Information Elicitation Mechanisms for Statistical Estimation

Yuqing Kong* Grant Schoenebeck† Biaoshuai Tao‡ Fang-Yi Yu§

Abstract

We study learning statistical properties from strategic agents with private information. In this problem, agents must be incentivized to truthfully reveal their information even when it cannot be directly verified. Moreover, the information reported by the agents must be aggregated into a statistical estimate. We study two fundamental statistical properties: estimating the mean of an unknown Gaussian, and linear regression with Gaussian error. The information of each agent is one point.

Our main results are two mechanisms for each of these problems which optimally aggregate the information of agents in the truth-telling equilibrium:

- A minimal (non-revelation) mechanism for large populations — agents need only report one value, but that value need not be their point.
- A mechanism for a small populations that is non-minimal — agents need to answer more than one question.

These mechanisms are "informed-truthful" mechanisms where reporting unaltered data (truth-telling) 1) forms a strict Bayesian Nash equilibrium and 2) has strictly higher welfare than any oblivious equilibrium where agents' strategies are independent of their private signals.

We also show a minimal revelation mechanism (each agent only reports her signal) for a restricted setting and use an impossibility result to prove the necessity of this restriction.

We build upon the peer prediction literature in the single-question setting; however, most previous work in this area focuses on discrete signals and homogeneous agents where as our setting is inherently continuous, and in the setting of linear regression, agents are heterogeneous.

*Peking University, yuqkong.kong@pku.edu.cn
†Univeristy of Michingan, schoeneb@umich.edu
‡Univeristy of Michingan, bstao@umich.edu
§Univeristy of Michingan, fayu@umich.edu

1 Introduction

Traditional statistical estimation approaches assume inputs are given and produce an output. However, increasingly, inputs must be obtained by eliciting information from a diverse set of users. In this new environment, gathering inputs can be at least as important and difficult as the computation itself. In many settings, users must be rewarded for their participation, and this is especially important when a representative sample is desired. However, rewarding users can create perverse incentives that lead to inaccurate reports, especially when the answers cannot be verified by the system. For example, agents may agree to participate due to the flat fee but then not take the time to report accurately or hide information due to privacy concerns.

This work develops systems to facilitate accurate statistical estimates by rewarding honest reporting—even when the information cannot be directly verified (peer prediction)—and then aggregating the results. In fact, we will see that often the aggregation and the reward go hand-in-hand: accurate rewards are produced by aggregating the information of other agents; but the truthfulness of the other agents relies on these accurate rewards. This creates a certain "full-pipeline" solution that integrates that information elicitation and aggregation into a single process.

We consider two fundamental statistical estimation problems: mean estimation and linear regression. In the mean estimation problem, the signal space is $\mathbb{R}^d$. Each agent i can choose to access a signal $s_i \in \mathbb{R}^d$ which is drawn from an unknown Gaussian distribution. The mean of the Gaussian is unknown to the agents, but believed to be drawn from a commonly known Gaussian prior distribution known to the agents, but not the mechanism. The mechanism's goal is to estimate the mean of the distribution by motivating each agent i to report some information about s_i truthfully.

For the linear regression problem, each agent i has a feature/attribute vector[1] $\phi(x_i) \in \mathbb{R}^{d+1}$ that is publicly known, and has access to a private signal $y_i \in \mathbb{R}$. The y_i point is from an unknown linear function applied to $\phi(x_i)$ plus some Gaussian error. While the linear function is not known it is selected from some common prior. The mechanism's goal is to estimate the linear function by motivating each agent i to report some information about y_i truthfully.

We seek to design information elicitation mechanisms that satisfy a number of useful properties including:

Informed Truthful: [17] Truth-telling is a strict Bayesian Nash equilibrium which both has the highest welfare (expected sum of agent payments) among all equilibria, and a strictly higher welfare than any oblivious equilibrium where agents' strategies are independent of their private signals.

Minimal: A mechanism is minimal if agents only need to report one value from the signal space. Additionally we say a mechanism is a **minimal revelation** mechanism if the point it requests from each agent is the agent's private information. If the one question the mechanism asks each agent may be about a different point (e.g. their best prediction for a peer's point), we call it a **minimal non-revelation** mechanism.

Small Group: The mechanism works for a small number of participants. In a large-group mechanism, we assume that the number of participants goes to infinity.

Detail-free: The mechanism need not know the common-prior of the agents.

Non-trivial Prior: Agents have some prior knowledge about the distribution of signals. In this case, the mechanism must be robust to an agent reporting a mixture of its signal and its prior when it is trying to predict the most likely signal of a peer.

As mentioned, we assume that the agents share a common prior, i.e. a common joint prior distribution over agents' private signals, which is a typical assumption when agents are only assigned one task [12, 13].

Peer-prediction information elicitation mechanisms are designed to craft payments solely based on agents' reports and have some "truth-rewarding" properties. Despite the great progress made in peer prediction literature [12, 13, 4, 17, 7], most previous work only considers the finite signal space, and when agents are only assigned one task, agents are usually assumed to be homogeneous. However, for many statistical

[1] In this work, we use this notion to simplify the matrix notation of Definition 6.1 where $\phi_j(x_i)$ represents the j-th attribute of agent i.

estimation settings, each agent's private information is from a continuous signal space (e.g., $\mathbb{R}^d$). Moreover, in the setting of linear regression, agents may be heterogeneous. A naive discretization approach [15] leads to impractically large finite space size; loses information, and fails to utilize the metric information to obtain stronger results. Thus, it is important to extend the aforementioned work to explicitly consider continuous signal spaces and to develop a new mechanism design approach in the continuous setting that can utilize the metric information.

Our Contributions Our main result is to develop two detail-free and informed truthful mechanisms for both mean estimation and linear regression, which optimally aggregate the information of agents to estimate the statistical property in the truth-telling equilibrium:

Proxy BTS A minimal non-revelation mechanism for large populations.

Disagreement A mechanism for small populations that is non-minimal.

We also ask if there exists a minimal revelation mechanism. In general, we show that the answer is no. However, for the special case where agents' knowledge of peer's private information comes solely from their own private information (not from the prior), we present **Jeffrey BTS**, a minimal revelation mechanism for small groups.

We summarize the properties of different mechanisms in the following table:

	Jeffrey BTS	Proxy BTS	Disagreement
Strict BNE	✓	✓	✓
informed truthful	✓	✓	✓
non-trivial prior		✓	✓
small group	✓		✓
minimal	✓	✓	

We note that the Jeffry BTS results only apply to mean estimation, and not to linear regression.

Structure of the paper We first review related work in section 1.1, and introduce our setting for information elicitation mechanisms and some basic game theory notions in section 2. We present our mechanisms for mean estimation problem in sections 3, 4, and 5, and the results of the mechanisms for linear regression problem are in section 6. Due to space, many of the proofs are omitted.

1.1 Related work

Peer prediction Miller et al. [12] introduce the peer prediction mechanism which is the first mechanism that has truth-telling as a strict Bayesian Nash equilibrium and does not need verification. However, their mechanism requires the full knowledge of the common prior and there exist some equilibria that are paid more than truth-telling. In particular, the oblivious equilibrium pays strictly more than truth-telling. Kong et al. [10] modify the original peer prediction mechanism such that truth-telling pays strictly better than any other equilibrium but still requires the full knowledge of the common prior. Prelec [13] designs the first detail-free peer prediction mechanism—Bayesian truth serum (BTS). BTS is also informed truthful. However, BTS is non-minimal: each agent needs to report the forecast in BTS, and requires an infinite number of participants. A series of works (e.g [14, 15, 18]) relaxes the large population requirement but loses the informed truthfulness property. Kong and Schoenebeck [8] propose a mechanism that is detail-free, informed truthful and works for a small population—the disagreement mechanism. In the above peer prediction literature, agents are assigned a single task and the common prior and the homogeneity of agents are usually assumed in the single-task setting. Our work also lies in the single-task setting and makes the common prior assumption. However, out study of the linear regression case allows agents to be heterogeneous.

Dasgupta and Ghosh [4] consider the setting where agents are assigned a batch of a priori similar tasks, which allows agents to have different priors. Their work requires each agent's private information to be a binary signal. Kong and Schoenebeck [7], Shnayder et al. [17] independently extend Dasgupta and Ghosh [4]'s work to multiple-choice questions.

In contrast with the general peer prediction literature, we consider a strategic statistical estimation setting and the private signal is from a continuous space.

