A   Further discussion on BPP payment

In this section, we discuss the connection of bonus-penalty payment and existing peer prediction mechanisms. First, if we substitute the third input with a uniformly random bit, denoted as $\hat{s}_k = Z \sim_u \{-1, 1\}$, the bonus-penalty payment simplifies to the *agreement mechanism* [62, 61, 63], one of the most basic peer prediction mechanisms,

$$\mathbb{E}\left[U^{BPP}(\hat{s}_i, \hat{s}_j, Z)\right] = \hat{s}_i \hat{s}_j = 2\mathbf{1}[\hat{s}_i = \hat{s}_j] - 1.$$

However, the agreement mechanism is not symmetrically strongly truthful, as all agents always reporting 1 and −1 can result in higher payments than truth-telling.

The bonus-penalty payment eq. (1) is originally proposed by [11, 57] for the multi-task setting. Our BPP mechanism in Mechanism 3 can be seen as a generalization of multi-task setting. In the multi-task setting, agents works on multiple tasks and for each task the private signals are jointly identically and independently (iid) sampled from a fixed distribution and the each agent's strategy also are iid. Take two agents (Isabel and Julia) and two tasks as an example: Isabel has a private signal (s_i^1, s_i^2) and reports $(\hat{s}_i^1, \hat{s}_i^2)$ and Julia has (s_j^1, s_j^2) and reports $(\hat{s}_j^1, \hat{s}_j^2)$ where (s_i^l, s_j^l) are iid from random vector (S_i, S_j). Isabel and Julia decide their reports on each task using random function $\sigma_i, \sigma_j : \{-1, 1\} \mapsto \{-1, 1\}$ respectively. Dasgupta and Ghosh [11] use the following payments for Isabel

$$\mathbf{1}[\hat{s}_i^1 = \hat{s}_j^1] - \mathbf{1}[\hat{s}_i^1 = \hat{s}_j^2] = \frac{1}{2}U^{BPP}\left(\hat{s}_i^1, \hat{s}_j^1, \hat{s}_j^2\right).$$

The payment is a special case of Mechanism 3 by taking the second input as $\hat{s}_j^1$ and the third input as $\hat{s}_j^2$. Additionally, S_j^1 uniform dominates S_j^2 for S_i^1 if and only if

$$\Pr[S_j = 1 \mid S_i = 1] > \Pr[S_j = 1], \text{ and } \Pr[S_j = -1 \mid S_i = -1] > \Pr[S_j = -1]$$

which is called *categorical signal distributions* [57].

Finally, similar to Shnayder et al. [57], we may extend to non-binary signal setting by extending the payment to

$$U^{BPP}(\hat{s}_i, \hat{s}_j, \hat{s}_k) = 2\left(\mathbf{1}[\hat{s}_i = \hat{s}_j] - \mathbf{1}[\hat{s}_i = \hat{s}_k]\right)$$

and the definition of uniform dominance to the following.

Definition A.1. Given a random vector $(S_i, S_j, S_k) \in \Omega^3$ on a discrete domain, we say S_j *uniformly dominates* S_k for S_i if

$$\Pr[S_j = s \mid S_i = s] - \Pr[S_k = s \mid S_i = s] > 0 \text{ and}$$
$$\Pr[S_j = s' \mid S_i = s] - \Pr[S_k = s' \mid S_i = s] < 0$$

for all $s, s' \in \Omega$ with $s \neq s'$.

However, the guarantee for truth-telling (*informed truthfulness*) is weaker than the binary setting.

Theorem A.2. *Given any discrete domain Ω, if for each agent i the associated agent j's signal uniformly dominates k's signal for i's signal (definition A.1), Mechanism 3's scheme is symmetrically informed truthful so that*

1. *truth-telling is a strict equilibrium, and*

2. *each agent's expected payment in truth-telling is no less than the payment in any other symmetric equilibria and strictly better than any uninformed equilibrium's.*

Proof. First truth-telling is a strict equilibrium, because if $S_i = s$,

$$\arg\max_{\hat{s}} \mathbb{E}\left[U^{BPP}(\hat{s}, S_j, S_k) \mid S_i = s\right]$$
$$= \arg\max_{\hat{s}} \Pr[S_j = \hat{s} \mid S_i = s] - \Pr[S_k = \hat{s} \mid S_i = s]$$
$$= s \qquad\qquad\qquad\qquad\qquad \text{(by definition A.1)}$$

15

Additionally, because $\Pr[S_j = s \mid S_i = s] - \Pr[S_k = s \mid S_i = s] > \Pr[S_j = s' \mid S_i = s] - \Pr[S_k = s' \mid S_i = s]$ for all $s' \neq s$, summing over all possible $s' \in \Omega$ on both sides gets $\Pr[S_j = s \mid S_i = s] - \Pr[S_k = s \mid S_i = s] > 0$ and

$$\mathbb{E}\left[U^{BPP}(S_i, S_j, S_k)\right] > 0.$$

For any informed equilibrium, by a direct computation $\mathbb{E}\left[U^{BPP}(\hat{S}_i, \hat{S}_j, \hat{S}_k)\right] = 0$.

Finally, we show that the truth-telling has the maximum expected payment for each agents. When all agent use a strategy $\sigma : \Omega \to \Omega$, agent i's expected payment is

$$\sum_{s_i, \hat{s}_i \in \Omega} \Pr[S_i = s_i] \sigma(s_i, \hat{s}_i) \mathbb{E}\left[U^{BPP}(\hat{s}_i, \hat{S}_j, \hat{S}_k) \mid S_i = s_i\right]$$

$$= 2 \sum_{s_i, \hat{s}_i \in \Omega} \Pr[S_i = s_i] \sigma(s_i, \hat{s}_i) \sum_{s \in \Omega} (\Pr[S_j = s \mid S_i = s_i] - \Pr[S_k = s \mid S_i = s_i]) \sigma(s, \hat{s}_i)$$

$$= 2 \sum_{s_i \in \Omega} \Pr[S_i = s_i] \sum_{\hat{s}_i, s \in \Omega} \sigma(s_i, \hat{s}_i) \sigma(s, \hat{s}_i) (\Pr[S_j = s \mid S_i = s_i] - \Pr[S_k = s \mid S_i = s_i])$$

Let $f_{s_i}(s) := \sum_{\hat{s}_i \in \Omega} \sigma(s_i, \hat{s}_i) \sigma(s, \hat{s}_i)$ which is between 0 and 1, because $f_{s_i}(s) \leq \sum_{\hat{s}_i \in \Omega} \sigma(s_i, \hat{s}_i) \sum_{\hat{s}_i \in \Omega} \sigma(s, \hat{s}_i) = 1$. Then the expectation becomes

$$\sum_{s_i, \hat{s}_i \in \Omega} \Pr[S_i = s_i] \sigma(s_i, \hat{s}_i) \mathbb{E}\left[U^{BPP}(\hat{s}_i, \hat{S}_j, \hat{S}_k) \mid S_i = s_i\right]$$

$$= 2 \sum_{s_i \in \Omega} \Pr[S_i = s_i] \sum_{s \in \Omega} \left(\Pr[S_j = s \mid S_i = s_i] - \Pr[S_k = s \mid S_i = s_i]\right) f_{s_i}(s)$$

$$\leq 2 \sum_{s_i \in \Omega} \Pr[S_i = s_i] \left(\Pr[S_j = s_i \mid S_i = s_i] - \Pr[S_k = s_i \mid S_i = s_i]\right)$$

$$= \mathbb{E}\left[U^{BPP}(S_i, S_j, S_k)\right]$$

The inequality holds because $f_{s_i} \in [0, 1]$ and definition A.1. Therefore, we complete the proof. $\square$

B   Proofs in Section 2: Bayesian SST model and other models

The proofs of propositions 2.3 and 2.5 are standard, and variations can be found in related literature. We include proofs here for completeness.

