

Online Appendix for “Algorithms to the Rescue: Market Mechanisms for Consensual Trading of Unbiased Individual Data”

Brian Birkhead

Coniak Ltd., 2 Titniss Cottages, London Road, Sunninghill, Berkshire SL5 0PP, United Kingdom,
brian.birkhead@coniak.co.uk

Ashkan Eshghi

Information Systems and Management, Warwick Business School, the University of Warwick, Coventry CV4 7AL, United Kingdom, ashkan.eshghi@wbs.ac.uk

Ram D. Gopal

Information Systems and Management, Warwick Business School, the University of Warwick, Coventry CV4 7AL, United Kingdom, ram.gopal@wbs.ac.uk

Hooman Hidaji

Haskayne School of Business, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada,
hooman.hidaji@haskayne.ucalgary.ca

Raymond A. Patterson

Haskayne School of Business, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada,
raymond.patterson@ucalgary.ca

Appendix: Proofs

Proof of Proposition 1:

Given the definition of data subjects’ expected utility in (1), if data subject i reports v'_i , then her expected utility is

$$U(v'_i) = \begin{cases} \gamma(\min_{j \neq i}(v'_j)) - \gamma(v_i), & \text{if } v'_i < \min_{j \neq i}(v'_j) \\ 0, & \text{otherwise} \end{cases}$$

Therefore, if data subject i truthfully reports her privacy concern, her utility is

$$U(v'_i = v_i) = \begin{cases} \gamma(\min_{j \neq i}(v'_j)) - \gamma(v_i), & \text{if } v_i < \min_{j \neq i}(v'_j) \\ 0, & \text{otherwise} \end{cases}$$

From (2) we have $\frac{\partial \gamma}{\partial v} > 0$, therefore, if $\min_{j \neq i}(v'_j) > v_i$, then $\gamma(\min_{j \neq i}(v'_j)) - \gamma(v_i) > 0$. This means that the utility of truth-telling is non-negative, that is $U(v'_i = v_i) \geq 0$. In other words, if the mechanism is incentive-compatible, then it is individually rational.

Now we only need to prove that truth-telling is the dominant strategy for all data subjects. We prove this by contradiction. Lets assume that $v'_i \neq v_i$, then either $v'_i > v_i$ or $v'_i < v_i$. We consider each case below.

If $v'_i > v_i$, that is, if data subject i over-reports her privacy concern, then

$$\left\{ \begin{array}{ll} \text{if } \min_{j \neq i}(v'_j) > v'_i, & \text{then, } U(v'_i) = U(v_i) = \gamma(\min_{j \neq i}(v'_j)) - \gamma(v_i) \\ \text{if } \min_{j \neq i}(v'_j) < v_i, & \text{then, } U(v'_i) = U(v_i) = 0 \\ \text{if } v_i < \min_{j \neq i}(v'_j) < v'_i, & \text{then, } U(v'_i) = 0, U(v_i) = \gamma(\min_{j \neq i}(v'_j)) - \gamma(v_i) > 0 \Rightarrow U(v'_i) < U(v_i) \end{array} \right.$$

Therefore, the strategy of over-reporting the privacy concern is dominated by the strategy of truth-telling.

On the other hand, if $v'_i < v_i$, that is, if data subject i under-reports her privacy concern, then

$$\left\{ \begin{array}{ll} \text{if } \min_{j \neq i}(v'_j) > v_i, & \text{then, } U(v'_i) = U(v_i) = \gamma(\min_{j \neq i}(v'_j)) - \gamma(v_i) \\ \text{if } \min_{j \neq i}(v'_j) < v'_i, & \text{then, } U(v'_i) = U(v_i) = 0 \\ \text{if } v'_i < \min_{j \neq i}(v'_j) < v_i, & \text{then, } U(v'_i) = \gamma(\min_{j \neq i}(v'_j)) - \gamma(v_i) < 0, U(v_i) = 0 \Rightarrow U(v'_i) < U(v_i) \end{array} \right.$$

Therefore, the strategy of under-reporting the privacy concern is also dominated by the strategy of truth-telling. This means that the mechanism is incentive-compatible, that is, the optimal strategy for all data subjects is truth-telling. ■

Proof of Proposition 2:

Let c_{RSP} be the compensation of the Fixed-Compensation method with equivalent Bias as RSP, which is zero. $B^{FC} = \frac{cn^\alpha - 1}{2}$, therefore, solving $B^{FC}(c = c_{RSP}) = 0$ for c_{RSP} yields $c_{RSP} = n^{-\alpha}$. Therefore, the Total Compensation of this method is $TC^{FC}(c = c_{RSP}) = TC^{FC}(c = n^{-\alpha}) = n^{1-\alpha}$. Total Compensation of RSP is $TC^{RSP} = \frac{n+2[\mathbf{H}_{N+2} - \mathbf{H}_{N+2-n}]}{2n^\alpha}$, thus,

$$TC^{FC}(c = c_{RSP}) - TC^{RSP} = n^{-\alpha} \underbrace{\left[\frac{n}{2} - \mathbf{H}_{N+2} + \mathbf{H}_{N+2-n} \right]}_{\psi}$$

To prove that the Total Compensation of RSP is less than the corresponding FC with equivalent Bias, we need to show that $\psi > 0$. ψ is strictly increasing in $N \in \mathbb{Z}_{>0}$ and $n \in \mathbb{Z}_{>0}$. Substituting both n and N with 1, we can find the lower-bound of ψ as $\frac{1}{2} - \mathbf{H}_3 + \mathbf{H}_2 = \frac{1}{6}$. Therefore, $\psi > 0 \forall \{N, n\} \in \mathbb{Z}_{>0}^2$.

Bias of the Fixed-Compensation method with equivalent Total Compensation as RSP is $B^{FC}(c = \frac{TC^{RSP}}{n}) = B^{FC}(c = \frac{n+2[\mathbf{H}_{N+2} - \mathbf{H}_{N+2-n}]}{2n^{1+\alpha}})$. Therefore,

$$B^{FC}(c = \frac{TC^{RSP}}{n}) = \frac{\mathbf{H}_{N+2} - \mathbf{H}_{N+2-n} - \frac{n}{2}}{2n} = \frac{-\psi}{2n} \neq 0 \quad \forall \{N, n\} \in \mathbb{Z}_{>0}^2$$

Therefore, $|B^{FC}(c = \frac{TC^{RSP}}{n})| > 0$, which means that $|B^{FC}(c = \frac{TC^{RSP}}{n})| > |B^{RSP}|$. ■

Proof of Proposition 3:

First, we compare the Total Cost of RSP with FCH. $\Phi^{FCH} = TC^{FCH} = n^{1-\alpha}$ and $\Phi^{RSP} = TC^{RSP} = \frac{n+2[\mathbf{H}_{N+2} - \mathbf{H}_{N+2-n}]}{2n^\alpha}$, thus,

$$\Phi^{FCH} - \Phi^{RSP} = n^{-\alpha} \underbrace{\left[\frac{n}{2} - \mathbf{H}_{N+2} + \mathbf{H}_{N+2-n} \right]}_{>0 \quad \forall \{N, n\} \in \mathbb{Z}_{>0}^2}$$

Therefore, the Total Cost of RSP is smaller than the Total Cost of FCH, independent of the per unit cost of bias (ω).

Second, we compare the Total Cost of RSP with FCL. $\Phi^{FCL} = TC^{FCL} + \omega|B^{FCL}|$, that is,

$$\Phi^{FCL} = \frac{n^{2-\alpha}}{N+1} + \omega \left| \frac{n-N}{2[N+1]} \right| = \frac{2n^{2-\alpha} + \omega[N-n]}{2[N+1]}$$

Solving $\Phi^{FCL} = \Phi^{RSP}$ for ω yields $\hat{\omega}_L = \frac{[N-2n+1]n+2[N+1][\mathbf{H}_{N+2} - \mathbf{H}_{N+2-n}]}{n^\alpha[N-n+1]}$.