Continuous signal space Goel and Faltings [5] propose a mechanism that elicits continuous valued multi-attribute personal data (e.g. body measurements). However, unlike us, they consider the setting where agents report multiple, say d, attributes and the continuous valued attributes are modeled by a mixture multidimensional normal distribution with K components and their results require that $d, K \geq 2$. They also require the number of agents to be infinite. We consider the setting where $d, K = 1$ and one of our proposed mechanisms is suitable for small populations.

Radanovic and Faltings [15] also consider the continuous signal space. They apply a discretization approach and ask the agents to report the forecast over all signals (non-minimal). However, while the discretization approach leads to large finite space size, which renders the forecast report impractical; loses information; and fails to utilize the metric information to obtain stronger results. In contrast, we utilize the metric information such that we are able to design minimal mechanisms and even in our non-minimal mechanism, the forecast report is practical (in addition to the signal report, agents are asked for their best guess for one signal report of another agent's signal, and how much they are "lying" in their reports).

Kong and Schoenebeck [9] consider eliciting agents' forecasts, which are also continuous values between 0 and 1. However, the strategic statistical estimation problems we study can come from high-dimensions, and lack the structure which is exploited by Kong and Schoenebeck [9].

Strategic machine learning The work of Cai, Daskalakis, and Papadimitriou [1] also studies the statistical inference from strategic sources and was an inspiration for this paper. One key difference is that in their model agents draw a signal with some noise (less noise is more costly), but then are assumed to truthfully report their received signal. We do not make the assumption that agents must report truthfully. This is especially significant in the case where agents have prior information about the signal that they could use to coordinate without procuring additional information. Their work focuses on incentivizing optimally cost-effective effort rather than truth-telling.

Chen and Liu [11] use machine learning techniques to incentivize agents to report truthfully on heterogeneous tasks. Conceptually, our linear regression techniques are similar in that they use information aggregation to accomplish information elicitation. However, their work is rather different as it focuses on the discrete setting.

Chen et al. [3] also consider the problem of strategy-proof linear regression. However, their setting is very different from ours. In their model, the agents care about the outcome of the learner, and this is what incentivizes them to truthfully (or un-truthfully) report. In our setting, agents are motivated by monetary incentives and are indifferent to the outcome of the learner.

Finally, there is a series of work on procuring data when that data has different costs [16, 1, 2, 19]. Our model is better suited to study procuring high-quality data rather than cost efficiency.

2 Preliminaries

Throughout the paper, we use n to denote the total number of agents. Let Ω_S be the signal space, and each agent i obtains a signal $s_i \in \Omega_S$.

2.1 Prior and Posterior

Before obtaining a signal, each agent believes that the set of n signals $(s_1, \dots, s_n)$ is sampled from a joint distribution $\mathbf{P}$ over Ω_S^n called *the prior* which is common knowledge. After agent i receives signal s_i, he updates his belief to the *posterior* $\mathbf{Q}_i(s_i)$ which is a distribution over the remaining $n-1$ signals $(s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$. We omit the input s_i and just write $\mathbf{Q}_i$ when there is no confusion.

Let $\mathcal{P}$ be a family of prior distributions. An important family of common priors we consider in this paper is the *Gaussian common prior*. First, we define a type of joint distribution called the *two-step Gaussian distribution*.

Definition 2.1. Given $m_0 \in \mathbb{R}^d$ and two positive definite matrices $\sigma^2, \tau^2 \in \mathcal{S}_{++}^d \subset \mathbb{R}^{d\times d}$ where $\mathcal{S}_{++}^d$ is the set of d-dimensional positive definite matrices, a *two-step Gaussian distribution* with parameters $(n, m_0, \sigma^2, \tau^2)$ is a joint distribution $(X_1, \ldots, X_n) \in \mathbb{R}^d$ defined as follows:

Step 1: A common mean (state) μ is sampled from $\mathcal{N}(m_0, \sigma^2)$.

Step 2: For all $i \in [n]$, $X_i \sim \mathcal{N}(\mu, \tau^2)$ are sampled independently and identically.

Definition 2.2. Consider the signal space $\Omega_S = \mathbb{R}^d$. A prior $\mathbf{P}$ is called a *Gaussian common prior* if $\mathbf{P}$ is a two-step Gaussian distribution with parameters $(n, m_0, \sigma^2, \tau^2)$ for certain $m_0 \in \mathbb{R}^d$ and positive definite matrices $\sigma^2, \tau^2 \in \mathcal{S}_{++}^d$. We denote a Gaussian common prior by $G(n, m_0, \sigma^2, \tau^2)$.

We summarize some related properties of the Gaussian common prior below.

Proposition 2.3. *Under a Gaussian common prior $G(n, m_0, \sigma^2, \tau^2)$, the marginal distribution for each s_i is $\mathcal{N}(m_0, \sigma^2 + \tau^2)$, and after agent i receives signal s_i, the posterior $\mathbf{Q}_i$ is a two-step Gaussian distribution with parameters*

$$G\left(n-1, (\sigma^{-2} + \tau^{-2})^{-1}(\tau^{-2}s_i + \sigma^{-2}m_0), (\sigma^{-2} + \tau^{-2})^{-1}, \tau^2\right).$$

We also consider a special case of the Gaussian common prior called the *Jeffreys prior* [6] in which the distribution of the common mean is arbitrarily close to the uniform distribution.

Definition 2.4. We call a prior $\mathbf{P}$ is called a *Jeffreys prior* if $\mathbf{P}$ is a two-step Gaussian distribution with parameters (n, m_0, ∞, τ^2)[2] for certain $m_0 \in \mathbb{R}^d$ and positive definite $\tau^2 \in \mathcal{S}_{++}^d$.

As an analogue to Proposition 2.3, we also have the following property of Jeffreys prior.

Proposition 2.5. *Under a Jeffreys prior $\mathbf{P} = G(n, m_0, \infty, \tau^2)$, after agent i receives signal s_i, the posterior $\mathbf{Q}_i$ is a two-step Gaussian distribution $G\left(n-1, s_i, \tau^2, \tau^2\right)$.*

2.2 Game Theory Basics

Informally, an *information elicitation mechanism* collects from each agent i a *report*, and rewards each agent i based on his report and the other agents' reports.

Definition 2.6 (Mechanism)**.** An *information elicitation mechanism* $\mathcal{M} = (\Omega_R, \pi_R, M_1, \ldots, M_n)$ specifies a space of allowed reports Ω_R, what kind of information should be contained in the report π_R, and rewards each agent i $M_i(\mathbf{r}) : \Omega_R^n \mapsto \mathbb{R}_{\geq 0}$ upon receiving the report collection $\mathbf{r} = (r_1, \ldots, r_n) \in \Omega_R^n$. We view π_R as a function $\pi_R : \mathcal{P} \times \Omega_S \mapsto \Omega_R$ which maps a prior distribution and a signal to a report which truthfully contains the required information.

Based on the nature of Ω_R and π_R, mechanisms can be classified into the following three types.

(Revelation) minimal $\Omega_R = \Omega_S$, and $\pi_R(\mathbf{P}, s_i) = s_i$. That is, the mechanism only collects each agent i's private signal s_i;

(Non-revelation) minimal $\Omega_R = \Omega_S$, and π_R can be arbitrary. That is, for each agent i, the mechanism collects a value $r \in \Omega_S$, which may be the private signal s_i, *or* anything else specified by the mechanism. For example, in the Gaussian common prior case, this may be the common mean of the posterior $\mathbf{Q}_i$ (which is $(\sigma^2 + \tau^2)^{-1}(\sigma^2 s_i + \tau^2 m_0)$ as given in Proposition 2.3);

Non-minimal Ω_R and π_R can be arbitrary. For example, a non-minimal mechanism can collect from agent i his private signal s_i and the mean of his posterior belief $\mathbf{Q}_i$. In the case $\mathbf{P}$ is a Gaussian common prior, we have $\Omega_R \in \mathbb{R}^d \times \mathbb{R}^d$ and $\pi_R(\mathbf{P}, s_i) = (s_i, \mu_i)$ where $\mu_i = (\sigma^{-2} + \tau^{-2})^{-1}(\tau^{-2}s_i + \sigma^{-2}m_0)$ (see Proposition 2.3). Our *disagreement mechanism* in Sect. 4 is an example of this.

[2]For any $\mu \in \mathbb{R}, p(\mu) \sim 1$ which can be seen as $\sigma = \infty$ in a two-step Gaussian distribution.