Proof of proposition 2.3. First given $\theta \in \mathbb{R}^{\mathcal{A}}$, for all distinct $a, a', a'' \in \mathcal{A}$, $\Pr[T_\theta(a, a') = 1], \Pr[T_\theta(a', a'') = 1] > 1/2$ implies that $\theta_a - \theta_{a'} > 0$ and $\theta_{a'} - \theta_{a''} > 0$ becuase F is strictly increasing and $F(0) = 1/2$. Because $\theta_a - \theta_{a''} = \theta_a - \theta_{a'} + \theta_{a'} - \theta_{a''} > \max(\theta_a - \theta_{a'}, \theta_{a'} - \theta_{a''})$, we have

$$\Pr[T_\theta(a, a'') = 1] = F(\theta_a - \theta_{a''})$$
$$> \max F(\theta_a - \theta_{a'}), F(\theta_{a'} - \theta_{a''})$$
$$= \max \Pr[T_\theta(a, a') = 1], \Pr[T_\theta(a', a'') = 1]$$

and thus T_θ is strongly stochastically transitive for all θ with distinct coordinates which happens surely as ν is non-atomic. Finally, since the distribution on θ is exchangeable on each coordinate, $\mathbb{E}\left[\mathbb{E}\left[T_\theta(a, a')\right]\right] = 0$ for all a, a'. $\square$

Proof of proposition 2.5. First given $\theta \in \Theta$, for all distinct $a, a' \in \mathcal{A}$, if the rank of a is higher than a',

$$\Pr[T_\theta(a, a') = 1] = h_\eta(\theta(a') - \theta(a) + 1) - h_\eta(\theta(a') - \theta(a))$$

where $h_\eta(x) = \frac{x}{1 - \exp(-\eta x)}$ by Busa-Fekete et al. [7].

Claim B.1. *For any $\eta > 0$ and $x \in \mathbb{Z}_{>0}$, the difference $h_\eta(x + 1) - h_\eta(x)$ is increasing and larger than $1/2$ where $h_\eta(x) = \frac{x}{1 - \exp(-\eta x)}$.*

16

By claim B.1, $\Pr[T_\theta(a, a') = 1], \Pr[T_\theta(a', a'') = 1] > 1/2$ implies that $\theta(a') - \theta(a) > 0$ and $\theta(a'') - \theta(a') > 0$. Thus, $\theta(a'') - \theta(a) > \max(\theta(a'') - \theta(a'), \theta(a'') - \theta(a'))$, and

$$\Pr[T_\theta(a, a'') = 1] = h(\theta(a'') - \theta(a) + 1) - h(\theta(a'') - \theta(a))$$
$$> \max h(\theta(a'') - \theta(a') + 1) - h(\theta(a'') - \theta(a')), h(\theta(a') - \theta(a) + 1) - h(\theta(a') - \theta(a))$$
$$= \max \Pr[T_\theta(a, a') = 1], \Pr[T_\theta(a', a'') = 1]$$

where the second inequality is due to claim B.1. Therefore, T_θ is strongly stochastically transitive for all θ. Finally, $\mathbb{E}\left[\mathbb{E}\left[T_\theta(a, a')\right]\right] = 0$ for all a, a' since θ is an uniform distribution on rankings. $\quad\square$

Proof of claim B.1. We first prove that the function $h_\eta(x) = \frac{x}{1-\exp(-\eta x)}$ is increasing and strictly convex on $x \geq 0$. Because $h_\eta(x) = \frac{1}{\eta} h_1(\eta x)$, for all η, x, it is sufficient to consider $\eta = 1$. First, $h_1'(x) = \frac{1-(x+1)e^{-x}}{(1-e^{-x})^2} > 0$, so h_1 is increasing. Second, as $h_1''(x) = \frac{e^{-x}((x-2)+(x+2)e^{-x})}{(1-e^{-x})^3}$, to show $h_1''(x) > 0$ for all $x > 0$, it is sufficient to show that $g(x) = (x - 2) + (x + 2)e^{-x} > 0$. Because $g(0) = 0$ and $g'(x) = 1 - (x + 1)e^{-x} > 0$, $g(x) > 0$ for all $x > 0$. Therefore, h_1 is strictly convex.

On the other hand, $h_\eta(x+2) - h_\eta(x+1) > h_\eta(x+1) - h_\eta(x)$ for all x by convexity, and $h_\eta(2) - h_\eta(1) = \frac{1}{1+e^{-\eta}} > \frac{1}{2}$ which completes the proof. $\quad\square$

C   Proofs in Section 3 and 4

C.1   Uniform dominance from Bayesian SST

Proof of lemma 4.2. With a prior similar assumption for Bayesian SST model, we only need to show

$$\Pr[S(a'', a') = 1 \mid S(a, a') = 1] > \Pr[S(a'', a) = 1 \mid S(a, a') = 1], \tag{5}$$

and the other case $\Pr[S(a'', a') = -1 \mid S(a, a') = -1] > \Pr[S(a'', a) = -1 \mid S(a, a') = -1]$ follows by symmetry. To prove eq. (5), we can rewrite the conditional probability in expectations of T_θ.

$$\Pr[S(a'', a') = 1 \mid S(a, a') = 1]$$
$$= \frac{\int \Pr[T_\theta(a'', a') = 1, T_\theta(a, a') = 1 \mid \theta] dP_\Theta}{\int \Pr[T_\theta(a, a') = 1 \mid \theta] dP_\Theta}$$
$$= \frac{\int \Pr[T_\theta(a'', a') = 1 \mid \theta] \Pr[T_\theta(a, a') = 1 \mid \theta] dP_\Theta}{\int \Pr[T_\theta(a, a') = 1 \mid \theta] dP_\Theta} \qquad \text{(conditional independent)}$$
$$= 2\int \Pr[T_\theta(a'', a') = 1 \mid \theta] \Pr[T_\theta(a, a') = 1 \mid \theta] dP_\Theta \qquad \text{(a prior similar)}$$
$$= 2\int \frac{\mathbb{E}\left[T_\theta(a'', a') \mid \theta\right] + 1}{2} \frac{\mathbb{E}\left[T_\theta(a, a') \mid \theta\right] + 1}{2} dP_\Theta \qquad \text{(binary value)}$$
$$= \frac{1}{2}\int \mathbb{E}\left[T_\theta(a'', a') \mid \theta\right] \mathbb{E}\left[T_\theta(a, a') \mid \theta\right] + \mathbb{E}\left[T_\theta(a'', a') \mid \theta\right] + \mathbb{E}\left[T_\theta(a, a') \mid \theta\right] + 1 dP_\Theta$$
$$= \frac{1}{2}\int \mathbb{E}\left[T_\theta(a'', a') \mid \theta\right] \mathbb{E}\left[T_\theta(a, a') \mid \theta\right] + 1 dP_\Theta. \qquad \text{(a prior similar)}$$

Claim C.1. *For any strongly stochastically transitive T_θ on $\mathcal{A}$, and distinct $a, a', a'' \in \mathcal{A}$*

$$\mathbb{E}\left[T_\theta(a, a') \mid \theta\right] \mathbb{E}\left[T_\theta(a'', a') \mid \theta\right] > \mathbb{E}\left[T_\theta(a, a') \mid \theta\right] \mathbb{E}\left[T_\theta(a'', a) \mid \theta\right].$$

With claim C.1, we have

$$\Pr[S(a'', a') = 1 \mid S(a, a') = 1] = \frac{1}{2}\int \mathbb{E}\left[T_\theta(a'', a') \mid \theta\right] \mathbb{E}\left[T_\theta(a, a') \mid \theta\right] + 1 dP_\Theta$$
$$> \frac{1}{2}\int \mathbb{E}\left[T_\theta(a'', a) \mid \theta\right] \mathbb{E}\left[T_\theta(a, a') \mid \theta\right] + 1 dP_\Theta = \Pr[S(a'', a) = 1 \mid S(a, a') = 1].$$

This completes the proof of eq. (5), and thus the uniform dominance. $\quad\square$

17

Proof of claim C.1. We let $Q(\alpha, \alpha') := \mathbb{E}\left[T_\theta(\alpha, \alpha') \mid \theta\right] = 2\Pr[T_\theta(\alpha, \alpha') = 1 \mid \theta] - 1$ for all α, α'. Note that $Q(\alpha, \alpha') > 0$ if and only if $\Pr[T_\theta(\alpha, \alpha') = 1 \mid \theta] > 1/2$ and $Q(\alpha, \alpha') = -Q(\alpha', \alpha)$.