Comparing $\hat{\omega}_L$ with $\hat{\omega}$, we have

$$\hat{\omega} - \hat{\omega}_L = 2n^{1-\alpha} - \hat{\omega}_L = \frac{2[N+1]}{n^\alpha[N-n]} \underbrace{\left[\frac{n}{2} - \mathbf{H}_{N+2} - \mathbf{H}_{N+2-n} \right]}_{>0 \quad \forall \{N, n\} \in \mathbb{Z}_{>0}^2}$$

Therefore, $\hat{\omega}_L < \hat{\omega}$. ■

Proof of Proposition 4:

Average cost of each method is calculated as $AC = \frac{TC}{n}$. Average cost of RSP is derived using the approximate value of Total Compensation. From Section 5.4 we know that $E(TC^{RSP}) = \frac{n^{1-\alpha}}{2} + \frac{\mathbf{H}_{N+2} - \mathbf{H}_{N+2-n}}{n^\alpha}$, and $\mathbf{H}_x - \mathbf{H}_y \approx \ln(x) - \ln(y) - \frac{[x-y]}{2xy}$. Therefore, $TC^{RSP} \approx \frac{n + \frac{1}{N+2} - \frac{1}{N+2-n} + 2[\ln(N+2) - \ln(N+2-n)]}{2n^\alpha}$. Average costs of the methods are as follows:

$$\begin{aligned} AC^{FCH} &= \frac{TC^{FCH}}{n} = \frac{1}{n^\alpha} \\ AC^{FCL} &= \frac{TC^{FCL}}{n} = \frac{1}{n^{1-\alpha}[N+1]} \\ AC^{RSP} &= \frac{TC^{RSP}}{n} \approx \frac{n + \frac{1}{N+2} - \frac{1}{N+2-n} + 2[\ln(N+2) - \ln(N+2-n)]}{2n^{\alpha+1}} \end{aligned}$$

To prove Proposition 4(a), it is sufficient to show that the partial derivative of average cost with respect to n is less than zero for FCH and greater than zero for FCL:

$$\frac{\partial AC^{FCH}}{\partial n} = \frac{\alpha}{n^{\alpha+1}} < 0 \quad \forall n \in \mathbb{Z}_{>0} \text{ \& } \alpha \in [0, 1]$$

$$\frac{\partial AC^{FCL}}{\partial n} = \frac{1 - \alpha}{n^\alpha [N + 1]} > 0 \quad \forall \{N, n\} \in \mathbb{Z}_{>0}^2 \text{ \& } \alpha \in [0, 1]$$

To prove Proposition 4(c), we first derive the partial derivative of average cost with respect to n for RSP:

$$\begin{aligned} \frac{\partial AC^{RSP}}{\partial n} &\approx \frac{n[1 - \frac{1}{[N+2-n]^2} + \frac{2}{N+2-n}] - [\alpha + 1][n + \frac{1}{N+2} - \frac{1}{N+2-n} + 2[\ln(N+2) - \ln(N+2-n)]]}{2n^\alpha} \\ \Rightarrow \frac{\partial AC^{RSP}}{\partial n} &\begin{cases} > 0, & \text{if } \alpha < \check{\alpha} \\ < 0, & \text{if } \alpha > \hat{\alpha} \end{cases}, \text{ where } \check{\alpha} \approx \frac{1}{N[N-1]} \text{ and } \hat{\alpha} \approx \frac{3}{4} \\ \text{if } \check{\alpha} < \alpha < \hat{\alpha}, &\text{ then } \frac{\partial AC^{RSP}}{\partial n} \begin{cases} < 0, & \text{if } \alpha < \tilde{n} \\ > 0, & \text{if } \alpha > \tilde{n} \end{cases}, \text{ where } \tilde{n} \approx N - \frac{1}{\alpha N}. \quad \blacksquare \end{aligned}$$

Proof of Proposition 5:

We know that $B^{COH} = B^{RSP} = 0$, therefore, $\Phi^{COH} = TC^{COH}$ and $\Phi^{RSP} = TC^{RSP}$. Let $\tau = \Phi^{COH} - \Phi^{RSP} = TC^{COH} - TC^{RSP}$, then,

$$\tau = n^{1-\alpha} \left[\frac{1+d}{2} \right] - \frac{n + 2[\mathbf{H}_{N+2} - \mathbf{H}_{N+2-n}]}{2n^\alpha} = \frac{nd - 2[\mathbf{H}_{N+2} + \mathbf{H}_{N+2-n}]}{2n^\alpha}$$

Solving $\tau(d = \hat{d}) = 0$ yields $\hat{d} = \frac{2[\mathbf{H}_{N+2} + \mathbf{H}_{N+2-n}]}{n}$. τ is increasing in d , therefore $\tau(d > \hat{d}) > 0$. This means that for any $d > \hat{d}$, the Total Compensation and Total Cost of COH is larger than the Total Compensation and Total Cost of RSP. \blacksquare