability he uses in equilibrium against $p_i = 1$ (while continuing to use his equilibrium strategy against $p_i = 0$). This leads us to a situation that is similar to the one in the paper. The proof of proposition 2 in the paper shows that OP could attain a higher payoff than this if the citizens used $t_0^p(\tau) = t_1^p(\tau) = 0$. □

Lemma 3 in the paper still applies unchanged in the current abstention setting. We can now also obtain the result for large n or high δ from the paper.

Proposition 12. *Assume that (i) OP strictly prefers M in the privacy case, (ii) δ is differentiable and strictly increasing in τ, i.e. $\delta'(\tau) > 0$ for all $\tau \in [\underline{\tau}, \bar{\tau}]$ and (iii) $\Gamma_\infty = \lim_{\theta_i \to \infty} \Gamma_{\theta_i}$ is a non-degenerate distribution in the sense that $\Gamma_\infty(\tau_i) > 0$ for all $\tau_i > \underline{\tau}$.*
1.) Privacy welfare dominates no privacy for large n 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.
2.) Let the disutility of a citizen facing action A by OP be $r\delta(\tau)$ (instead of $\delta(\tau)$). For r sufficiently large, privacy welfare dominates no privacy.

Proof of proposition 12. As in the paper, the "influence" of a single player is $1/n$ and therefore approaches zero which allows us to show that for sufficiently large n only mixed equilibria exist.

If for n sufficiently large the equilibrium is such that two of the three groups (voters of 0, a, 1) are treated in the same way by OP, then the analysis of the paper applies and proposition 3 in the paper yields the result. Hence, assume that for arbitrarily large

n we can find equilibria in which the three groups are treated differently. By (i) and lemma 11, OP has to play M with probability 1 against $p_i = 0$ in this case. If OP uses a truly mixed strategy against $p_i = 1$ (and therefore by lemma 11 also uses a truly mixed strategy against $p_i = a$), then OP has to be indifferent between his equilibrium strategy and playing M against all groups. Hence, OP's payoff under no privacy would be the same as under privacy but expected citizen payoffs are clearly lower. Hence, we can restrict ourselves to the case where OP plays A with probability 1 against $p_i = 1$. This implies that either $\Delta_{1a} \geq 1/2$ or $\Delta_{a0} \geq 1/2$. For concreteness, say $\Delta_{1a} \geq 1/2$. The proof of proposition 3 in the paper shows that t_1' becomes arbitrarily steep as $n \to \infty$ and shows that this implies that $v_1 \to \tau^E$ (where τ^E is the unconditional expected value of τ_i) which by (i) is strictly negative. This contradicts that OP plays A against $p_i = 1$ in equilibrium.[4] This proves (1).

The proof of (2) is analogous to the steps above and the proof in the paper. □

The correlation result (proposition 4 in the paper) goes through if we assume that for $\lambda = 1$ there is a unique equilibrium in which OP plays A against $p_i = 1$ and M against $p_i = 0$ and no one abstains.

4. Extension: 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 2.3 in the paper 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.[5] 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

[4]For the case where $\Delta_{a0} \geq 1/2$, the proof of proposition 3 in the paper shows that the expected τ_i given $\theta_i \geq t_0(\tau_i)$ approaches τ^E. This implies that OP's best response against at least on of the two groups $p_i = a$ and $p_i = 1$ is M which contradicts that $\Delta_{a0} \geq 1/2$.

[5]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 paragraph "Majority rule and $t = 0$ are optimal if the planner could choose q and t" below for the – somewhat technical – 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 4 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 4 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 4.

Majority rule and $t = 0$ are optimal if the planner could choose q and t It remains to show that in the hypothetical problem in which the planner could choose both t and q, he would optimally choose $t = 0$ and majority rule. The planner's expected payoff in this problem is

$$V(q,t) = \sum_{m=0}^{n} \binom{n}{m} \Phi(t)^{n-m}(1-\Phi(t))^m q_m \left((n-m)\mathbb{E}[\theta|\theta < t] + m\mathbb{E}[\theta|\theta \geq t]\right)$$

where q_j denotes the probability that $p = 1$ is chosen if excatly j individuals choose $p_i = 1$. From here it is already obvious that $q_0 = 0$ and $q_n = 1$ as $\mathbb{E}[\theta|\theta < t] < 0$ and $\mathbb{E}[\theta|\theta \geq t] > 0$ for any t. Furthermore, the optimal q will be a cutoff rule where $q_m = 0$ if $m < \hat{m}$ and $q_m = 1$ if $m \geq \hat{m}$ for some $\hat{m}$. For the cutoff $\hat{m}$, we have $dV/d\hat{m} \geq 0$ which

is equivalent to

$$\begin{aligned} (n-\hat{m})\mathbb{E}[\theta|\theta<t] + \hat{m}\mathbb{E}[\theta|\theta\geq t] &\geq 0 \\ \Leftrightarrow (n-\hat{m})\frac{\int_{-\infty}^{t}\theta\, d\Phi(\theta)}{\Phi(t)} + \hat{m}\frac{\int_{t}^{\infty}\theta\, d\Phi(\theta)}{1-\Phi(t)} &\geq 0 \\ \Leftrightarrow -(n-\hat{m})(1-\Phi(t)) + \hat{m}\Phi(t) &\geq 0 \end{aligned} \tag{3}$$

where we use $-\int_{-\infty}^{t}\theta\, d\Phi(\theta) = \int_{t}^{\infty}\theta\, d\Phi(\theta) > 0$ by the fact that the expectation of a standard normal random variable is zero.