By symmetry, let $Q(a, a') > 0$. It is sufficient to show that

$$Q(a'', a') > Q(a'', a).$$

If $Q(a', a'') > 0$, by definition 2.1 $Q(a, a'') > Q(a', a'') > 0$ so $Q(a'', a') > Q(a'', a)$. Now consider $Q(a', a'') < 0$. If $Q(a'', a) < 0$, $Q(a'', a') > 0 > Q(a'', a)$. If $Q(a'', a) > 0$, we have $Q(a'', a) > 0, Q(a, a') > 0$, and thus $Q(a'', a') > Q(a'', a)$ by definition 2.1 □

C.2 Uniform dominance and weak notions of stochastic transitivity

There are weaker forms of stochastic transitivity, raising the question of whether they are sufficient for uniform dominance as in lemma 4.2. We show that general weak stochastic transitivity is not sufficient. Additionally, we show that although the noisy sorting model from [5] is only weakly stochastically transitive but does not satisfy definition 2.1, it exhibits uniform dominance.

Definition C.2 ([13]). A stochastic comparison function, $T : \mathcal{A}^2 \to \{-1, 1\}$, is *weakly stochastically transitive* if for all $a, a', a'' \in \mathcal{A}$ with $\Pr[T(a, a') = 1] > 1/2$ and $\Pr[T(a', a'') = 1] > 1/2$,

$$\Pr[T(a, a'') = 1] > 1/2.$$

Compared to definition 2.1, the weak stochastic transitivity only require the item a is favorable than a''. Below we provide a simple weakly stochastically transitive example with a prior similar property that does not satisfy the uniform dominance in eq. (5).

Example C.3. Consider the set of three items and Θ consists of all ranking on $\mathcal{A}$ with uniform prior where θ maps each items to its value. Given $\theta \in \Theta$ so that if $\theta(a) > \theta(a') > \theta(a'')$,

$$\Pr[T_\theta(a, a') = 1] = \Pr[T_\theta(a', a'') = 1] = 0.9 \text{ and } \Pr[T_\theta(a, a'') = 1] = 0.6.$$

Note that the model is weakly stochastically transitive, because an item with a larger value is more favorable and the weak stochastic transitivity is reduced to transitivity on the values. However, the model is not strongly stochastically transitive, because $\Pr[T_\theta(a, a'') = 1] = 0.6 < \max\{\Pr[(T(a, a') = 1], \Pr[(T(a', a'') = 1]]\} = 0.9$. Finally, as the rank θ has a uniform prior, the model satisfies a prior similar assumption.

To conclude the example, we show that eq. (5) does not hold for the above model. By direct computation over all six possible ranking θ, we have

$$\Pr[S(a'', a') = 1 \mid S(a, a') = 1]$$

$$= \frac{1}{2}\int \mathbb{E}\left[T_\theta(a'', a') \mid \theta\right]\mathbb{E}\left[T_\theta(a, a') \mid \theta\right] + 1 \, dP_\Theta$$

$$= \frac{1}{2}\left(1 - \frac{64}{6}\right),$$

but $\Pr[S(a'', a) = 1 \mid S(a, a') = 1] = \frac{1}{2}\left(1 + \frac{64}{6}\right)$. Therefore, we have $\Pr[S(a'', a') = 1 \mid S(a, a') = 1] < \Pr[S(a'', a) = 1 \mid S(a, a') = 1]$, and show that eq. (5) does not hold.

Though the above example shows that weak stochastic transitivity is not sufficient.[5] Below we show a popular weakly stochastically transitive model in Braverman and Mossel [5] has uniform dominance as in lemma 4.2.

Example C.4. Let Θ be the set of rankings on $\mathcal{A}$ and $\eta > 0$ be a parameter. Given a uniformly distributed reference ranking $\theta \in \Theta$, the noise ranking model [5] ensures that for all $\theta(a) > \theta(a')$

$$\Pr[T_\theta(a, a') = 1] = \frac{1}{2} + \eta$$

Note that the above model does not satisfy the strict inequality in definition 2.1, but by direct computation, $\Pr[S(a'', a') = 1 \mid S(a, a') = 1] = \frac{1}{2}\left(1 + \frac{4\gamma^2}{3}\right)$ and $\Pr[S(a'', a) = 1 \mid S(a, a') = 1] = \frac{1}{2}\left(1 - \frac{4\gamma^2}{3}\right)$, which satisfies lemma 4.2.

---

[5]In the above example, we can also decrease 0.9 to a smaller number that satisfies both uniform dominance and weak stochastic transitivity.

18

C.3 Symmetrically strongly truthful from uniform dominance

Proof of lemma 4.3. Suppose $S_i = 1$. Because $\Pr[S_j = 1|S_i = 1] > \Pr[S_k = 1|S_i = 1]$, $\Pr[S_j = -1|S_i = 1] < \Pr[S_k = -1|S_i = 1]$. Therefore, $\arg\max_{\hat{s} \in \{-1,1\}} \Pr[S_j = \hat{s}|S_i = 1] - \Pr[S_k = \hat{s}_i|S_i = 1] = 1$. Identical argument holds for the case of $S_i = -1$ which completes the proof.

Additionally, the expected payment of truth-telling is

$$\mathbb{E}\left[U^{BPP}(S_i, S_j, S_k)\right] = \sum_a \Pr[S_i = s_i] \sum_{s_j, s_k} \Pr[S_j = s_j, S_k = s_k \mid S_i = s_i] U^{BPP}(s_i, s_j, s_k)$$

$$= 2 \sum_a \Pr[S_i = s_i] \sum_{s_j, s_k} \Pr[S_j = s_j, S_k = s_k \mid S_i = s_i](\mathbf{1}[s_i = s_k] - \mathbf{1}[s_i = s_k])$$

$$= 2 \sum_a \Pr[S_i = s_i] \left(\Pr[S_j = s_i \mid S_i = s_i] - \Pr[S_k = s_i \mid S_i = s_i]\right)$$

$$> 0$$

The last inequality holds due to definition 4.1. $\square$

Proof of lemma 4.4. As σ is uninformed, let $\mu(s) = \sigma(s,s) = \sigma(-s,s)$ and $\mu(-s) = \sigma(s,-s) = \sigma(-s,-s)$ for all s.

$$\mathbb{E}\left[U^{BPP}(\hat{s}_i, \hat{S}_j, \hat{S}_k) \mid S_i = s_i\right] = \sum_{\hat{s}_j, \hat{s}_k} \mu(\hat{s}_j)\mu(\hat{s}_k)U^{BPP}(\hat{s}_i, \hat{S}_j, \hat{S}_k) = \sum_{\hat{s}_j, \hat{s}_k} \mu(\hat{s}_j)\mu(\hat{s}_k)(\hat{s}_i\hat{s}_j - \hat{s}_i\hat{s}_k) = 0$$