Definition 2.7 (Strategy)**.** Given a mechanism $\mathcal{M}$, the *strategy* of each agent i in the mechanism $\mathcal{M}$ is a mapping $\theta_i : \mathcal{P} \times \Omega_S \mapsto \Delta_{\Omega_R}$ from his prior and obtained signal to a probability distribution over Ω_R, and a collection of agents' strategies $\boldsymbol{\theta} := (\theta_1, \theta_2, \dots, \theta_n)$ is called a *strategy profile*.

In this work, the report spaces Ω_R are the Euclidean space $\mathbb{R}^r$ for some constant r, and we only consider the agents' strategies that are distributions with finite second moments: for each agent i and $s_i \in \Omega_S$, the random vector $\theta_i(\boldsymbol{P}, s_i) \in \mathbb{R}^r$ has a finite second moment $\mathbb{E}[\theta_i(\boldsymbol{P}, s_i) \cdot \theta_i(\boldsymbol{P}, s_i)^\top]$.[3]

A strategy profile $\boldsymbol{\theta}$ is *truth-telling* if each θ_i is a pure strategy and $\theta_i(\mathbf{P}, s_i) = \pi_R(\mathbf{P}, s_i)$. On the other hand, a strategy θ_i is *oblivious* if θ_i does not depend on the signals: for any $s_1, s_2 \in \Omega_S$ and any $\mathbf{P} \in \mathcal{P}$, we have $\theta_i(\mathbf{P}, s_1) = \theta_i(\mathbf{P}, s_2)$.

Definition 2.8 (Utility and social welfare)**.** Given a mechanism $\mathcal{M}$, under a common prior $\mathbf{P}$, for a strategy profile $\boldsymbol{\theta}$, the *prior expected utility* of agent i is

$$u_i(\mathbf{P}, \boldsymbol{\theta}) \triangleq \mathop{\mathbb{E}}_{\mathbf{s} \sim \mathbf{P}; \mathbf{r} \sim \boldsymbol{\theta}(\mathbf{P}, \mathbf{s})} \left[M_i(\mathbf{r}) \right],$$

the *welfare* is the sum of all agents' prior expected utilities:

$$w(\mathbf{P}, \boldsymbol{\theta}) = \sum_{i=1}^{n} u_i(\mathbf{P}, \boldsymbol{\theta}),$$

and the *average welfare* is $\frac{1}{n} \sum_{i=1}^{n} u_i(\mathbf{P}, \boldsymbol{\theta})$.

Bayes Nash equilibrium A *Bayes Nash equilibrium* under common prior $\mathbf{P}$ consists of a strategy profile $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ such that no agent can obtain a strictly higher utility by changing his strategy, given the strategies of the other agents. A formal definition is shown below.

Definition 2.9 (Bayes Nash equilibrium)**.** The strategy profile $\boldsymbol{\theta}$ is a *Bayes Nash equilibrium* with respect to an information elicitation mechanism $\mathcal{M}$ under common prior $\mathbf{P}$ if for each agent i and for all $\theta_i' \neq \theta_i$ almost surely (see Definition 2.12), we have

$$u_i(\mathbf{P}, \boldsymbol{\theta}) \geq u_i(\mathbf{P}, \theta_i', \boldsymbol{\theta}_{-i}). \tag{1}$$

It is a *strict Bayes Nash equilibrium* if the above inequality is strict for each agent i.

2.3 Mechanism Design Goals

A mechanism is *(strictly) truthful* if the truth-telling strategy profile is a (strict) Bayes Nash equilibrium under every prior $\mathbf{P} \in \mathcal{P}$. A stronger goal is to design a mechanism that is (strictly) truthful and *the truth-telling profile has the maximum welfare.* The ultimate goal we want from a mechanism is the *informed truthfulness*, which says that it is strictly truthful, the truth-telling profile has the maximum welfare, and each agent receives strictly more in a truth-telling profile than in an oblivious strategy profile.

Definition 2.10. A mechanism $\mathcal{M} = (\Omega_R, \pi_R, M_1, \dots, M_n)$ is *informed truthful* under a prior family $\mathcal{P}$ if it is strictly truthful, the truth-telling profile has the maximum welfare, and for any $i \in [n]$ and $\mathbf{P} \in \mathcal{P}$,

$$u_i(\mathbf{P}, \boldsymbol{\theta}) > u_i(\mathbf{P}, \boldsymbol{\theta'}),$$

where $\boldsymbol{\theta}$ is the truth-telling profile and $\boldsymbol{\theta'}$ is any oblivious Bayes Nash equilibrium profile.

We also regard $\mathcal{M}$ as informed truthful if it is strictly truthful, the truth-telling profile has the maximum welfare, and there are no oblivious Bayes Nash equilibrium.

Definition 2.11. We say that a mechanism $\mathcal{M}$ is an *optimal estimator* of a statistic, if, in the truthful equilibrium, it can reproduce any estimate of that statistic as if it were given access to all the private signals of the agents.

[3]The restriction to finite second moment is largely a technicality as highly diffuse strategies are unlikely to pay well in our mechanisms.

2.4 Mathematical Notations and Equalities

Definition 2.12 (Equal almost surely)**.** Given two random variables X and Y, $X = Y$ *almost surely* (a.s.) if $\Pr[X \in A] = \Pr[Y \in A]$ for all Borel-measurable sets A.

Proposition 2.13 (Sherman-Morrison formula)**.** *If A is an invertible matrix, and u, v are two vectors,* $(A^{-1} + uv^\top)^{-1} = A - \frac{Auv^\top A}{1+v^\top Au}$.

Proposition 2.14 (Bias variance decomposition)**.** *Given a random vector x and constant matrix A,*

$$\mathbb{E}[x^\top Ax] = \operatorname{Tr}(A \operatorname{Cov}(x)) + \mathbb{E}[x]^\top A\, \mathbb{E}[x],$$

where $\operatorname{Cov}(x) = \mathbb{E}[(x - \mathbb{E}[x])^\top (x - \mathbb{E}[x])]$ *is the covariance matrix of x.*

Proposition 2.14 implies the following proposition, which is used multiple times in our paper.

Proposition 2.15. *Given an arbitrary distribution $D \in \Delta_{\mathbb{R}^d}$ and an arbitrary constant positive definite matrix A, define $\phi : \mathbb{R}^d \mapsto \mathbb{R}$ as $\phi(x) = \mathbb{E}_{s\sim D}\left[(x - s)^\top A(x - s)\right]$. Then $\phi(\cdot)$ is continuous, and has unique minimizer $x = \mathbb{E}_{s\sim D}[s]$ with* $\min \phi(x) = \operatorname{Tr}(A \operatorname{Cov}(s))$*.*

Proof. Available in the appendix. □

The proof will be in the full version.

3 The Proxy Bayesian Truth Serum for Gaussian Prior

In this section, we propose a minimal mechanism called *the proxy Bayesian truth serum* (the proxy BTS), which is informed truthful for Gaussian common priors. From each agent, the proxy BTS elicits either the agent's private signal or the agent's posterior expectation. We first present the mechanism in a way that collects *both* the private signal and posterior expectation from each agent, and then show how to make it minimal by collecting *either* the private signal *or* the posterior expectation.

Theorem 3.1. *Let $n \to \infty$ and assume the Gaussian common prior $G(n, m_0, \sigma^2, \tau^2)$ in $\mathbb{R}^d$ defined in Definition 2.2. The mean estimation mechanism $\mathcal{M}_{\text{proxy}}$ in Algorithm 1 is an optimal estimator. Additionally, $\mathcal{M}_{\text{proxy}}$ is informed truthful.*

Remark 3.2. Because Mechanism 1 is decomposible, to make it minimal, for agent i we can collect s_i with probability 0.5, or t_i otherwise, and the property of informed truthfulness still holds.

The main idea of Mechanism 1 is to simulate the Bayesian truth serum (BTS) mechanism in [13] at the truth-telling equilibrium, so Mechanism 1 has truth-telling equilibrium as a strict Bayesian Nash equilibrium. Intuitively, in the BTS mechanism, the agents are asked to provide both the signals and the posterior believes of other agents' signals, and their payments can be decomposed into two parts, *the prediction score* and *the information score*.

Prediction score is based on how *accurate* the reported prediction is. The first term $(-(\hat{s}_j - \hat{t}_i)^\top L^{-1}(\hat{s}_j - \hat{t}_i))$ is larger when the agent's prediction $\hat{t}_i$ is closer to the reference agent's reported signal s_j. Moreover, the value of the first term is essentially the log-likelihood of $\hat{t}_i$ with respect to the Gaussian distribution $\mathcal{N}(\hat{s}_j, L)$. Thus, it is maximized from the prospective of agent i when the reported prediction is the maximum likelihood estimator of the mean—$\mathbb{E}[s_j \mid s_i]$.