This allows us rewrite V.

$$\begin{aligned} V(q,t) &= \sum_{m=1}^{n-1}\Phi(t)^{n-m}(1-\Phi(t))^{m}q_m\frac{n!}{m!(n-m-1)!}\mathbb{E}[\theta|\theta<t] \\ &\quad + \sum_{m=1}^{n-1}\Phi(t)^{n-m}(1-\Phi(t))^{m}q_m\frac{n!}{(m-1)!(n-m)!}\mathbb{E}[\theta|\theta\geq t] + (1-\Phi(t))^{n}n\mathbb{E}[\theta|\theta\geq t] \\ &= \sum_{m=1}^{n-1}\Phi(t)^{n-m-1}(1-\Phi(t))^{m}q_m\frac{n!}{m!(n-m-1)!}\int_{-\infty}^{t}\theta\, d\Phi(\theta) \\ &\quad + \sum_{m=1}^{n-1}\Phi(t)^{n-m}(1-\Phi(t))^{m-1}q_m\frac{n!}{(m-1)!(n-m)!}\int_{t}^{\infty}\theta\, d\Phi(\theta) + (1-\Phi(t))^{n-1}n\int_{t}^{\infty}\theta\, d\Phi(\theta) \\ &= \Phi(t)^{n-\hat{m}}(1-\Phi(t))^{\hat{m}-1}\frac{n!}{(\hat{m}-1)!(n-\hat{m})!}\int_{t}^{\infty}\theta\, d\Phi(\theta) \end{aligned} \tag{4}$$

where the last equality uses (i) that $q_m = 0$ if $m < \hat{m}$ and $q_m = 1$ if $m \geq \hat{m}$, (ii) that $\int_{-\infty}^{\infty}\theta\, d\Phi(\theta) = 0$ and (iii) that the term m term from the first sum "fits together" with the $m+1$ term in the second sum and cancel each other out using (ii).

Given the optimal $\hat{m}$, the optimal t has to satisfy the first order condition $\partial V/\partial t = 0$ which can be rewritten as

$$\frac{\partial V}{\partial t} = \phi(t)\frac{n!\Phi(t)^{n-\hat{m}-1}(1-\Phi(t))^{\hat{m}-2}}{(\hat{m}-1)!(n-\hat{m})!}\left[(n-\hat{m})(1-\Phi(t))\int_{t}^{\infty}\theta\, d\Phi(\theta) - (\hat{m}-1)\Phi(t)\int_{t}^{\infty}\theta\, d\Phi(\theta) - t\Phi(t\right. \tag{5}$$

However, it is worthwhile to go back to (4). This can be rewritten as

$$\begin{aligned} V(q,t) &= n\binom{n-1}{\hat{m}-1}\Phi(t)^{n-\hat{m}}(1-\Phi(t))^{\hat{m}-1}\int_{t}^{\infty}\theta\, d\Phi(\theta) \\ &= \binom{\tilde{n}}{\tilde{m}}\Phi(t)^{\tilde{n}-\tilde{m}}(1-\Phi(t))^{\tilde{m}}n\int_{t}^{\infty}\theta\, d\Phi(\theta) \end{aligned}$$

where $\tilde{n} = n-1$ and $\tilde{m} = \hat{m}-1$. The latter expression consists of the probability of $\tilde{m}$ hits according to the binomial distribution with $\tilde{n}$ draws and probability $1-\Phi(t)$. As the integral is positive but does not depend on $\tilde{m}$, the optimal $\tilde{m}$ is the one that maximizes

this probability. A binomial distribution has its maximal probability at its mode which in this case is $\lfloor(1-\Phi(t))*(\tilde{n}+1)\rfloor$. Consequently, the optimal $\hat{m} = 1+\tilde{m}$ is $1+\lfloor(1-\Phi(t))*n\rfloor$. Plugging this back into V, we obtain a maximization problem over one variable t:

$$\max_t \binom{\tilde{n}}{\lfloor(1-\Phi(t))*n\rfloor} \Phi(t)^{\tilde{n}-\lfloor(1-\Phi(t))*n\rfloor}(1-\Phi(t))^{\lfloor(1-\Phi(t))*n\rfloor} n \int_t^\infty \theta \, d\Phi(\theta).$$

This objective function is unfortunately somewhat ill behaved; see the graph of the objective funtion for $n = 21$ in figure 1.