The first equality holds as the reports are independent of signals. $\square$

Proof of lemma 4.5.

$$\mathbb{E}_{P,\sigma}\left[U^{BPP}(\hat{s}_i, \hat{S}_j, \hat{S}_k) \mid S_i = s_i\right]$$

$$= \sum_{s_j, s_k, \hat{s}_j, \hat{s}_k} \Pr[S_j = s_j, S_k = s_k \mid S_i = s_i]\sigma(s_j, \hat{s}_j)\sigma(s_k, \hat{s}_k)U^{BPP}(\hat{s}_i, \hat{s}_j, \hat{s}_k)$$

$$= 2 \sum_{s_j, s_k, \hat{s}_j, \hat{s}_k} \Pr[S_j = s_j, S_k = s_k \mid S_i = s_i]\sigma(s_j, \hat{s}_j)\sigma(s_k, \hat{s}_k)\left(\mathbf{1}[\hat{s}_i = \hat{s}_j] - \mathbf{1}[\hat{s}_i = \hat{s}_k]\right)$$

$$\text{(by eq. (1))}$$

$$= 2 \sum_{s_j, \hat{s}_j} \Pr[S_j = s_j \mid S_i = s_i]\sigma(s_j, \hat{s}_j)\mathbf{1}[\hat{s}_i = \hat{s}_j] - 2 \sum_{s_k, \hat{s}_k} \Pr[S_k = s_k \mid S_i = s_i]\sigma(s_k, \hat{s}_k)\mathbf{1}[\hat{s}_i = \hat{s}_k]$$

$$= 2 \sum_{s, \hat{s}} (\Pr[S_j = s \mid S_i = s_i] - \Pr[S_k = s \mid S_i = s_i])\sigma(s, \hat{s})\mathbf{1}[\hat{s}_i = \hat{s}] \quad \text{(renaming dummy variables)}$$

$$= 2 \sum_{s} (\Pr[S_j = s \mid S_i = s_i] - \Pr[S_k = s \mid S_i = s_i])\sigma(s, \hat{s}_i)$$

Let $\delta = \Pr[S_j = s_i \mid S_i = s_i] - \Pr[S_k = s_i \mid S_i = s_i] > 0$, because S_j uniformly dominates S_k for S_i. Additionally, $\Pr[S_j = -s_i \mid S_i = s_i] - \Pr[S_k = -s_i \mid S_i = s_i] = 1 - \Pr[S_j = s_i \mid S_i = s_i] - 1 + \Pr[S_k = s_i \mid S_i = s_i] = -\delta$. We have

$$\mathbb{E}_{P,\sigma}\left[U^{BPP}(\hat{s}_i, \hat{S}_j, \hat{S}_k) \mid S_i = s_i\right]$$

$$= 2 \sum_{s} (\Pr[S_j = s \mid S_i = s_i] - \Pr[S_k = s \mid S_i = s_i])\sigma(s, \hat{s}_i)$$

$$= 2\delta \left(\sigma(s_i, \hat{s}_i) - \sigma(-a, \hat{s}_i)\right),$$

so $\arg\max_{\hat{s}_i \in \{-1,1\}} \mathbb{E}_{P,\sigma}\left[U^{BPP}(\hat{s}_i, \hat{S}_j, \hat{S}_k) \mid S_i = s_i\right] = \arg\max_{\hat{s}_i \in \{-1,1\}} \{\sigma(s_i, \hat{s}_i) - \sigma(-s_i, \hat{s}_i)\}$ which completes the proof. $\square$

D Proofs in Section 5.1

Before diving into the proof, we introduce some notations. We further introduce Ising models with bias parameter $\alpha \in \mathbb{R}_{\geq 0}^V$ in addition to $\boldsymbol{\beta}$ where

$$H(\mathbf{s}) = \sum_{i,j \in V} \beta_{i,j} s_i s_j + \sum_{i \in V} \alpha_i s_i$$

19

and $\Pr_{\alpha,\beta}[\mathbf{S} = \mathbf{s}] \propto \exp(H(\mathbf{s}))$, for all configuration $\mathbf{s}$. Given $i \in V$, let the expectation and ratio be

$$\nu_i(\alpha, \beta) = \mathbb{E}_{\alpha,\beta}[S_i] = \Pr_{\alpha,\beta}[S_i = 1] - \Pr_{\alpha,\beta}[S_i = -1] \text{ and } \rho_i(\alpha, \beta) = \frac{\Pr_{\alpha,\beta}[S_i = 1]}{\Pr_{\alpha,\beta}[S_i = -1]}$$

respectively which are monotone to each other. We will omit α, β when clear. Given a subset $U \subseteq V$, $\mathbf{s}_U \in \{-1, 1\}^U$ is a configuration over the nodes in U, and $\mathbf{s}_U = 1$ if $x_\iota = 1$ for all $\iota \in U$. We write $\Pr[\cdot | \mathbf{S}_U = \mathbf{s}_U]$, $\nu_{i|\mathbf{S}_U=\mathbf{s}_U}$, and $\rho_{i|\mathbf{S}_U=\mathbf{s}_U}$ for the conditional probability, expectation and ratio when the configuration in U is fixed as specified by $\mathbf{s}_U$.

A lower bound for LHS   Informally, we want to lower bound the correlation between adjacent i and j (friends). Note that as we remove edges (setting coordinates of β to zeros), the correlation should decrease, and the smallest correlation between neighboring nodes i and j happens when $E = \{(i, j)\}$. Lemma D.2 formalizes this idea using the following monotone inequality [44, Theorem 17.2].

Theorem D.1 (Griffiths' inequality). *For any $i \in V$, $\nu_i(\alpha, \beta) = \mathbb{E}_{\alpha,\beta}[S_i]$ is non-negative and non-decreasing in all $\beta_{j,k} \geq 0$ and $\alpha_j \geq 0$ with $j, k \in V$.*

Lemma D.2. *Given V and $i, j \in V$, for all α, β, and β', if $\beta'_e = \beta_e$ when $e = (i, j)$ and $\beta'_e = 0$ otherwise, we have*

$$\nu_{i|S_j=1}(\alpha, \beta) \geq \nu_{i|S_j=1}(\alpha, \beta') \text{ and } \rho_{i|S_j=1}(\alpha, \beta) \geq \rho_{i|S_j=1}(\alpha, \beta').$$

Proof. First, note that we can write the conditional expectation $\mathbb{E}_{\alpha,\beta}\left[S_i \mid S_j = 1\right]$ as marginal expectation. Formally, consider α^η so that $\alpha_\iota^\eta = \alpha_\iota$ if $\iota \neq j$ and $\alpha_j^\eta = \alpha_j + \eta$. Because $\eta \rightarrow \alpha^\eta$ is non-decreasing, $\eta \rightarrow \nu_i(\alpha^\eta, \beta)$ is non-decreasing by theorem D.1. In addition, $\Pr_{\alpha^\eta,\beta}[S_i \mid S_j = s] = \Pr_{\alpha,\beta}[S_i \mid S_j = s]$ for all s, and $\Pr_{\alpha^\eta,\beta}[S_j = -1] = O(e^{-2\eta})$, so

$$\nu_{i|S_j=1}(\alpha, \beta) = \mathbb{E}_{\alpha,\beta}[S_i \mid S_j = 1] = \lim_{\eta\to+\infty} \nu_i(\alpha^\eta, \beta).$$

Similarly,

$$\nu_{i|S_j=1}(\alpha, \beta') = \mathbb{E}_{\alpha,\beta'}[S_i \mid S_j = 1] = \lim_{\eta\to+\infty} \nu_i(\alpha^\eta, \beta').$$

