Supplementary material to
An Informational Theory of Privacy

Ole Jann and Christoph Schottmüller

February 20, 2017

1. Extension: 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 1. *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[1]

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

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

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

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

$$
\begin{aligned}
v_0 &= \int_{\mathbb{R}} \tau\, d\beta_0(\tau) \\
v_1 &= \int_{\mathbb{R}} \tau\, d\beta_1(\tau).
\end{aligned}
$$

Lemma 2. *In every perfect Bayesian equilibrium (without privacy), $v_1 \ge v_0$.*

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

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

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

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 1. *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 3. *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 $\mathbb{E}[\theta|\theta_i] > 0$.[2] 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

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

(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 2. *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.

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 4. *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 2: 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 3 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 4 of the paper and its proof go through without change. Proposition 5 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 3 (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. □

2. 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 3 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 5. *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 5. 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 5 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 6. *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 6. 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 $\int_{\underline{\tau}}^{\overline{\tau}} \tau_i \, d\Gamma_{\theta_i} < 0$ 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 4 (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 4: 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 in the paper. □

Proposition 5. *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 5. 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 6 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 of 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 6. *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 6. 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 6, 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 6 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.[3] This proves (1).

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

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.

3. Extension: Endogenous Information Aggregation Process q

This section contains some additional analysis concerning the extension described in section 5.1. It remained 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

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

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.

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

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

[3] 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 $\Delta_{a0} \geq 1/2$.

rewritten as

$$\frac{\partial V}{\partial t} = \frac{n!\Phi(t)^{n-\hat{m}-1}(1-\Phi(t))^{\hat{m}-2}}{(\hat{m}-1)!(n-\hat{m})!}\left[(n-\hat{m})(1-\Phi(t))\int_t^\infty \theta\, d\Phi(\theta) - (\hat{m}-1)\Phi(t)\int_t^\infty \theta\, d\Phi(\theta) - t\Phi(t)(1-\Phi(t)\right] = 0. \tag{5}$$

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 either $\lfloor(1-\Phi(t))*\tilde{n}\rfloor$ or $\lceil(1-\Phi(t))*\tilde{n}\rceil$. Consequently, the optimal $\hat{m} = 1+\tilde{m}$ is either $1+\lfloor(1-\Phi(t))*(n-1)\rfloor$ or $1+\lceil(1-\Phi(t))*(n-1)\rceil$ (depending on which of the two is the mode). Plugging this back into V, we obtain a maximization problem over one variable t:

$$\max_t \max\{B_{\tilde{n},1-\Phi(t)}(\lfloor(1-\Phi(t))\tilde{n}\rfloor), B_{\tilde{n},1-\Phi(t)}(\lceil(1-\Phi(t))\tilde{n}\rceil)\}\int_t^\infty \theta\, d\Phi(\theta)$$

where $B_{\tilde{n},1-\Phi(t)}(k) = \binom{\tilde{n}}{k}\Phi(t)^{\tilde{n}-k}(1-\Phi(t))^k$. This objective function is unfortunately somewhat ill behaved; see the graph of the objective funtion for $n = 21$ in figure 1.