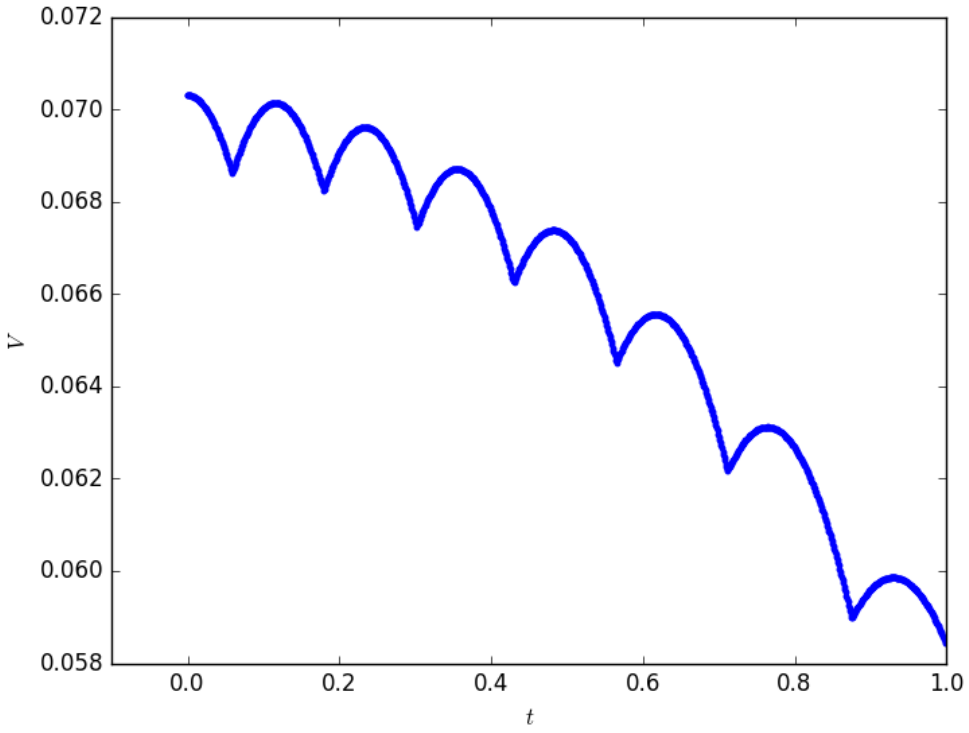

Figure 1: Objective as funtion of t for $n = 21$.

We claim that the solution to the problem is $t = 0$ and majority voting (i.e. $\hat{m} = (n+1)/2$ given our assumption that n is odd). Note that this solution does indeed satisfy the necessary conditions (3) and (5). Admittedly, these conditions are necessary and not sufficient. However, numerical analysis is very easy for a given n. We verified numerically that $t = 0$ is optimal for all odd n between 1 and 100000.

To give a more analytical idea why this is true, return to (5). Let us denote by $R(t) \in [-1+\Phi(t), \Phi(t)]$ the difference between mode and mean. That is, $R(t) = \tilde{n}(1-\Phi(t)) - \lfloor n(1-\Phi(t))\rfloor$. Using R and letting $\hat{m}$ be the optimal one, we get from (5)

$$\left.\frac{\partial V}{\partial t}\right|_{\hat{m}} = \phi(t)\frac{n!\Phi(t)^{n-\hat{m}-1}(1-\Phi(t))^{\hat{m}-2}}{(\hat{m}-1)!(n-\hat{m})!}\left[R(t)\int_t^\infty \theta d\Phi * -t\Phi(t)(1-\Phi(t))\right].$$

Clearly, the fraction in front of the brackets is strictly positive and we will now concentrate on the term in brackets. For $t = 0$, we have $R = 0$ (recall that n is odd and therefore

$(n-1)/2$ is an integer which is both mean and mode) and both terms in brackets are zero. For t slightly above 0, the mode will not change but the mean $(1-\Phi(t))(n-1)$ gets smaller which means that $R(t)$ is negative. Therefore, V has a local maximum at $t=0$. As we increase t, R will fluctuate around zero.[6] If we abstract from R (and treat it as zero for now), it is clear that $t=0$ is optimal as V' is negative for all $t>0$. R creates some wave like fluctuations around this downward sloping function. In figure 2, we plot $\partial V/\partial t|_{\hat{m}}$. From this figure it is already clear that the integral over this function from zero to $t>0$ will be negative and figure 3 shows exactly this. Note that n enters the term in brackets only indirectly through R, i.e. n only determines the frequency with which R fluctuates but does not change the qualitative conclusions.