On the other hand, consider β^λ so that $\beta_e^\lambda = \beta_e$ if $e \neq (i, j)$ and $\beta_{i,j}^\lambda = \beta_{i,j} + \lambda$. By theorem D.1, $\nu_i(\alpha^\eta, \beta^\lambda)$ is non-decreasing in λ for all η. Because $\beta^0 = \beta'$ and $\beta^1 = \beta$, we have

$$\nu_{i|S_j=1}(\alpha, \beta') = \lim_{\eta\to+\infty} \nu_i(\alpha^\eta, \beta') \leq \lim_{\eta\to+\infty} \nu_i(\alpha^\eta, \beta) = \nu_{i|S_j=1}(\alpha, \beta)$$

Because $\rho_i = \frac{1+\nu_i}{1-\nu_i}$ is monotone in ν_i, the second part follows. $\qquad\square$

Given β' defined in lemma D.2, by some direct computation with $\alpha = 0$

$$\rho_{i|S_j=1}(\alpha, \beta) \geq \rho_{i|S_j=1}(\alpha, \beta') = e^{2\alpha_i+2\beta_{i,j}} = e^{2\underline{\beta}}. \tag{6}$$

An upper bound for RHS   Now, we need to upper bound the correlation between non-adjacent i and k (non-friends). We will use Weitz's self-avoiding walks reduction [65] to upper bound the correlation on general graph G by the correlation on trees.

Given a general graph G, and an arbitrary node i, we can construct the Self Avoiding Walk Tree of G rooted at i, denoted $T_{SAW}(G, i)$, so that $\Pr[S_i = 1 \mid \mathbf{S}_U = \mathbf{s}_U]$ is the same in G as in the tree. We outline the construction. $T_{SAW}(G, i)$ enumerates all self-avoiding walks in G starting at i which terminates when it revisits a previous node (closes a cycle). Then, $T_{SAW}(G, i)$ introduces a leaf with a certain boundary condition. The self-avoiding walk never revisits a node immediately, so there all the leaves with fixed boundary conditions are at least three hops away from node i. Note that if G has maximum degree d, T_{SAW} is a d-ary tree.

Theorem D.3 ([65]). *For any α, β, node $i \in V$, and configuration $\mathbf{s}_U$ on $U \subset V$,*

$$\Pr_{\alpha,\beta}[S_i = 1 \mid \mathbf{S}_U = \mathbf{s}_U] = \Pr_{T_{SAW}(G,i)}[S_i = 1 \mid \mathbf{S}_U = \mathbf{s}_U].$$

20

First, with the above theorem, we have $\nu_{i|S_k=1}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbb{E}_{\alpha,\beta}[S_i \mid S_k = 1] = \mathbb{E}_{T_{SAW}(G,i)}[S_i \mid S_k = 1]$. By the monotone property in theorem D.1, setting all two-hop neighbors U in $T_{SAW}(G,i)$ to 1 (recalled that any boundary conditions for $T_{SAW}(G,i)$ being at least three hops away) increases the conditional expectation,

$$\mathbb{E}_{T_{SAW}(G,i)}[S_i \mid S_k = 1] \le \mathbb{E}_{T_{SAW}(G,i)}[S_i \mid \mathbf{S}_U = 1, S_k = 1].$$

Let T be the tree by truncating $T_{SAW}(G,i)$ at level 2. By the Markov property of Ising models, the expectation is equal to the expectation on T.

$$\mathbb{E}_{\alpha,\beta}[S_i \mid S_k = 1] \le \mathbb{E}_{T_{SAW}(G,i)}[S_i \mid \mathbf{S}_U = 1] = \mathbb{E}_T[S_i \mid \mathbf{S}_U = 1]. \tag{7}$$

Finally, we can recursively compute the probability ratio ρ_i (and thus expectation ν_i) on trees. Specifically, given a rooted tree T', we define $\rho_{T'}$ as the ratio of probabilities for the root to be $+1$ and -1 respectively, and $\rho_{T'|\mathbf{S}_U=\mathbf{s}_U}$ for the ratio of conditional probabilities. As stated in the following lemma, it is well known (see, for example, [22]) that the ratio of each node can be computed recursively over the children's ratio.

Lemma D.4. *Given a tree T rooted at i with parameter $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and boundary condition $\mathbf{s}_U$,*

$$\rho_{T|\mathbf{S}_U=\mathbf{s}_U} = e^{2\alpha_i} \prod_{l=1}^{d} \frac{\rho_{T_l|\mathbf{S}_U=\mathbf{s}_U} e^{2\beta_{i,j_l}} + 1}{e^{2\beta_{i,j_l}} + \rho_{T_l|\mathbf{S}_U=\mathbf{s}_U}}$$

where $j_1, \ldots, j_d$ are children of i and T_l is the subtree rooted at j_l for all l.

By the monotone property in theorem D.1, the maximum of right-hand side of eq. (7) happens when T is a complete d-ary tree with $\boldsymbol{\beta} = \overline{\beta}$. Therefore,

$$\rho_{i|S_k=1}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \le \left(\frac{e^{2(d+1)\overline{\beta}} + 1}{e^{2\overline{\beta}} + e^{2d\overline{\beta}}} \right)^d. \tag{8}$$

Finally, with eqs. (6) and (8), we have $\rho_{i|S_j=1}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \ge e^{2\underline{\beta}} \ge \left(\frac{e^{2(d+1)\overline{\beta}}+1}{e^{2\overline{\beta}}+e^{2d\overline{\beta}}} \right)^d \ge \rho_{i|S_k=1}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ which implies eq. (4).

Remark D.5. Note that for any graph G there exists small enough $\overline{\beta}, \underline{\beta}$ so that the condition in theorem 5.1 is satisfied, because the inequality become equality when $\overline{\beta} = \underline{\beta} = 0$, and $\frac{\partial}{\partial \beta} \frac{2\beta}{d} > 0 = \frac{\partial}{\partial \beta} \ln \frac{e^{2(d+1)\beta}+1}{e^{2\beta}+e^{2d\beta}}$.

The bound between β and d is necessary as shown in fig. 3. On the other hand, by the Markov property of the Ising model, the majority of all neighbor's signals is a sufficient statistic, and we can show the majority of all neighbor's signals are uniformly dominant to a non-neighbor's signal. Therefore, we can get a symmetrically strongly truthful mechanism by replacing j's reports with the majority of reports from i's neighbors.

E Proof of Theorem 5.3

The sufficient condition is done by lemma 4.3, because

$$\begin{aligned} &\underset{\hat{s}_i \in \{-1,1\}}{\arg\max} \mathbb{E}\left[\lambda U^{BPP}(\hat{s}_i, S_j, S_k) + \mu(S_j, S_k) \mid S_i = s_i \right] \\ &= \underset{\hat{s}_i \in \{-1,1\}}{\arg\max} \mathbb{E}\left[\lambda U^{BPP}(\hat{s}_i, S_j, S_k) \mid S_i = s_i \right] \\ &= \underset{\hat{s}_i \in \{-1,1\}}{\arg\max} \mathbb{E}\left[U^{BPP}(\hat{s}_i, S_j, S_k) \mid S_i = s_i \right] && (\lambda > 0) \\ &= s_i && \text{(by lemma 4.3)} \end{aligned}$$

For the necessary, given U, define $D(s_j, s_k) = \frac{1}{2}\left(U(1, s_j, s_k) - U(-1, s_j, s_k) \right)$ and $\mu(s_j, s_k) = \frac{1}{2}(U(1, s_j, s_k) + U(-1, s_j, s_k))$ for all s_j and s_k in $\{-1, 1\}$. Hence

$$U(s_i, s_j, s_k) = s_i \cdot D(s_j, s_k) + \mu(s_j, s_k), \forall s_i, s_j, s_k \in \{-1, 1\} \tag{9}$$