Information score is based on how *surprisingly common* an agent's reported signal $\hat{s}_i$ is. The Equation (7) has two terms. The first term $(-(\hat{s}_i - \mu_s)^\top \Sigma_s^{-2}(\hat{s}_i - \mu_s))$ is large when the reported signal $\hat{s}_i$ is close to the average reported signal μ_s, and the second term $(+(\hat{s}_i - \hat{t}_j)^\top L^{-1}(\hat{s}_i - \hat{t}_j))$ is large if $\hat{s}_i$ is far from the reported prediction $\hat{t}_j$. Therefore the first term can be interpreted as how *common* the reported signal is, and the second term is how *surprising* it is.

Algorithm 1 Proxy-BTS mechanism $\mathcal{M}_{\text{proxy}} = (\Omega_R, \pi_R, \mathbf{M})$

1: Generate two disjoint groups A, B with sizes equal to $\lfloor \frac{n}{2} \rfloor$ and $n - \lfloor \frac{n}{2} \rfloor$.

2: Each agent $i \in A$ reports a signal and a prediction of the posterior mean $r_i = (\hat{s}_i, \hat{t}_i)$. That is $\Omega_R = \mathbb{R}^d \times \mathbb{R}^d$ and $\pi_R(\mathbf{P}, s_i) = (s_i, \mathbb{E}[\mu \mid s_i])$. If $|A| < |B|$ randomly remove an agent in group B and give it 0 payment.

3: For each agent $i \in A$, choose a reference agent $j \in B$ uniformly at random (and vice versa), and all agents in $A_{-i} \triangleq A \setminus \{i\}$ as i's competitors. Based on the signal report $\{\hat{s}_k\}_{k \in A_{-i}}$ the prediction report $\{\hat{t}_k\}_{k \in A_{-i}}$, calculate the sample mean and covariance—

$$\mu_s(i) \triangleq \frac{1}{|A_{-i}|} \sum_{k \in A_{-i}} \hat{s}_k \text{ and } \mu_t(i) \triangleq \frac{1}{|A_{-i}|} \sum_{k \in A_{-i}} \hat{t}_k \tag{2}$$

$$\Sigma_s(i)^2 \triangleq \frac{1}{|A_{-i}| - 1} \sum_{k \in A_{-i}} (\hat{s}_k - \mu_s(i))(\hat{s}_k - \mu_s(i))^\top \tag{3}$$

$$\Sigma_t(i)^2 \triangleq \frac{1}{|A_{-i}| - 1} \sum_{k \in A_{-i}} (\hat{t}_k - \mu_t(i))(\hat{t}_k - \mu_t(i))^\top \tag{4}$$

If (3) and (4) are positive definite, let $\Sigma_s(i), \Sigma_t(i)$ be the positive square root them and drop i when there is no ambiguity. If not, skip to step 5. With the above values, compute

$$K \triangleq \Sigma_s(\Sigma_s^{-1}\Sigma_t^2\Sigma_s^{-1})^{1/2}\Sigma_s^{-1} \text{ and } L \triangleq (K + \mathbb{I})\Sigma_s^2 \tag{5}$$

4: Depending on $\hat{t}_i$, the agent i's prediction score is

$$\text{PS}_i(\mathbf{r}) = -(\hat{s}_j - \hat{t}_i)^\top L^{-1}(\hat{s}_j - \hat{t}_i) + (\hat{s}_j - \mu_s)^\top \Sigma_s^{-2}(\hat{s}_j - \mu_s) \tag{6}$$

and depending on $\hat{s}_i$ the information score is

$$\text{IS}_i(\mathbf{r}) = -(\hat{s}_i - \mu_s)^\top \Sigma_s^{-2}(\hat{s}_i - \mu_s) + (\hat{s}_i - \hat{t}_j)^\top L^{-1}(\hat{s}_i - \hat{t}_j) \tag{7}$$

5: The reward for agent i is

$$M_i(\mathbf{r}) = \begin{cases} -100, \text{ if (3) or (4) are singular for some agent.} \\ \text{IS}_i(\mathbf{r}) + \text{PS}_i(\mathbf{r}) \text{ otherwise.} \end{cases}$$

The normalization factors L^{-1} and Σ_s^{-2} are chosen carefully such that telling the truth maximizes the information score when other agents tell the truth. While the mechanism is analogous to Bayesian Truth Serum [13], the proof is not. Agents have a much larger range of strategies in the continuous case that must be considered.

The following Lemma shows the limits and convergence properties of the statistics in Mechanism 1.

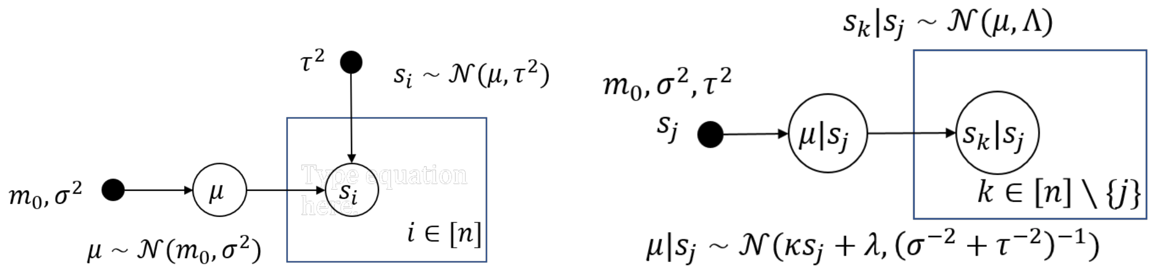
Lemma 3.3. *Under truth telling strategy profile, for all i, $\theta_i(s_i) = (s_i, \mathbb{E}[s_{new}|s_i])$, where s_{new} is the signal of an agent that is not i, as $n \to \infty$, for three distinct agents i, j and k,*

$$\mu_s(i) \xrightarrow{p} \mathbb{E}[s_k] \tag{8}$$

$$\mu_t(i) \xrightarrow{p} \mathbb{E}_{s_j}[\mathbb{E}[s_k|s_j]] \tag{9}$$

$$\Sigma_s(i)^2 \xrightarrow{p} \text{Cov}[s_k] = \tau^2 \tag{10}$$

$$\Sigma_t(i)^2 \xrightarrow{p} \text{Cov}[t_k] = \text{Cov}_{s_j}[\mathbb{E}[s_k|s_j]] \triangleq \sigma_t^2. \tag{11}$$



(a) prior (b) posterior

Figure 1: (a) Graphical model for the prior of the two-step Gaussian distribution with parameters $(n, m_0, \sigma^2, \tau^2)$. (b) Graphical model for the posterior of the agent j with signal s_j. The distribution of $\mu \mid s_j$ is a Gaussian with mean $\kappa s_j + \lambda$ and variance $\left(\sigma^{-2} + \tau^{-2}\right)^{-1}$. The distribution of $s_k \mid s_j$ is also a Gaussian with mean $\kappa s_j + \lambda$ and variance Λ as defined in Lemma 3.3.

Let $\kappa \triangleq (\sigma^{-2} + \tau^{-2})^{-1}\tau^{-2}$ *and* $\lambda \triangleq (\mathbb{I} - \kappa)m_0$.

$$\mathbb{E}[s_k|s_j] = \kappa s_j + \lambda \text{ and } \operatorname{Cov}_{s_j}[\mathbb{E}[s_k|s_j]] = \kappa \operatorname{Cov}_{s_j}[s_k]\kappa^\top = \kappa\tau^2\kappa^\top. \tag{12}$$

With equation (10) *and* (11), $\Sigma_s(i)$ *and* $\Sigma_t(i)$ *are invertible with probability* 1, *and*

$$K \xrightarrow{p} \kappa \triangleq (\sigma^{-2} + \tau^{-2})^{-1}\tau^{-2} \tag{13}$$

$$L \xrightarrow{p} \Lambda \triangleq \operatorname{Cov}[s_k|s_j] = (\sigma^{-2} + \tau^{-2})^{-1} + \tau^2 = (\kappa + \mathbb{I})\tau^2 \tag{14}$$

where $\mathbb{I}$ *is the identity matrix.*

The proof is in appendix. Now we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. **Strictly truthful:** For the first part of the statement, we want to show each agent i plays

$$\theta_i(s_i) = (s_i, \mathbb{E}[s_{\text{new}}|s_i]) \tag{15}$$

forms a strict Bayesian Nash equilibrium where s_{new} is the signal from a new agent that is not i.