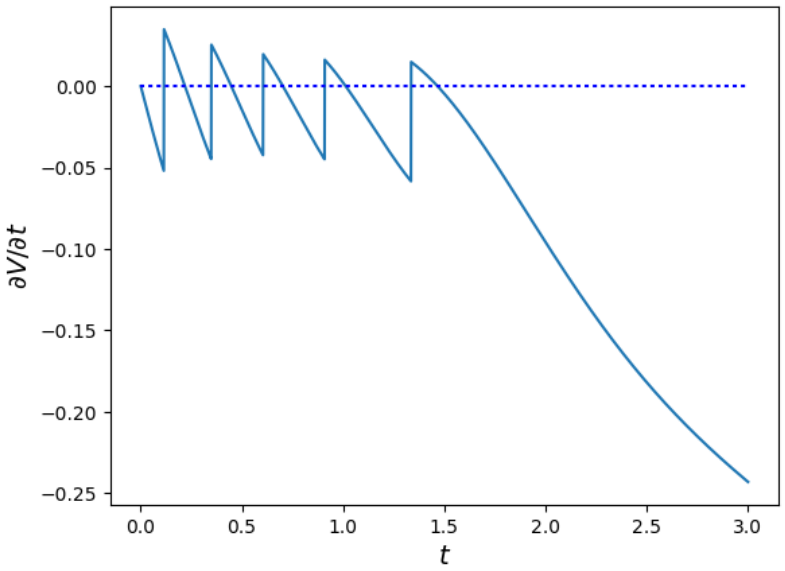

Figure 2: $V'(t)$ for $n=11$.

5. Extension: General Information Aggregation Process q

We give modified proofs for the case of a more general information aggregation process q that is strictly increasing, point-symmetric around 0.5, and s-shaped, i.e. weakly convex in $[0,0.5]$ and weakly concave in $[0.5,1]$.

Lemma 12. *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: $t^p(\tau_i)=0$.*

Proof. For $\theta_i > (\max_{\tau_i}\delta(\tau_i))\,/\,(min_k\{q(k7n)-q((k-1)/n):\ k\in\{1,\dots,n\}\})$, it is a dominant action to choose $p_i=1$. Similarly, for $\theta_i < -(\max_{\tau_i}\delta(\tau_i))\,/\,(min_k\{q(k/n)-q((k-1)/n):$

[6]The fluctuations of R are constant in the the hit rate. However, the hit rate is $1-\Phi(t)$ and this means that the fluctuations get smaller in t as $1-\Phi(t)$ is convexly decreasing in $t\geq 0$.

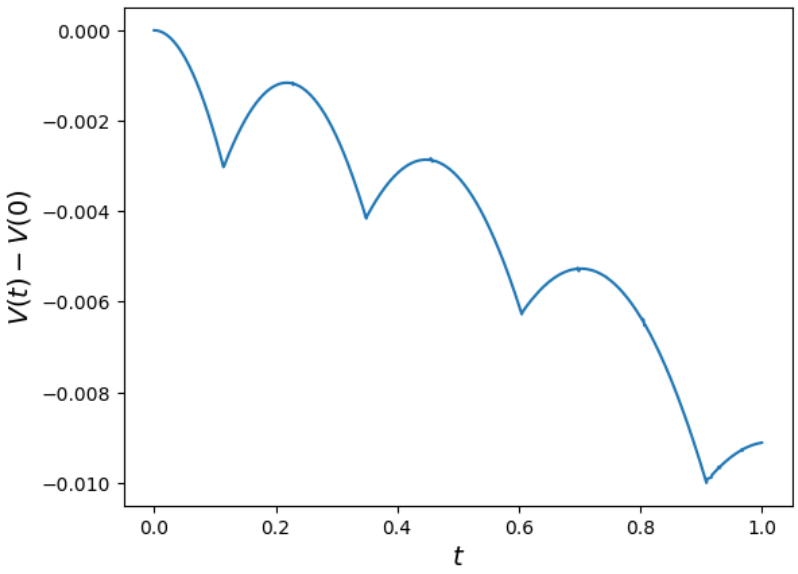

Figure 3: This graph integrates the graph in figure 2 from zero to t and therefore shows for each t by how much this t is worse than $t = 0$. $V(t) - V(0) = \int_0^t \partial V/\partial t|_{m^*}\, dt$.

$k \in \{1, \ldots, n\}\}$), it is a dominant action to choose $p_i = 0$. Write the expected utility difference of playing $p_i = 1$ and playing $p_i = 0$ as

$$-\delta(\tau_i)\Delta + \theta_i * \sum_{k=1}^{n} ((q(k/n) - q((k-1)/n)) * prob(k-1)) \tag{6}$$

where $prob(k-1)$ is i's belief that exactly $k-1$ other citizens will choose $p_j = 1$ and $\Delta \in [-1, 1]$ is the difference between the (believed) probability that OP plays A when facing a citizen who has played $p_i = 1$ and a citizen who has played $p_i = 0$. Clearly, (6) is strictly increasing and continuous in θ_i. As it is optimal to play $p_i = 1$ ($p_i = 0$) 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 dominance regions above establish that an interior cutoff exists.) 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 $\theta_i = 0$ as the sum is clearly positive (recall that the cumulative distribution function q was strictly increasing by assumption). Consequently, $t^p(\tau_i) = 0$. □

Lemma 13. *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 $t^{np}(\tau) > 0$.*

Proof. 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. Suppose cutoff $t^* \geq 0$ is consumer surplus optimal. A

necessary condition for optimality is the following: Say an individual has type $\theta_i = t^*$, then his choice must be consumer surplus neutral. That is, whether he chooses $p_i = 0$ or $p_i = 1$ must lead to the same expected consumer surplus (conditional on his own type being $\theta_i = t^*$). If this condition was not satisfied, either in- or decreasing t^* will increase expected consumer surplus thereby contradicting the optimality of t^*. We will show that the only t^* satisfying this necessary condition is $t^* = 0$.