21
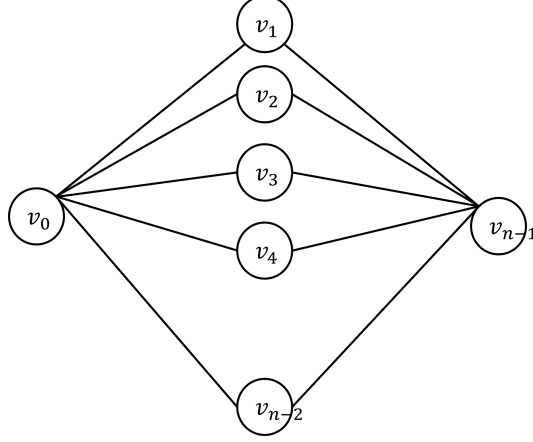
Figure 3: As fixing any $\beta, \bar{\beta}$, we can construct a simple graph with $V = \{v_0, \ldots, v_{n-1}\}$ and $E = \{(v_0, v_l), (v_l, v_{n-1}) : l = 1, \ldots, n-2\}$ where agent v_0 and v_{n-1} are not connected but share $n-2$ common friends. We can show that the correlation between S_0 and S_{n-1} converge to 1 as the number of common friends d increases, while the correlation between S_0 and S_1 is bounded away from 1.

Given a joint distribution satisfying definition 4.1, we let $p^{s_i}(s_j, s_k) = \Pr[S_j = s_j, S_k = s_k \mid S_i = s_i]$ and additionally write $p^{s_i} = \begin{bmatrix} p^{s_i}(1,1) & p^{s_i}(1,-1) \\ p^{s_i}(-1,1) & p^{s_i}(-1,-1) \end{bmatrix}$. Then definition 4.1 ensures that

$$p^1(1,-1) > p^1(-1,1) \text{ and } p^{-1}(1,-1) < p^{-1}(-1,1).$$

Because U is truthful for all uniformly dominant tuples, we have

$$\begin{aligned} 0 &< \mathbb{E}\left[U(1, S_j, S_k) \mid S_i = 1\right] - \mathbb{E}\left[U(-1, S_j, S_k) \mid S_i = 1\right] = 2 \sum_{s_j, s_k} D(s_j, s_k) p^1(s_i, s_j) \\ 0 &> \mathbb{E}\left[U(1, S_j, S_k) \mid S_i = -1\right] - \mathbb{E}\left[U(-1, S_j, S_k) \mid S_i = -1\right] = 2 \sum_{s_j, s_k} D(s_j, s_k) p^{-1}(s_i, s_j). \end{aligned} \tag{10}$$

Suppose the following are true

$$D(1,-1) = -D(-1,1) > 0 \tag{11}$$
$$D(1,1) = D(-1,-1) = 0 \tag{12}$$

Let $\lambda = D(1,-1) > 0$. By eqs. (11) and (12), we have

$$\begin{aligned} U(s_i, s_j, s_k) &= s_i \cdot D(s_j, s_k) + \mu(s_j, s_k) && \text{(by eq. (9))} \\ &= \lambda \cdot s_i(s_j - s_k) + \mu(s_j, s_k) && \text{(by eqs. (9) and (11))} \end{aligned}$$

which completes the proof. Thus, we will construct three joint distributions satisfying definition 4.1 to prove eqs. (11) and (12).

The first joint distribution $p_1^{s_i}(s_j, s_k)$ with $0 < \delta \le 1/2$

$$p^1 = \begin{bmatrix} 0 & 1/2 + \delta \\ 1/2 - \delta & 0 \end{bmatrix} \text{ and } p^{-1} = \begin{bmatrix} 0 & 1/2 - \delta \\ 1/2 + \delta & 0 \end{bmatrix}.$$

Then eq. (10) on the first distribution reduces to

$$0 < D(1,-1)p_1^1(1,-1) + D(-1,1)p_1^1(-1,1) = \frac{1}{2}(D(1,-1) + D(-1,1)) + \delta(D(1,-1) - D(-1,1))$$

$$0 > D(1,-1)p_1^{-1}(1,-1) + D(-1,1)p_1^{-1}(-1,1) = \frac{1}{2}(D(1,-1) + D(-1,1)) - \delta(D(1,-1) - D(-1,1)).$$

As we take δ to zero, we prove $D(1,-1) = -D(-1,1)$. Then plugging in with nonzero δ, we have $D(1,-1) > 0$ and complete the proof of eq. (11).

22

The second joint distribution $p_2^{s_i}(s_j, s_k)$ with $0 \leq \epsilon \leq 1$ is

$$p^1 = \begin{bmatrix} 1-\epsilon & \frac{3}{4}\epsilon \\ \frac{\epsilon}{4} & 0 \end{bmatrix} \text{ and } p^{-1} = \begin{bmatrix} 1-\epsilon & \frac{\epsilon}{4} \\ \frac{3\epsilon}{4} & 0 \end{bmatrix}.$$

With eq. (11), eq. (10) reduces to

$$0 < (1-\epsilon)D(1,1) + \frac{\epsilon}{4}(D(1,-1) - D(-1,1))$$

$$0 > (1-\epsilon)D(1,1) - \frac{\epsilon}{4}(D(1,-1) - D(-1,1)).$$

By taking ϵ to zero, we prove $D(1,1) = 0$. We can prove $D(-1,-1) = 0$ using the similar trick and complete the proof of eq. (12).

F   Additional empirical results

F.1   Comparison data

Here we test if the dataset satisfy transitivity property. We denote the proportion of rankings such that item a is higher than item a' in the dataset by $p_{a>a'}$. If $p_{a>a'} > 1/2$, $p_{a'>a''} > 1/2$, and $p_{a>a''} > 1/2$, we say the triple of items $\{a, a', a''\}$ empirically satisfies transitivity. If $p_{a>a'} > 1/2$, $p_{a'>a''} > 1/2$, and $p_{a>a''} > \max\{p_{a>a'}, p_{a'>a''}\}$, we say the triple of items $\{a, a', a''\}$ empirically satisfies strong transitivity. We first test the transitivity of the SUSHI subdataset selected in section 6.1. We find that 100% of the item triples empirically satisfy transitivity, and 69.17% of the item triples empirically satisfy strong transitivity. This suggests that our transitivity assumption for the comparison data is mostly aligned.

Moreover, we conducted an experiment on the entire SUSHI dataset without any selection criteria and demonstrated the results in fig. 4. Observe that the ECDF of payments from original human users also dominates the payments under the uninformed strategy and the unilateral deviating strategy. This is consistent with our experimental results in section 6.1. However, there are two minor difference. First the separation of truth-telling from the other two is slightly less prominent than fig. 1 with the selection criteria. This may be due to a slightly lower degree of transitivity across agents with different backgrounds. In particular, we found the average value of $p_{a>a''} - \max\{p_{a>a'}, p_{a'>a''}\}$ is 0.0559 without the selection criteria which is less than 0.0604 with the selection criteria in fig. 1. Second, the fraction of agents receiving positive payments is slightly higher than in fig. 1 (0.785 and 0.763 respectively). This aligned with or empirical (strong) transitively which are 1 and 0.7667 compared to the above 1 and 0.69117. Furthermore, we also conducted experiments on other groups of users by changing the selection criteria. Those interested can refer to fig. 5, fig. 6 and table 1 for the results, which further verify the effectiveness of our mechanism.