A key observation is that in a Gaussian common prior, for each agent i, after receiving signal s_i, its estimation for s_{new} is an affine transformation of its signal— $\mathbb{E}[s_{\text{new}} \mid s_i] = \kappa s_i + \lambda$ where κ, λ are defined in Lemma 3.3.

Because this mechanism is decomposable— the report of signal $\hat{s}$ only affects the information score and the report of the prediction $\hat{t}$ only affects the prediction score, we can discuss the strategy of the signal and prediction separately. Note that for agent i the probability that (3) or (4) is singular is 0 when agents are truthtelling, so the best response is maximizing its $\mathrm{IS} + \mathrm{PS}$.

Prediction score: Suppose every other agent uses the truth telling strategy defined in equation (15). For the prediction score in (6), note that $\arg\max_{\hat{t}_i} \mathbb{E}[\mathrm{PS}_i(\mathbf{r})|s_i]$ is equal to

$$\arg\min_{\hat{t}_i} \mathbb{E}\left[(s_j - \hat{t}_i)^\top L^{-1}(s_j - \hat{t}_i)|s_i\right] \tag{16}$$

Now we study how to minimize the quadratic form $(\hat{s}_j - \hat{t}_i)^\top L^{-1}(\hat{s}_j - \hat{t}_i)$. Firstly, because (10) and (11) are consistent estimators and L is derived by a continuous function of Σ_s and Σ_t, so L also converges in probability with limit Λ in (14). Therefore we can consider $\mathbb{E}\left[(s_j - \hat{t}_i)^\top \Lambda^{-1}(s_j - \hat{t}_i)|s_i\right]$ instead as $n \to \infty$.

A strategy for $\hat{t}$ is a random function from $\mathbb{R}^d$ to $\mathbb{R}^d$. Fixing an s_i, the action for agent i is a distribution in $\Delta_{\mathbb{R}^d}$. Because Λ is positive definite and $\hat{t}_i$ and s_j are independent conditioned on s_i, by proposition 2.15, the best response is

$$\hat{t}_i = \mathbb{E}[s_j|s_i] \text{ with probability } 1 \tag{17}$$

Information score: Similarly, we can decompose the strategy for signal $\hat{s}_i \triangleq D_s(s_i) + \epsilon_s(s_i)$ into the expectation $D_s(s_i) = \mathbb{E}[\hat{s}_i]$ and noise $\epsilon_s(s_i) = \hat{s}_i - D_s(s_i)$. We first show that it is sufficient to consider deterministic strategies. By (7), $\mathbb{E}[\mathrm{IS}_i(\mathbf{r})|s_i]$ is equal to

$$
\begin{aligned}
&= \mathbb{E}\left[(\hat{s}_i - \hat{t}_j)^\top L^{-1}(\hat{s}_i - \hat{t}_j) - (\hat{s}_i - \mu_s)^\top \Sigma_s^{-2}(\hat{s}_i - \mu_s)|s_i\right] \\
&= \mathbb{E}[(D_s(s_i) + \epsilon_s(s_i) - \hat{t}_j)^\top L^{-1}(D_s(s_i) + \epsilon_s(s_i) - \hat{t}_j) - (D_s(s_i) + \epsilon_s(s_i) - \mu_s)^\top \Sigma_s^{-2}(D_s(s_i) + \epsilon_s(s_i) - \mu_s)|s_i] \\
&= \mathbb{E}\left[(D_s(s_i) - \hat{t}_j)^\top L^{-1}(D_s(s_i) - \hat{t}_j)\right. \\
&\quad - (D_s(s_i) - \mu_s)^\top \Sigma_s^{-2}(D_s(s_i) - \mu_s) \\
&\quad \left. + \operatorname{Tr}\left((L^{-1} - \Sigma_s^{-2}) \operatorname{Cov}(\epsilon_s(s_i))\right) \middle| s_i\right]. \qquad \text{(Proposition 2.14)}
\end{aligned}
$$

Because the noise term ϵ_s only appears in $\operatorname{Tr}\left((L^{-1} - \Sigma_s^{-2}) \operatorname{Cov}(\epsilon_s(s_i))\right)$, it is sufficient to consider which ϵ_s minimizes this quadratic form. Note that $(L^{-1} - \Sigma_s^{-2})$ is symmetric and

$$
(L^{-1} - \Sigma_s^{-2}) = \Sigma_s^{-1}\left(\left((\Sigma_s^{-1}\Sigma_t^2\Sigma_s^{-1})^{1/2} + \mathbb{I}\right)^{-1} - \mathbb{I}\right)\Sigma_s^{-1}. \tag{18}
$$

By the definition Σ_s and Σ_t are positive semi-definite and invertible with probability 1, so $\Sigma_s^{-1}\Sigma_t^2\Sigma_s^{-1}$ is positive definite and $\left(\left((\Sigma_s^{-1}\Sigma_t^2\Sigma_s^{-1})^{1/2} + \mathbb{I}\right)^{-1} - \mathbb{I}\right)$ is negative definite. Thus by Equation (18) $(L^{-1} - \Sigma_s^{-2})$ is negative definite. That is, for all $x \in \mathbb{R}^d, x^\top(L^{-1} - \Sigma_s^{-2})x < 0$. Moreover, the trace of a negative definite matrix is negative. Therefore, unless $\operatorname{Cov}(\epsilon_s(s_i)) = 0$, this term, $\operatorname{Tr}\left((L^{-1} - \Sigma_s^{-2}) \operatorname{Cov}(\epsilon_s(s_i))\right) \left| s_i\right]$, is negative, and so is maximized at $\epsilon_s = 0$. Thus it is sufficient to consider deterministic strategies.

If i plays a deterministic strategy, the signal best-response against truth-telling problem reduces to an unconstrained optimization problem:

$$
\max_x \mathbb{E}[\mathrm{IS}((x, t_i), r_{-i})|s_i],
$$

which can be easily solved by calculus. First, we compute the first order derivative

$$
\begin{aligned}
\nabla \mathbb{E}[\mathrm{IS}((x, t_i), r_{-i})|s_i] =& 2\,\mathbb{E}\left[-\Sigma_s^{-2}(x - \mu_s) + L^{-1}(x - \hat{t}_j)|s_i\right] \\
=& 2\,\mathbb{E}\left[(L^{-1} - \Sigma_s^{-2})x - (L^{-1}\hat{t}_j - \Sigma_s^{-2}\mu_s)|s_i\right]
\end{aligned} \tag{19}
$$

and the second order derivatives is negative definite by (18) because

$$
\nabla^2 \mathbb{E}[\mathrm{IS}((x, t_i), r_{-i})|s_i] = 2\,\mathbb{E}\left[(L^{-1} - \Sigma_s^{-2})|s_i\right].
$$

Therefore, the information score is strictly concave and x^* maximizes the information score if and only if the first order derivative is zero. By Equation (19) the first order derivative is zero if $\mathbb{E}\left[(L^{-1} - \Sigma_s^{-2})|s_i\right] x^* = \mathbb{E}\left[(L^{-1}\hat{t}_j - \Sigma_s^{-2}\mu_s)|s_i\right]$. By Lemma 3.3, L and Σ_s do not depend on s_i. Thus, the maximum happens at x^* if and only if

$$
\mathbb{E}\left[L^{-1} - \Sigma_s^{-2}\right] x^* = \mathbb{E}\left[L^{-1}\right] \mathbb{E}\left[\hat{t}_j|s_i\right] - \mathbb{E}\left[\Sigma_s^{-2}\right] \mathbb{E}\left[\mu_s|s_i\right] \tag{20}
$$

we have the last equality. Now let compute the value of $\mathbb{E}\left[\mu_s|s_i\right]$ and $\mathbb{E}\left[\hat{t}_j|s_i\right]$. The sample mean μ_s is an unbiased estimator for $\mathbb{E}[\mu|s_i]$, so by equation 12 in Lemma 3.3

$$
\mathbb{E}\left[\mu_s|s_i\right] = \kappa s_i + \lambda. \tag{21}
$$

Because every other agents tell the truth,

$$
\mathbb{E}\left[\hat{t}_j|s_i\right] = \mathbb{E}\left[\mathbb{E}\left[\mu|s_j\right]|s_i\right] = \mathbb{E}\left[\kappa s_j + \lambda|s_i\right] = \kappa^2 s_j + (\kappa + \mathbb{I})\lambda. \tag{22}
$$

Combining (20), (21), (22), and (18) and taking $n \to \infty$ we have

$$
\mathbb{E}\left[-\Sigma_s^{-2}(K + \mathbb{I})^{-1}K\right] x^* = \mathbb{E}\left[-\Sigma_s^{-2}(K + \mathbb{I})^{-1}K\right] s_i.
$$

Since $\mathbb{E}\left[-\Sigma_s^{-2}(K + \mathbb{I})^{-1}K\right]$ is invertible by (18), $x^* = s_i$ which completes the proof.