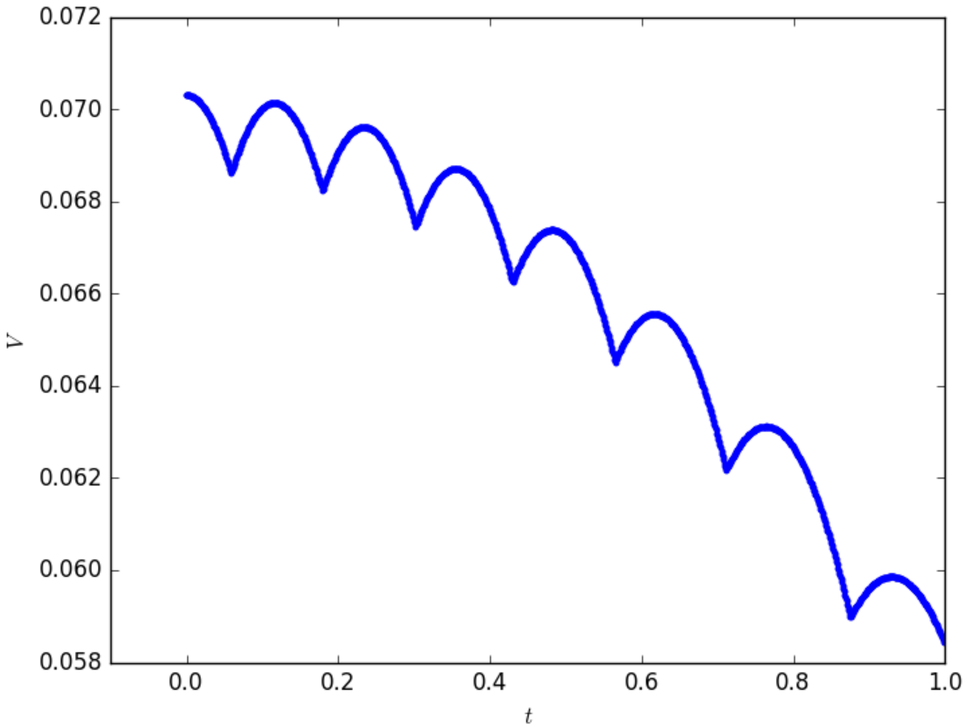

Figure 1: Objective as funtion of t for $n = 21$.

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 $R(t) \in (-1, 1)$ the difference between the optimal $\hat{m}$ and $1 + (1 - \Phi(t))(n - 1)$. That is, R is the difference between mode and mean of $B_{\tilde{n},1-\Phi(t)}$. Plugging the optimal $\hat{m}$ into (5) we get (dropping function arguments to save space)

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

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) 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 positive. Therefore, V has a local maximum at $t = 0$. As we increase t, R will fluctuate around zero. 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 $\partial V/\partial t|_{m^*}$. 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.

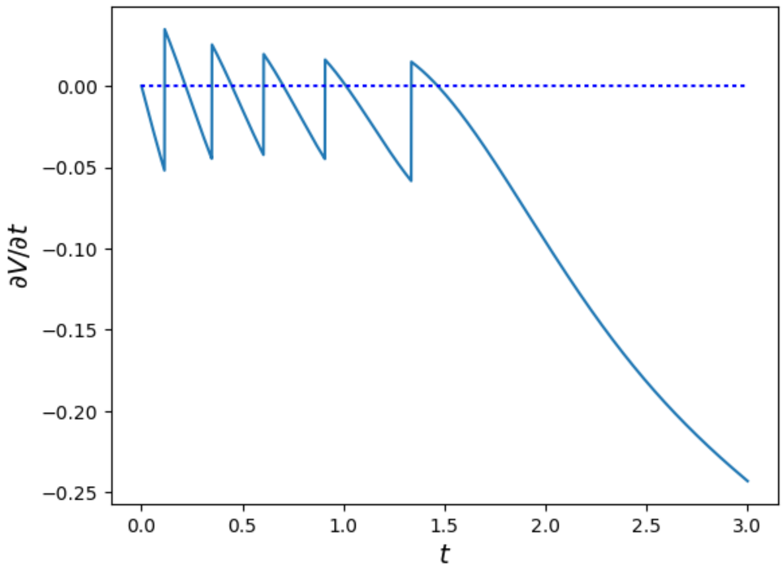

Figure 2: $V'(t)$ for $n = 11$.

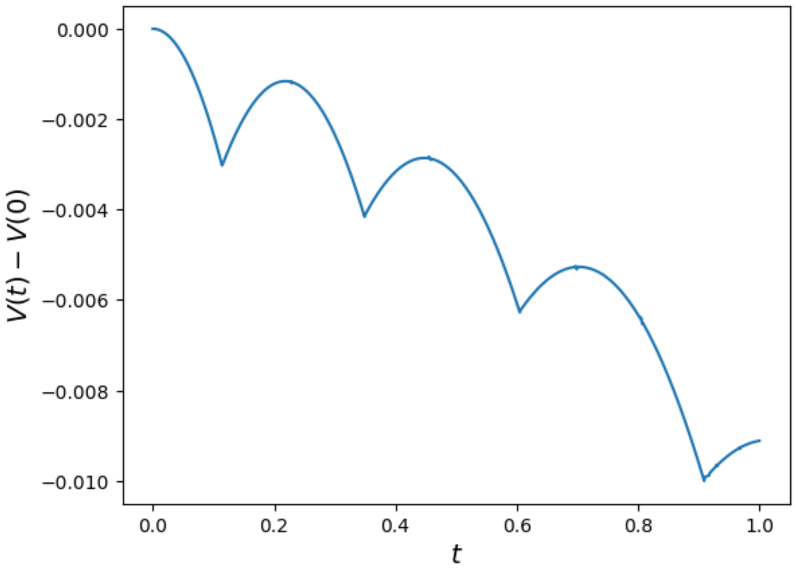

Figure 3: This graph integrates the graph in figure 2 from zero to t and therefore shows for each t by how much this t is worse than $t = 0$. $V(t) - V(0) = \int_0^t \partial V/\partial t|_{m^*}\, dt$.