Selection criteria	Number of users	Average utility	Fraction of positive utility
All (No selection)	5000	0.138	78.5%
Female, 30-49, Kanto/Shizuoka	249	0.137	76.3%
Male, 30-49, Kanto/Shizuoka	185	0.167	82.2%
Female, 5-29, Kanto/Shizuoka	146	0.175	84.2%
Female, 50+, Kanto/Shizuoka	26	0.13	80.8%
Female, 30-49, Tohoku	30	0.174	83.3%
Female, 30-49, Hokuriku	23	0.105	69.6%

Table 1: Summary of truth-telling utility in appendix F.1.

F.2   Networked data

Alongside fig. 2, Figure 7 and table 2 present empirical results for the top five popular artists in the dataset, excluding Lady Gaga, who are Britney Spears, Rihanna, The Beatles, and Katy Perry. All these settings show similar results. However, the Beatles' data is less conclusive as the payment distribution under the uninformed strategy profile is close to the truth-telling. This observation is also documented in Daskalakis et al. [12] which notes that the Ising model performs much better for

23
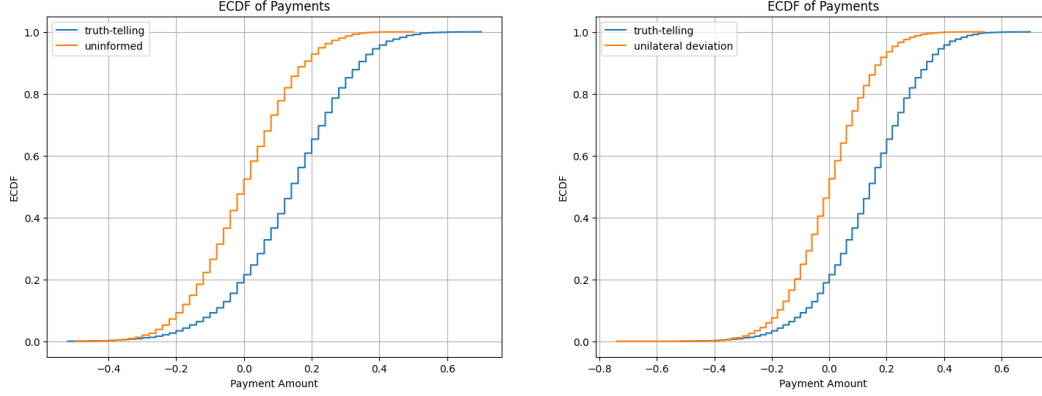
Figure 4: ECDF comparisons on all users without any selection.

rock artists than for pop artists. The authors conjecture that this may be due to the highly divisive popularity of pop artists like Lady Gaga and Britney Spears, whose listeners may form dense cliques within the graph.

Note that there is a buck of agent with a payment of around 0.5 under the truth-telling. This is because many non-listeners have no listener friends, and payment is $1 - [(1 - p) - p] = 2p$ is twice the popularity $p \approx 0.25$. Moreover, the jump is most minor for the Beatles, and indicates less agreement between non-listeners. Additionally, by the definition of bonus-penalty payment, we can see the payment of deviation is the minus of the truthful payment, so that the ECDF is symmetric around $(0, 0.5)$.

Artists	Fraction of listener	Average utility	Fraction of positive utility
Lady Gaga	32.2%	0.37	76%
Britney Spears	27.6%	0.420	82.6%
Rihanna	25.6%	0.422	83.4%
The Beatles	25.4%	0.137	68.5%
Katy Perry	25.0%	0.361	79.9%

Table 2: Summary of truth-telling utility in appendix F.2.

Figure 8 further shows the scatter plot of average payment and fraction of agents with positive payments across the top fifty popular artists where all settings have more than 60% percent of agents get positive payment. However, for less popular artists, the performance of our mechanism declines. This is expected, as we cannot provide effective incentives when only one agent listens to an artist.

24
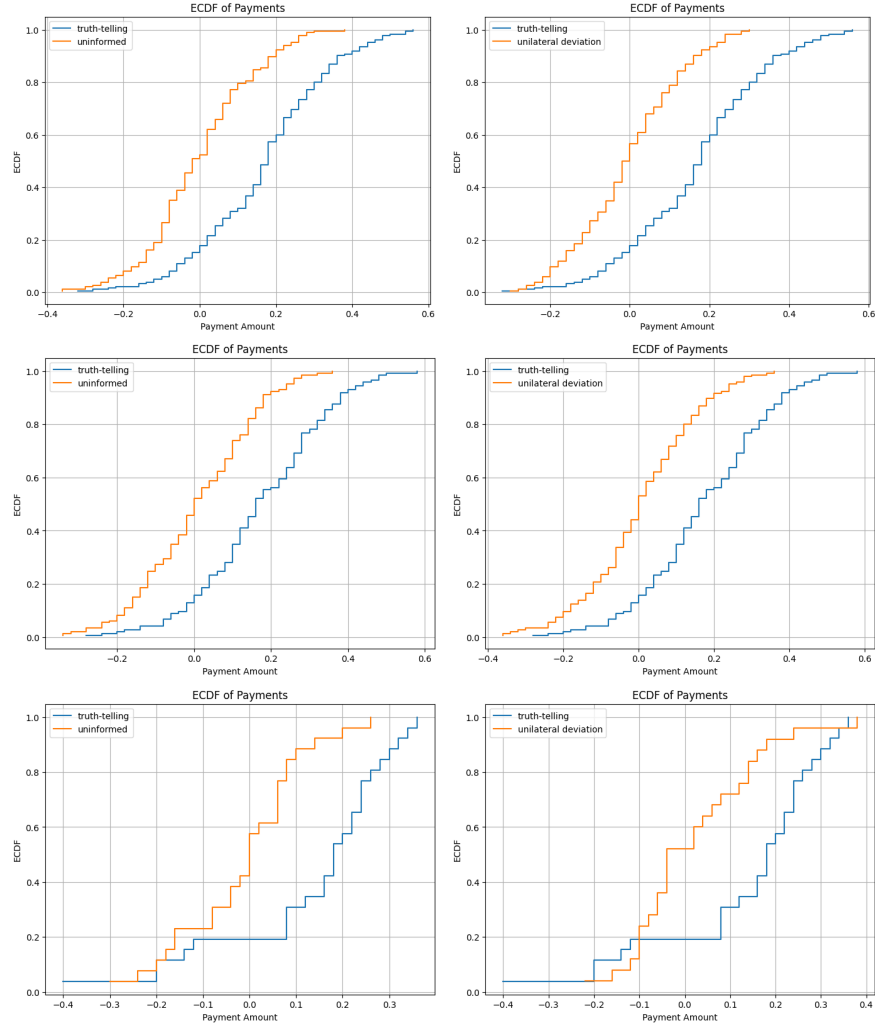
Figure 5: In each of the rows, we present the ECDF comparisons after changing the selection criteria for the user group as follows: from female to male, from ages 30–49 to ages 5–29, from ages 30–49 to ages 50+, respectively.