The expected difference of consumer surplus when choosing $p_i = 1$ and $p_i = 0$ is (where we write t instead of t^* to shorten notation)

$$t+\sum_{l=0}^{n-1}\binom{n-1}{l}\Phi(t)^l(1-\Phi(t))^{n-1-l}(q((l+1)/n)-q(l/n))\left(l\mathbb{E}[\theta|\theta<t]+(n-1-l)\mathbb{E}[\theta|\theta>t]\right) \tag{7}$$

where l is the number of others choosing $p_i = 0$ (according to the cutoff strategy t). For linear q with slope α, the expression in (7) can be simplified easily (using $\Phi(t)\mathbb{E}[\theta|\theta < t]+(1-\Phi(t))\mathbb{E}[\theta|\theta > t] = 0$ which follows from the fact that θ has unconditional expected value of zero):

$$\begin{aligned} t + \alpha\sum_{l=0}^{n-1}\binom{n-1}{l}\Phi(t)^l(1-\Phi(t))^{n-1-l}\left(-l\frac{1-\Phi(t)}{\Phi(t)}+(n-1-l)\right)\mathbb{E}[\theta|\theta>t] \\ t+\frac{\alpha}{\Phi(t)}\mathbb{E}[\theta|\theta>t]\sum_{l=0}^{n-1}\binom{n-1}{l}\Phi(t)^l(1-\Phi(t))^{n-1-l}\left(-l+\Phi(t)(n-1)\right) \\ = t+\frac{\alpha}{\Phi(t)}\mathbb{E}[\theta|\theta>t]\left(-(n-1)\Phi(t)+\Phi(t)(n-1)\right) = t \end{aligned}$$

where we use the fact that the expected value of a binomial distribution with hit rate $\Phi(t)$ and $n-1$ draws is $(n-1)\Phi(t)$. Clearly, the necessary condition for optimality can only be satisfied for $t = 0$ with linear q.

Back to (7) with general q functions. First, consider the term $l = (n-1)/2$ (in case n is odd). For this term $l = (n-1-l)$ and as $\mathbb{E}[\theta|\theta < t]+\mathbb{E}[\theta|\theta < t] \geq 0$ by $t \geq 0$, $p_i = 1$ will lead to a higher expected consumer surplus in this case. For $l < (n-1)/2$, we clearly have $l\mathbb{E}[\theta|\theta < t]+(n-1-l)\mathbb{E}[\theta|\theta < t] > 0$ by $t > 0$ and again $p_i = 1$ increases expected consumer surplus. However, for $l > (n-1)/2$ the opposite might be the case. Hence, we have to weigh terms with different l against each other. In particular, we will consider the terms $l > (n-1)/2$ and $n-1-l < (n-1)/2$ jointly. By the assumption that q is point symmetric around $1/2$, $q((l+1)/n)-q(l/n) = q((n-1-l+1)/n)-q((n-1-l)/n)$. Furthermore, the binomial coefficient is symmetric around the mean which means that also $\binom{n-1}{l} = \binom{n-1}{n-1-l}$. Consequently, we can write the sum of the two terms corresponding to l and $n-1-l$ as follows (using $z = 2l-n+1$ and dropping the argument of $\Phi(t)$ to

save space)

$$\binom{n-1}{l}\Phi^{n-1-l}(1-\Phi)^{n-1-l}(q((l+1)/n)-q(l/n))$$
$$\{\mathbb{E}[\theta|\theta<t]\left(l\Phi^z+(n-1-l)(1-\Phi)^z\right)+\mathbb{E}[\theta|\theta>t]\left((n-1-l)\Phi^z+l(1-\Phi)^z\right)\}.$$