Truth-telling has the maximum welfare: We want to show the social welfare of truth-telling equilibrium is better than (or equal to) all other Bayesian Nash equilibria.

There are two cases: if there exists i such that (3) or (4) is singular, the average welfare is -100; otherwise, because the reference agents are chosen uniformly at random, the welfare is $\sum \mathbb{E}[\text{IS} + \text{PS}] = 0$. Additionally, by equations (10) and (11), as $n \to \infty$, (3) or (4) are invertible with probability 1, so the average welfare of truth-telling is 0 which is greater than or equal to the welfare in the above two cases.

Informed truthful: Suppose there is an oblivious equilibrium with strategy profile of signals $\boldsymbol{\theta}^{(s)}$. By definition of oblivious, $\theta_i^{(s)}(s_i)$ is independent to s_i, so $\{\hat{s}_i\}$ are independent samples from random variables $\{\theta_i\}$ respectively. For the sake of contradiction, suppose all $\{\Sigma_t(i)\}$ are invertible, as otherwise the agent welfare is -100. Without loss of generality, there exists $i \in A$ such that $\hat{s}_i \neq \mathbb{E}_{j\in B}[\theta_j]$ with probability 1. If this were not the case, all the signals would be constants, and then the coveriance matrix would be singular, which we assumed was not the case. However, then this cannot be an equilibrium, because, by Equation (17), $s_i^* \triangleq (\hat{s}_i + \mathbb{E}_{j\in B}[\theta_j])/2$ has larger expected payment than reporting $\hat{s}_i$. Therefore, there exists $i, \Sigma_t(i)$ is singular and the average welfare is -100. □

4 The Disagreement Mechanism

In this section, we present an informed truthful non-minimal mechanism for a small number of agents, called *the disagreement mechanism.* The spirit of our mechanism is similar a mechanism in the discrete case [15, 8] and borrows its name from the mechanism in Kong and Schoenebeck [8]. However, much additional effort is needed to adapt it into our continuous setting.

The intuition of the disagreement mechanism is as follows: the mechanism collects from each agent a report containing his posterior belief and his private signal; it rewards each agent based on how well his reported posterior predicts another agent's signal, and it penalizes each agent if his reported signal is inconsistent to his reported posterior; in particular, the mechanism will use other agents' signals to infer a mapping (presumably from the prior) of private signals to posterior predictions of the mean. Then he will punish an agents in the measure that its reports are not consistent with this mapping.

The mechanism $\mathcal{M}$ is shown in Algorithm 2. Here is a verbal summary of what the mechanism does. The mechanism partitions agents into three groups A, B and C with sufficiently large sizes, and chooses $d+1$ reference agents in each group. Specifically, all reports from group A (B, C) are compared to the $d+1$ reference agents in group B (C, A). We say B, C, A are the reference groups of A, B, C respectively. Each agent is required to report a signal $\hat{s}_i$, the mean $\hat{\mu}_i$ of his posterior belief $\mathbf{Q}_i$, and an *amount of untruthfulness* $\xi_i \in \mathbb{R}_{\geq 0}$ where agent i confesses how untruthful his report is, with $\xi_i = 0$ being completely truthful. The prediction score measures how well $\hat{\mu}_i$ predicts the signal reported by the first reference agent in the reference group. Specifically, we compute the squared distance between $\hat{\mu}_i$ and the reference agent's reported signal, with the amount of untruthfulness ξ_i punished (Step 3). The inconsistency score measures how consistent $\hat{s}_i$ and $\hat{\mu}_i$ are (Step 4). In particular, to compute the inconsistency score for an agent $i \in A$, we first use the reports for the $d+1$ reference agents from the reference group B to infer a bijection between the private signals and posterior belief of the mean that is consistent with the prior distribution $\mathbf{P}$. The mechanism then checks whether an agent's posterior mean $\hat{\mu}_i$ is indeed that which should be inferred from its reported signal $\hat{s}_i$, and punish agent i for the amount of inconsistency. Notice that we do not punish an agent i for inconsistency if he confesses a positive amount of untruthfulness $\xi_i > 0$. The payment to each agent i is then computed by the sum of the two scores (Step 5), and we normalize the payments such that the sum of all the n agents' payments is zero (Step 6). Finally, we punish all agents in each group if they report the same posterior (Step 7).[4]

Theorem 4.1. *Assume a Gaussian prior $G(n, m_0, \sigma^2, \tau^2)$ in $\mathbb{R}^d$ defined in Definition 2.2. When the number of agents $n \geq 3d+3$, the mean estimation mechanism $\mathcal{M}$ is an optimal estimator. Additionally, $\mathcal{M}$ is informed truthful.*

The proof is in the appendix.

[4]An alternative design to the present mechanism would be to impose a sufficiently large enough punishment for each agent in group A (B, C) where all agents in group B (C, A) report the same *signal*—a more common design in the past literature. This also yields an informed truthful mechanism. However, this variant of the mechanism allows oblivious Nash equilibria, whereas the mechanism described in Algorithm 2 forbids all oblivious Nash equilibria, which is, in some sense, stronger.

Algorithm 2 The disagreement mechanism $\mathcal{M}(\Omega_R, \pi_R, M_1, \ldots, M_n)$

1: Each agent i reports a signal $\hat{s}_i \in \mathbb{R}^d$, the mean $\hat{\mu}_i$ of his posterior belief $\mathbf{Q}_i$, and an *amount of untruthfulness* $\xi_i \in \mathbb{R}_{\geq 0}$ where agent i confesses how untruthful his report is, with $\xi_i = 0$ being completely truthful. That is, $\Omega_R = \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}_{\geq 0}$ and $\pi_R(\mathbf{P}, s_i) = (s_i, \mu_i, 0)$.

2: Partition the agents into three groups A, B and C of sizes at least $d+1$. Let $\{a_0, a_1, \ldots, a_d\}$, $\{b_0, b_1, \ldots, b_d\}$ and $\{c_0, c_1, \ldots, c_d\}$ be the first $d+1$ agents in the three groups respectively.

3: Each agent i is assigned a *prediction score* defined as

$$\mathrm{PS}_i(\mathbf{r}) = \begin{cases} -\|\hat{s}_{b_0} - \hat{\mu}_i\|^2 - \xi_i & \text{if } i \in A \\ -\|\hat{s}_{c_0} - \hat{\mu}_i\|^2 - \xi_i & \text{if } i \in B \\ -\|\hat{s}_{a_0} - \hat{\mu}_i\|^2 - \xi_i & \text{if } i \in C \end{cases}.$$

4: Each agent i is also assigned an *inconsistency score* computed as follows:

- for each $j \in \{1, \ldots, d\}$, let $t_{a_j} = \hat{s}_{a_j} - \hat{s}_{a_0}$ and $T_a = [t_{a_1} t_{a_2} \cdots t_{a_d}]$; define T_b, T_c similarly. Let $\nu_{a_j} = \hat{\mu}_{a_j} - \hat{\mu}_{a_0}$ and $U_a = [\nu_{a_1} \nu_{a_2} \cdots \nu_{a_d}]$; define U_b, U_c similarly;
- the inconsistency score for each $i \in A$ is

$$\mathrm{IS}_i(\mathbf{r}) = \begin{cases} -\left\|U_b T_b^{-1}(\hat{s}_i - \hat{s}_{b_0}) - (\hat{\mu}_i - \hat{\mu}_{b_0})\right\| & \text{if } T_b \text{ is invertible and } \xi_i = 0 \\ 0 & \text{otherwise} \end{cases};$$

define the inconsistency score for agents in group B and C similarly (The report of each agent in B is compared to the first $d+1$ reports in C, and the report of each agent in C is compared to the first $d+1$ reports in A).