25
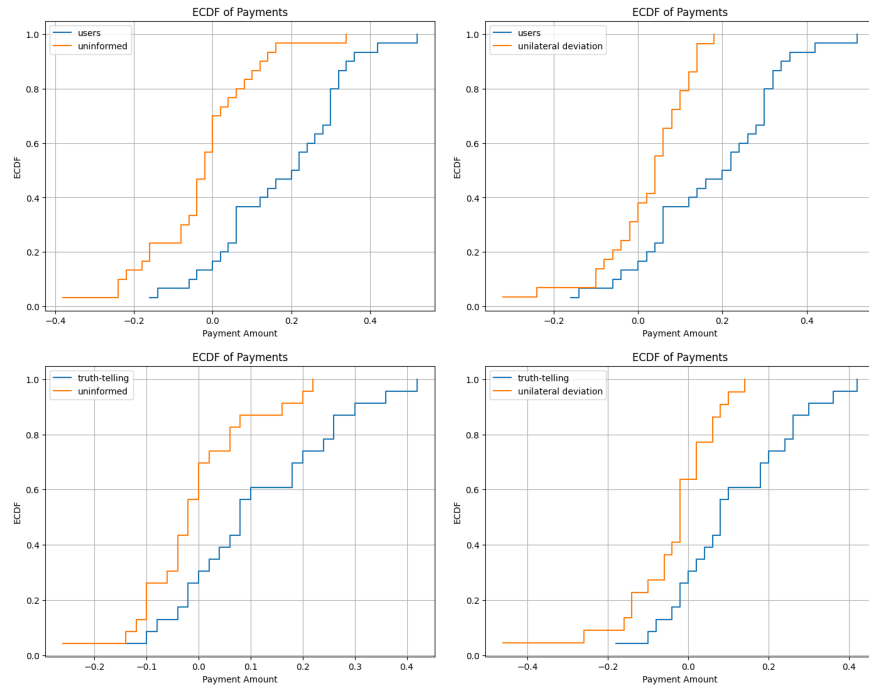
Figure 6: In each of the rows, we present the ECDF comparisons after changing the location criteria for the user group as follows: from mostly living in Kanto or Shizuoka to Tohoku until age 15, and from mostly living in Kanto or Shizuoka to Hokuriku until age 15, respectively.

26
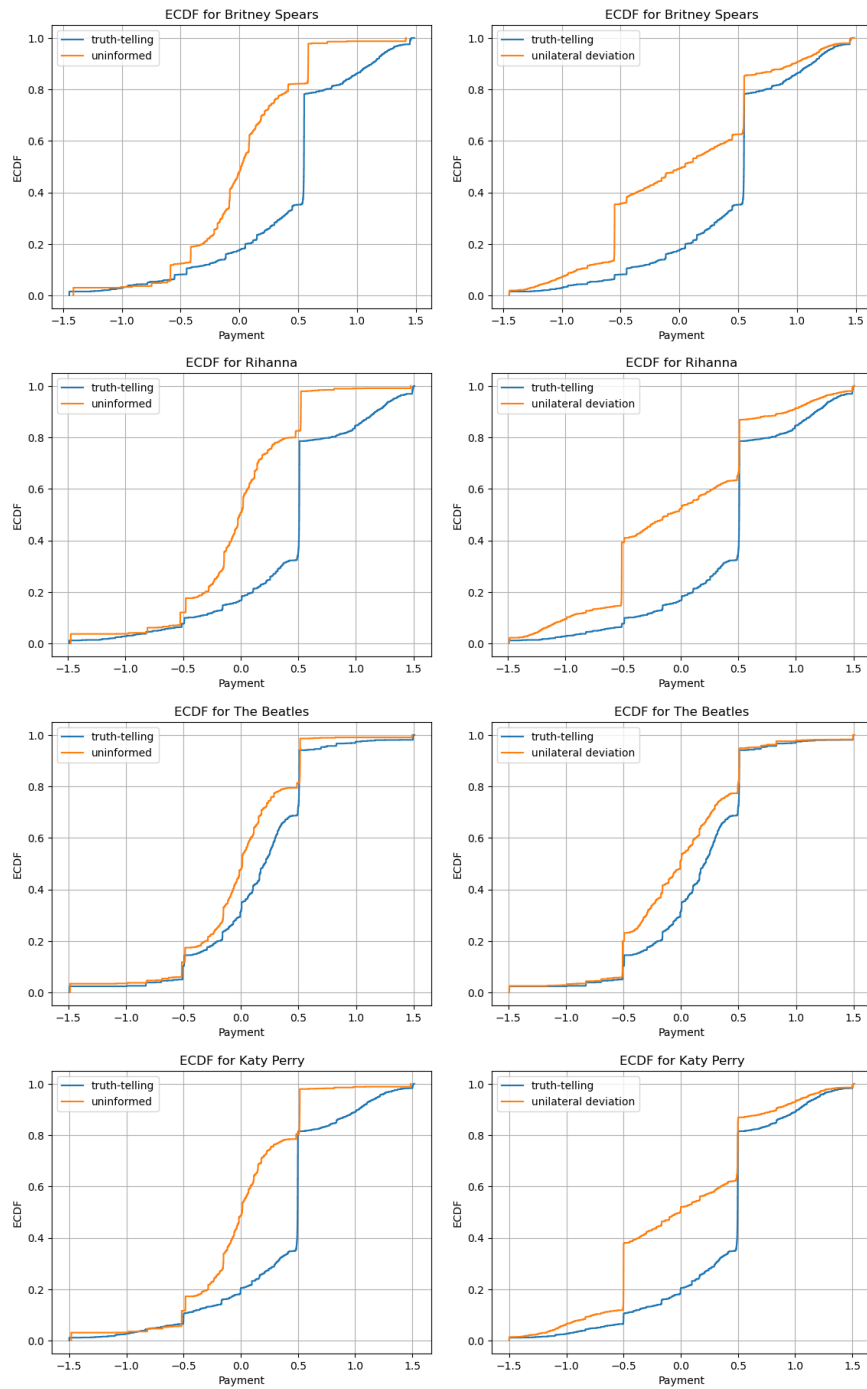
Figure 7: Last.fm dataset for other top five popular artists excluding Lady Gaga.

27
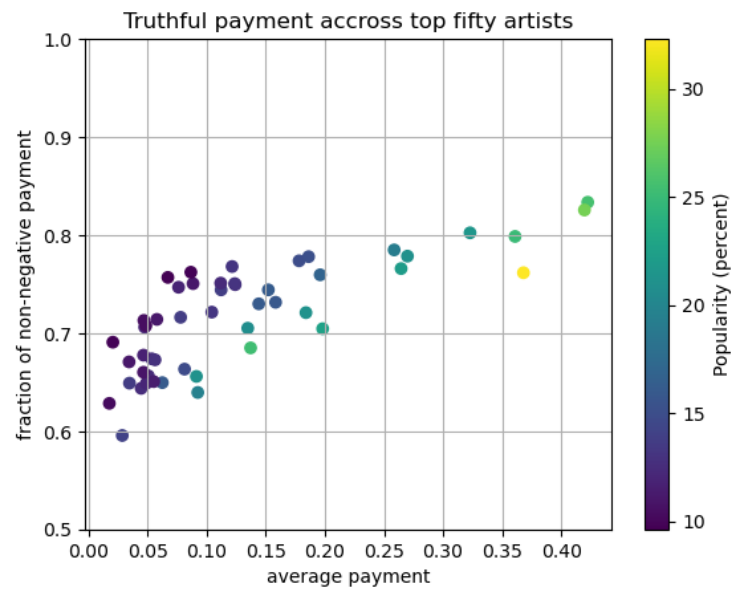
Figure 8: Average payment and fraction of positive payment under the truth-telling across top fifty popular artists.

28

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS paper checklist",**
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our claims accurately reflect the contributions and scope of the paper.

   Guidelines:
   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In section 7, we discuss potential future research directions, which are the limitations of our current work.

29

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: We present all the assumptions. The complete proofs are provided in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The code is uploaded in the supplementary material. All the information required to reproduce the experimental results is provided.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

30

- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code is uploaded in the supplementary material.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

31

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Most of these details are explained in section 6 and in the provided code.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: We believe the error bars are not relevant to our empirical metrics.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: We believe the computer resources are not relevant to our main contributions.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

32

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The research conducted in our paper conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: Our work contributes to the theory of information elicitation. We discussed the applicability and limitations for elicitation settings.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: We believe this paper poses no such risks.

33

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The datasets used in this paper are mentioned with URLs and the licenses and terms of use are properly respected.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

34

Answer: [NA]

Justification: The paper does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

35