Note that $\Phi\mathbb{E}[\theta|\theta<t]+(1-\Phi)\mathbb{E}[\theta|\theta>t]=0$ as the unconditional expected value of θ is zero. Plugging this into the previous expression gives

$$\begin{aligned}
&\binom{n-1}{l}\Phi^{n-1-l}(1-\Phi)^{n-1-l}(q((l+1)/n)-q(l/n))\\
&\left\{\frac{1}{\Phi}\mathbb{E}[\theta|\theta>t]\left(-l(1-\Phi)\Phi^z-(n-1-l)(1-\Phi)^{z+1}+(n-1-l)\Phi^{z+1}+l\Phi(1-\Phi)^z\right)\right\}\\
&\quad=\binom{n-1}{l}\Phi^{n-1-l}(1-\Phi)^{n-1-l}(q((l+1)/n)-q(l/n))\frac{1}{\Phi}\mathbb{E}[\theta|\theta>t]\\
&\quad\left\{(n-1)\left(\Phi^{z+1}-(1-\Phi)^{z+1}\right)-l\left(\Phi^{z+1}-(1-\Phi)^{z+1}+\Phi^z(1-\Phi)-\Phi(1-\Phi)^z\right)\right\}\\
&\quad=\binom{n-1}{l}\Phi^{n-1-l}(1-\Phi)^{n-1-l}(q((l+1)/n)-q(l/n))\frac{1}{\Phi}\mathbb{E}[\theta|\theta>t]\\
&\qquad\left\{\frac{n-1}{2}\left(\Phi^{z+1}-(1-\Phi)^{z+1}-\Phi^z(1-\Phi)+\Phi(1-\Phi)^z\right)\right.\\
&\qquad\left.-\frac{z}{2}\left(\Phi^{z+1}-(1-\Phi)^{z+1}+\Phi^z(1-\Phi)-\Phi(1-\Phi)^z\right)\right\}\\
&\quad=\binom{n-1}{l}\Phi^{n-1-l}(1-\Phi)^{n-1-l}(q((l+1)/n)-q(l/n))\frac{1}{\Phi}\mathbb{E}[\theta|\theta>t]\\
&\qquad\left\{\frac{n-1}{2}\left(\Phi^z+(1-\Phi)^z\right)(2\Phi-1)-\frac{z}{2}\left(\Phi^z-(1-\Phi)^z\right)\right\}.
\end{aligned}$$

We will show below that there is a cutoff z such that the term in curly brackets is strictly negative iff z above cutoff (given $\Phi>1/2$). In this case, we can argue that S-shaped q will put relatively more weight on positive terms than on negative ones (compared to linear q). As the sum was zero with linear q it will now be positive which contradicts that the necessary condition for optimality is met for $t>0$.

It remains to show that – for given $\Phi\in[1/2,1]$ and n – there exists a cutoff $\bar{z}$ such that the term in curly brackets is positive for $z\leq\bar{z}$ and negative if $z>\bar{z}$. Note that the term in curly brackets is positive for $z=1$ and therefore it is sufficient to show that it has (at most) one zero if viewed as a function of z. To this end, let $g(z)=(n-1)\left(\Phi^z+(1-\Phi)^z\right)(2\Phi-1)-z\left(\Phi^z-(1-\Phi)^z\right)=\Phi^z(A-z)+(1-\Phi)^z(A+z)$ where $A=(n-1)(2\Phi-1)\geq 0$. Then $g(z)=0$ implies that

$$\frac{\Phi}{1-\Phi}=\left(\frac{A+z}{z-A}\right)^{1/z}. \tag{8}$$

By $1 \geq \Phi \geq 1/2$, we have $\Phi/(1-\Phi) \geq 1$ and therefore (8) can only hold if $(A+z)/(z-A) \geq 1$. In particular, this implies $z > A$ at every zero of g. We will now compute g' and show that g' is negative at every zero of g. As g is continuous and differentiable, this implies that g can have only one zero.

$$\begin{aligned} g'(z) = \frac{d\, e^{\frac{log((A+z)/(z-A))}{z}}}{dz} &= \left(\frac{A+z}{z-A}\right)^{1/z} \frac{\frac{z-A}{z+A}\frac{-2A}{(z-A)^2}z - log\left(\frac{A+z}{z-A}\right)}{z^2} \\ &= \left(\frac{A+z}{z-A}\right)^{1/z-1} \frac{-2az - (z-A)^2\frac{A+z}{z-A}log\left(\frac{A+z}{z-A}\right)}{z^2(z-A)^2} \\ &= \left(\frac{A+z}{z-A}\right)^{1/z-1} \frac{-2az - (z^2-A^2)log\left(\frac{A+z}{z-A}\right)}{z^2(z-A)^2}. \end{aligned}$$

At every z at which $g(z) = 0$ we must have $g'(z) < 0$ because we established above that at such z we have $z > A$ and $(A+z)/(z-A) \geq 1$. □

Given the two lemmas above, the proofs of propositions 3-5 will go through without change with one small exception: For the proof of proposition 3 it is necessary to show that t^{np} becomes arbitrarily steep as $n \to \infty$ if $\Delta = 1$. We will show this here: Using (6) and assuming $\Delta = 1$, we can write

$$t^{np}(\tau_i) = \frac{\delta(\tau_i)}{\sum_{k=1}^{n}(q(k/n) - q((k-1)/n) * prob(k-1))}.$$

As q is assumed to be continuously differentiable on $[0,1]$, its slope q' attains its maximum on $[0,1]$ which we denote by ζ. The denominator of the fraction above is bounded from above by $\sum_{k=1}^{n}(\zeta/n) * prob(k-1) = \zeta/n \sum_{k=1}^{n} prob(k-1) = \zeta/n$ which converges to zero as $n \to \infty$. Hence, t^{np} becomes arbitrarily steep as $n \to \infty$.