5: The *payment before normalization* for agent i is

$$\bar{M}_i(\mathbf{r}) = \mathrm{PS}_i(\mathbf{r}) + \mathrm{IS}_i(\mathbf{r}).$$

6: Normalize the payments for each agent i as follows:

$$M_i(\mathbf{r}) = \begin{cases} \bar{M}_i(\mathbf{r}) - \frac{1}{|A|}\sum_{j \in B} \bar{M}_j(\mathbf{r}) & \text{if } i \in A \\ \bar{M}_i(\mathbf{r}) - \frac{1}{|B|}\sum_{j \in C} \bar{M}_j(\mathbf{r}) & \text{if } i \in B \\ \bar{M}_i(\mathbf{r}) - \frac{1}{|C|}\sum_{j \in A} \bar{M}_j(\mathbf{r}) & \text{if } i \in C \end{cases}.$$

7: If all the agents in group A report the same posterior mean $\hat{\mu}_i$, then update the score for each $i \in A$

$$M_i(\mathbf{r}) \leftarrow M_i(\mathbf{r}) - 100.$$

Do the same for agents in group B and C.

5 Minimal Revelation Mechanisms

In this section, we study minimal revelation mechanisms which ask agents to report private signals only. In Sect. 5.1, we present two strong impossibility results showing that there is no revelation minimal mechanism that can achieve truthfulness. However, under the assumption that agents have a Jeffery prior (Definition 2.4), we present a mechanism called *the metric mechanism* that is revelation minimal and informed truthful.

5.1 Impossibility Results

The theorem below, whose proof is deferred to the full version, shows that, minimal revelation mechanisms cannot achieve even the weakest truthful property—truth-telling as a strict Bayes Nash equilibrium, if no additional assumption is made about the prior distribution (other than it is a Gaussian common prior). Notice that having truth-telling as a (weak) Bayes Nash equilibrium is trivial: we can pay each agent a fixed amount regardless of what he reports.

Theorem 5.1. *For any number n of agents, there is no minimal revelation mechanism such that the truth-telling profile $\{\theta_i(\mathbf{P}, s_i) = s_i\}$ is a strict Bayes Nash equilibrium.*

Proof of Theorem 5.1. Suppose for the sake of contradiction that such a mechanism $\mathcal{M}$ exists. We consider one-dimensional case where each $s_i \in \mathbb{R}$. Recall that $M_i : \mathbb{R}^n \mapsto \mathbb{R}^+$ is the payment function for agent i. We consider the following two different possible priors (with the same standard deviation for the conjugate and the likelihood): $\mathbf{P}^{(1)} = (n, m_0^{(1)} = 0, \sigma^2 = 1, \tau^2 = 1)$ and $\mathbf{P}^{(2)} = (n, m_0^{(2)} = 0.5, \sigma^2 = 1, \tau^2 = 1)$. We consider the two scenarios: $\mathbf{P}^{(1)}$ is the common prior and agent 1 receives signal $s_1^{(1)} = 1$, and $\mathbf{P}^{(2)}$ is the common prior and agent 1 receives signal $s_1^{(2)} = 0.5$. Applying the results in Proposition 2.3 in one-dimensional case, we can see that the posterior distributions for agent 1 are the same in both scenarios:

$$\mathbf{Q}_{-1} = G\left(n-1, \frac{1^2 \cdot 1 + 1^2 \cdot 0}{1^2 + 1^2}, \frac{1^2 \cdot 1^2}{1^2 + 1^2}, 1\right) = G\left(n-1, \frac{1}{2}, \frac{1}{2}, 1\right).$$

Assuming the remaining $n-1$ agents report their private signals truthfully, agent 1's expected payment (in both scenarios) by reporting $x \in \mathbb{R}$ can be represented by a function $\phi : \mathbb{R} \mapsto \mathbb{R}^+$ defined as

$$\phi(x) = \mathop{\mathbb{E}}_{(s_2,\ldots,s_n)\sim\mathbf{Q}_{-1}} \left[M_1(x, s_2, \ldots, s_n)\right].$$

Since $\mathcal{M}$ has truth-telling as the strict Bayes Nash equilibrium, reporting $x = s_1^{(1)} = 1$ should be the unique maximizer for ϕ in the first scenario: $\phi(1) > \phi(x)$ for any $x \neq 1$. This, in particular, implies $\phi(1) > \phi(0.5)$. Thus, agent 1 will receive higher expected payment if he misreports his signal to be 1 instead of the true signal $s_1^{(2)} = 0.5$ in the second scenario. $\square$

If we consider the *nonzero-effort* setting where each agent i needs to spend a positive amount of effort $c_i > 0$ to obtain signal s_i and assume quasi-linear utility functions (i.e., for each agent i, his utility is given by his payment minus the effort c_i he spent), the following theorem shows that we cannot even have truth-telling being a (weakly) Bayes Nash equilibrium. We defer the proof to the appendix.

Theorem 5.2. *For any number n of agents and in the nonzero-effort setting with any $c_1, \ldots, c_n \in \mathbb{R}^+$, there is no minimal revelation mechanism such that truth-telling profile $\{\theta_i(\mathbf{P}, s_i) = s_i\}$ is a Bayes Nash equilibrium.*

5.2 Minimal Revelation Mechanism for Jeffreys Prior

In this subsection, we present *the metric mechanism* which is minimal revelation and informed truthful if agents have a Jeffreys prior as in Definition 2.4.

Theorem 5.3. *Given any Gaussian prior $G(n, m_0, \infty, \tau^2)$ with $n \geq 4$ defined in Definition 2.2, the minimal revelation mechanism $\mathcal{M}_{\text{metric}}$ is an optimal estimator, and is, in addition, informed truthful.*

Algorithm 3 A metric mechanism $\mathcal{M}_{\text{metric}} = (\Omega_R, \pi_R, \mathbf{M})$

1: Mechanism generates (predetermined) two disjoint groups A, B with size equal to $\lfloor \frac{n}{2} \rfloor$ and $n - \lfloor \frac{n}{2} \rfloor$.
2: Each agent i reports a signal $\hat{s}_i \in \mathbb{R}^d$ where $\Omega_R = \mathbb{R}^d$ and $\pi_R(\mathbf{P}, s_i) = s_i$. If $|A| < |B|$ we randomly remove an agent in group B and give it 0 payment.
3: For each agent $i \in A$ we random choose a reference agent $j \in B$ and a competitor $k \in A$ (and vice versa for B) such that i, j, k are distinct. The payment to agent i is

$$M_i(\hat{\mathbf{s}}) = \left\{ \begin{array}{ll} -100 & \text{if } \hat{s}_i = \hat{s}_k \\ \|\hat{s}_j - \hat{s}_k\|_2^2 - \|\hat{s}_j - \hat{s}_i\|_2^2 & \text{otherwise.} \end{array} \right. \tag{23}$$

The proof is in the appendix, however, the intuition is that here an agent's private signal and posterior mean are identical. Thus rewarding the agents for reporting a signal close to the posterior mean (as the mechanism does), also rewards them for playing their private signal.

Remark 5.4. There is a simple mechanism that achieves the same property. Instead of rewarding agent i having the distance $\|\hat{s}_i - \hat{s}_j\|$ smaller than $\|\hat{s}_k - \hat{s}_j\|$, we can reward i for being strictly "between" j and k:

$$M_i(\hat{\mathbf{s}}) = \begin{cases} 1, & \text{if } \|\hat{s}_j - \hat{s}_i\| < \|\hat{s}_j - \hat{s}_k\| \\ 0, & \text{otherwise.} \end{cases}$$

This naturally gives the oblivious equilibrium 0 social welfare.

6 Linear Regression

In linear regression with respect to a linear function $\phi : \mathbb{R}^d \mapsto \mathbb{R}^{d+1}$ called the *feature extractor*, each agent has a public independent variable $x_i \in \mathbb{R}^d$ and a private signal $y_i \in \Omega_S = \mathbb{R}$ correlated with $\phi(x_i)$. In this section, we consider *the Gaussian linear prior* $G_L(n, \beta_0, \sigma^2, \tau^2)$ with the design matrix Φ, which is defined as follows.

Definition 6.1 (Gaussian linear model prior)**.** Consider the signal space $\Omega_S = \mathbb{R}$. Given $n \in \mathbb{N}$, $\beta_0 \in \mathbb{R}^{d+1}$, a positive number $\tau \in \mathbb{R}_+$ and a positive definite matrix $\sigma^2 \in \mathcal{S}_{++}^{d+1}$, a prior $G_L(n, \beta_0, \sigma^2, \tau^2)$ called a *Gaussian linear prior with design matrix* $\Phi \in \mathbb{R}^{n \times (d+1)}$,

$$\Phi = \begin{bmatrix} \phi(x_1)^\top \\ \phi(x_2)^\top \\ \vdots \\ \phi(x_n)^\top \end{bmatrix} = \begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \dots & \phi_d(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \dots & \phi_d(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(x_n) & \phi_1(x_n) & \dots & \phi_d(x_n) \end{bmatrix},$$

is a joint distribution over $(y_1, \dots, y_n)$ defined in the following two steps.