6. Extension: Privacy as the Result of Information Design

Suppose that we give OP the possibility to choose a signal technology that informs him of each citizen's decision p_i. However, OP has to choose this technology before the game, and once chosen, it becomes common knowledge, so that individuals can adjust their choices accordingly. It is up to OP to choose whether this signal technology should be noisy or not.

If OP chooses a perfectly revealing signal or a perfectly uninformative signal, we are back in the two cases of our main analysis. This section generalizes proposition 3 by establishing that OP cannot do better than choosing privacy if n (or δ) is sufficiently large. That is, privacy may endogenously emerge even if the information flow is under OP's control. We assume throughout this section $\delta' > 0$ and $\mathbb{E}[\tau_i] < 0$.

As OP has only two actions, it is without loss of generality to consider a binary signal

technology that sends a signal in $\{A, M\}$. A signal technology for citizen i consists of two probabilities ρ_i^0 and ρ_i^1 such that ρ_i^j is the probability that OP receives signal A after i chooses action j and signal M with the complementary probability $1 - \rho_i^j$. Due to the revelation principle, it is without loss of generality to restrict ourselves to obedient signal technologies, i.e. technologies such that OP plays A (M) if he receives the signal A (M). For simplicity, we will consider only the case where the same signal technology is used for all individuals; that is, ρ_i^j does not depend on i and we can write ρ^j instead.

For the remainder of this section, let Δ denote the expected probability of being treated aggressively after choosing $p_i = 1$ minus the expected probability of being treated aggressively after choosing $p_i = 0$ given a certain signal technology (and OP obedience). That is, $\Delta = \rho^1 - \rho^0$. With this slight change in notation an individual's equilibrium strategy is still given by the cutoff $t(\tau_i) = \Delta n \delta(\tau_i)$.

If Δ is close to zero, i.e. if $\rho^1 \approx \rho^0$, OP's belief about i's type τ_i will (almost) not depend on the signal OP receives and will therefore be (almost) equal to his prior $\mathbb{E}[\tau]$. That is, there exists an $\varepsilon > 0$ such that OP's belief is below 0 for both signals if $\Delta \leq \varepsilon$. In this case, obedience is violated (unless $\rho^0 = \rho^1 = 0$) as OP prefers M even when receiving signal A. Hence, obedience constrains the choice of signal technologies to signal technologies with $\Delta > \varepsilon$ (unless $\rho^0 = \rho^1 = 0$ which is equivalent to the privacy case).

If there is an equilibrium that does not correspond to privacy, the previous paragraph implies that $t(\tau_i) > \varepsilon n \delta(\tau_i)$ because $\Delta > \varepsilon$. As $\varepsilon > 0$, this lower bound implies that t becomes arbitrarily high and steep as n grows large. Following the proof of proposition 3, this implies that $\mathbb{E}[\tau_i | \tau_i \geq t(\tau_i)] < 0$ for sufficiently large n. As OP's belief about τ_i in any signal technology is bounded from above by $\mathbb{E}[\tau_i | \tau_i \geq t(\tau_i)]$, OP consequently prefers M over A when receiving signal A for sufficiently large n. This contradicts obedience and we have therefore established that equilibrium play is equivalent to the privacy case when n is sufficiently large. A similar argument holds for sufficiently high r where the disutility of being treated aggressively is denoted by $r\delta(\tau_i)$.

7. Extension: Optional Privacy

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.[7] 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 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

[7] 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.

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.

In general, the problem lies in the fact that choosing privacy is, in itself, informative about the individual's type. In addition, any individual that chooses not to have privacy is revealing information about those who still choose privacy, and hence exerts a sort of informational externality. Choosing privacy becomes less informative, of course, if it is a one-time choice and there are choices to be made and many interactions. This dilutes the informational content of choosing privacy – however, it is still informative about an individual's type.

8. Extension: 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 (7), 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.[8] 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 only exists a mixed equilibrium in which OP sometimes gathers information and treats

[8] 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.

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.*

9. Extension: 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 following.

Here we analyze the extension with a defensive action. That is each individual can choose in his interaction with OP to play the defensive action D (or alternatively opt for "not D"). The defensive action reduces OP's payoff and is a best response to A but is not a best response to M.

What we want to show in this section is that the model can easily be extended by introducing a defensive action such that privacy could lead to (i) OP being *strictly* better off with privacy while (ii) individuals being 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 and this is what we are going to do. Suppose $\tau_i \in \{\underline{\tau}, \bar{\tau}\}$, that is, τ_i can have only one of two values. Furthermore, assume that the probability that $\tau_i = \bar{\tau}$ equals

$$\gamma_{\theta_i} = \begin{cases} 0.7 - \frac{0.3}{\theta_i + 1} & \text{if } \theta_i \geq 0 \\ 0.1 - \frac{0.3}{\theta_i - 1} & \text{if } \theta_i < 0. \end{cases}$$

That is, the probability of a high $\bar{\tau}$ is increasing in θ_i and is point symmetric around $(0, 0.4)$. We take $\underline{\tau} = -2$, $\bar{\tau} = 3$ and $\delta(\tau_i)$ as given as in table 1.

action/type	$\underline{\tau}$	$\bar{\tau}$
not D	-0.1	-0.125
D	0.0	-0.025

Table 1: $-\delta(\tau_i)$ depending on whether the defensive action is taken.

If an individual takes action D and OP plays M, his payoff is -0.1, that is, the costs of the action D are 0.1. Note that the individual wants to play D if the chance of playing A is higher than 1/2. OP's payoffs are reduced by 1 if an individual plays D (for simplicity the payoff reduction is assumed to be independent of OP's action).

Under privacy, it is an equilibrium that every individual chooses $p_i = 1$ if and only if $\theta_i \geq 0$ while in the second stage OP plays M and no individual takes the action D. Without privacy, this is no longer an equilibrium as OP prefers to deviate by playing A against all individuals choosing $p_i = 1$: The probability that an individual is of type $\tau_i = \bar{\tau}$ given $\theta_i \geq 0$ (and therefore $p_i = 1$) is

$$\frac{\int_0^\infty \gamma_{\theta_i} \, d\Phi(\theta_i)}{2} \approx 0.51$$

which implies that OP's best response is A.

The equilibrium in the case without privacy is as follows: Individuals use cutoff strategies characterized by cutoffs $t(\underline{\tau}) = 0$ and $t(\bar{\tau}) = N * 0.025$. In the second stage, those individuals that chose $p_i = 1$ will play D. OP plays A against all individuals that chose $p_i = 1$ and M otherwise. To see that this is an equilibrium, note that an individual of type $(\theta_i, \tau_i) = (0, \underline{\tau})$ is indeed indifferent between choosing $p_i = 0$ and not playing D, which gives a payoff of 0 as OP will play M, and $p_i = 1$ and playing D which also gives a payoff of 0 as OP will then play A. Similarly, an individual of type $(\theta_i, \tau_i) = (0.025N, \bar{\tau})$ is indifferent between choosing $p_i = 0$ and not playing D and choosing $p_i = 1$ and playing D. The reason is that choosing $p_i = 1$ increases the probability of $p = 1$ being chosen by $1/N$ and therefore increases the expected payoff of an individual with $\theta_i = 0.025N$ by 0.025. However, the down side of choosing $p_i = 1$ is that the payoff in the interaction stage is

0.025 lower as $-\delta(\bar{\tau}) = -0.025$ (when playing D). For OP, the probabilities

$$\begin{aligned} prob(\tau_i = \bar{\tau}|p_i = 0) &= \frac{\int_{-\infty}^{0.025N} \gamma_{\theta_i}\, d\Phi(\theta_i)}{\int_{-\infty}^{0.025N} \gamma_{\theta_i}\, d\Phi(\theta_i) + \int_{-\infty}^{0} 1 - \gamma_{\theta_i}\, d\Phi(\theta_i)} \\ prob(\tau_i = \bar{\tau}|p_i = 1) &= \frac{\int_{0.025N}^{\infty} \gamma_{\theta_i}\, d\Phi(\theta_i)}{\int_{0.025N}^{\infty} \gamma_{\theta_i}\, d\Phi(\theta_i) + \int_{0}^{\infty} 1 - \gamma_{\theta_i}\, d\Phi(\theta_i)} \end{aligned}$$

are such that playing A (M) against those that chose $p_i = 1$ ($p_i = 0$) is optimal, i.e. $prob(\tau_i = \bar{\tau}|p_i = 0) \leq 0.4 \leq prob(\tau_i = \bar{\tau}|p_i = 1)$, if $N \leq 22$.[9]

OP's expected payoffs in the equilibrium without privacy are $-0.043 * N$ while OP's profits with privacy are zero. Individuals are strictly worse off if they have $\theta_i > 0$: The reasons are (i) that some are chilled and therefore expect a lower payoff from the information aggregation in stage 1, (ii) those that are not chilled have to endure action A by OP (and have to bear the costs of the defensive action). Individuals with $\theta_i < 0$ benefit from the chilling of others as this chilling implies that their personally preferred alternative is more likely to be implemented. Note, however, that – by the symmetry of the setup – this only offsets the first negative effect on those with $\theta_i > 0$ (in expectation, e.g. behind the veil of ignorance). The second negative effect on those with $\theta_i > 0$ lowers the expected payoff of any individual.

References

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

[9]For $N > 22$, no pure strategy equilibria exist without privacy and OP will therefore be indifferent between privacy and no privacy – cf. proposition 3 in the paper.