Step 1 The true state β is sampled from $\mathcal{N}(\beta_0, \sigma^2)$.

Step 2 For each $i \in [n]$, sample $y_i \sim \mathcal{N}(\beta^\top \phi(x_i), \tau^2)$ independently.

Proposition 6.2. *Under the Gaussian linear prior $G_L(n, \beta_0, \sigma^2, \tau^2)$ with design matrix Φ, after agent i with attribute $\phi(x_i)$ receives signal y_i, we have the following:*

1. *The marginal distribution of the state is $\mathcal{N}(\beta_i, \Sigma_i)$, where*

$$\Sigma_i^{-1} = \sigma^{-2} + \frac{1}{\tau^2}\phi(x_i)\phi(x_i)^\top \quad \textit{and} \quad \beta_i = \Sigma_i \left(\sigma^{-2}\beta_0 + \frac{y_i}{\tau^2}\phi(x_i) \right).$$

Moreover, $\Sigma_i = \sigma^2 - \frac{\sigma^2\phi(x_i)\phi(x_i)^\top\sigma^2}{\tau^2 + \phi(x_i)^\top\sigma^2\phi(x_i)}$ by Proposition 2.13.

2. *The posterior $\mathbf{Q}_i$ is $G_L(n, \beta_i, \Sigma_i, \tau^2)$ with design matrix $\Phi_{[n]\setminus\{i\}}$ which is a $(n-1)\times(d+1)$ submatrix of Φ without the i-th column. In particular, the marginal posterior of reference agent $j \neq i$ is*

$$\mathcal{N}\left(\beta_i^\top \phi(x_j), \phi(x_i)^\top \Sigma_i \phi(x_j) + \tau^2\right).$$

For information elicitation mechanisms, we consider two scenarios for how agents acquire samples (x_i, y_i):

Active learning/optimal experimental design The learning algorithm is able to interactively query the agents to obtain the desired outputs at new data points chosen by the algorithm—each agent i is assigned an x_i by the learning algorithm, and privately learns y_i from a conditional distribution $P(\cdot|x_i)$.

Passive learning The data for learning are obtained without reference to the learning algorithm—each agent i has a public x_i derived from a distribution $P(\cdot)$ and learns y_i from a conditional distribution $P(\cdot|x_i)$ privately;

Note that in the active setting, if we are not concerned with sample complexity, we can use our mechanism for mean estimation on each point individually by requesting multiple y_i estimates for each point x_i. However, this will be inefficient, and we would like the mechanism to be able to, given data set $\{x_i\}_{i\in[m]}$, use only m agents. A similar technical difficulty arises in the passive model, where we might not be able to query two points with the same x_i. It turns out that our solution for the passive model will port naturally to the active setting without this blow-up.

For the passive model, the data are obtained from some unknown distribution. Informally, given a public attribute profile $\Phi = (\phi(x_1), \dots, \phi(x_n))^\top \in \mathbb{R}^{n\times(d+1)}$ which is sampled from some random vector $\phi(X) \in \Delta_{\mathbb{R}^{d+1}}$, a mechanism for linear regression collects from each agent i with attribute $\phi(x_i)$ a report $r_i \in \Omega_R$, which may contain the signal y_i or some additional information (more on this later), where we denote by Ω_R the space of all possible reports, and rewards each agent i based on the collection of n reports $\mathbf{r} = (r_1, \dots, r_n)$.

Definition 6.3 (mechanism for linear regression)**.** Given n agents with a public attribute profile Φ, a *mechanism for linear regression* $\mathcal{M} = (n, \Phi, \Omega_R, \pi_R, M_1, \dots, M_n)$ has a report space Ω_R, a report specification π_R, and rewards each agent i, $M_i(\mathbf{r}, \Phi)$, upon receiving the report collection $\mathbf{r} = (r_1, \dots, r_n)_{\forall i: r_i\in\Omega_R}$.

We view π_R as a function $\pi_R : \mathcal{P}\times\Omega_\Phi\times\Omega_S \mapsto \Omega_R$ which maps a prior distribution, attribute profile Φ, and a signal y to a report the containing correct required information.

We propose counterparts of the proxy-BTS mechanism and the disagreement mechanism under the Gaussian linear prior. Note that our mechanisms can be translated to active learning setting as long as the public control variables $x_1, \dots, x_n$ are non-degenerate, meaning that the linear space spanned by $\phi(x_1), \dots, \phi(x_n)$ has full rank.

6.1 Proxy BTS for Gaussian Linear Prior

Here we propose a minimal non-revelation mechanism for Gaussian linear prior, which collects either the private signal y_i or posterior expectation of the state $\mathbb{E}[\beta|y_i]$ from each agent.

We present the mechanism that collects both the private signal and the posterior expectation from each agent, and applying the same adjustment in Remark 3.2 can make it minimal.

Theorem 6.4. *Under Gaussian prior $G(n, \beta_0, \sigma^2, \tau^2)$ with public attribute profile Φ sampled from some non-degenerate distribution over $\mathbb{R}^{d+1}$ such that $\mathrm{Cov}(\phi(X)) \in \mathcal{S}_{++}^{d+1}$ and the forth moment is bounded*[5]*, $\mathcal{M}_{\text{proxy}}$ in Algorithm 4 is informed truthful.*

The main idea of Mechanism 4 is identical to Mechanism 1, which simulates the Bayesian truth serum (BTS) in [13] mechanism at the truth-telling equilibrium as $n \to \infty$. On the other hand, if agents apply oblivious strategy, Mechanism 4 can detect and punish everyone. An important observation is that, instead of collecting prediction to all signals $y_1, \dots, y_n$, Mechanism 4 collects agent i's the prediction of the state vector β, and computes the prediction to all signals from the prediction of the state vector β. This is possible because of linear structure and the controlled variables Φ are public.

[5] $\mathbb{E}[|\prod_{0\le j\le d}\phi(X)_j^{k_j}|] < \infty$ for all $(k_0, k_1, \dots, k_d) \in \mathbb{N}_0^{d+1}$ where $\sum k_j \le 4$, if $n \to \infty$

Algorithm 4 Linear Proxy-BTS mechanism $\mathcal{M}_{\text{proxy}} = (n, \Phi, \Omega_R, \pi_R, \mathbf{M})$

1: Generate two disjoint groups A, B with size equal to $\lfloor \frac{n}{2} \rfloor$ and $n - \lfloor \frac{n}{2} \rfloor$.

2: Each agent $i \in A$ with attribute $\phi(x_i)$ reports a signal and a prediction of the posterior state vector, $r_i = (\hat{y}_i, \hat{\beta}_i)$. That is $\Omega_R = \mathbb{R} \times \mathbb{R}^{d+1}$ and $\pi_R(\mathbf{P}, y_i) = (y_i, \mathbb{E}[\beta \mid y_i])$. If $|A| < |B|$, we randomly remove an agent in group B and give it 0 payment.

3: For agent $i \in A$, the mechanism chooses a reference agent $j \in B$ uniformly at random (and vise versa). Based on the signal report $\{\hat{y}_k\}_{k \in A \setminus \{i\}}$, calculate

$$b(i) \triangleq \left(\Phi_{-i}^\top \Phi_{-i}\right)^{-1} \Phi_{-i}^\top \hat{y}_{-i}, \tag{24}$$

$$t(i)^2 \triangleq \frac{1}{|A| - 2} \sum_{k \in A \setminus \{i\}} \left(\hat{y}_k - b(i)^\top \phi(x_k)\right)^2 \tag{25}$$

where $\Phi_{-i} \triangleq [\phi(x_k)^\top]_{k \in A \setminus \{i\}}$ is a submatrix of Φ, and $y_{-i} \triangleq [y_k]_{k \in A \setminus \{i\}}$.

Based on the signal and prediction report $\{\hat{\beta}_k\}_{k \in A \setminus \{i\}}$, let $S_0(i) \in \mathbb{R}^{(d+1) \times (d+1)}$ and $b_0(i) \in \mathbb{R}^{d+1}$ be the (least squares) solution of the following linear systems

$$\forall k \in A \setminus \{i\}, (\hat{y}_k - \hat{\beta}_k^\top \phi(x_k)) S_0(i)^2 \phi(x_k) = t(i)^2 (\hat{\beta}_k - b_0(i)). \tag{26}$$

Finally, compute $S(i) = S_0(i)^2 - \frac{(S_0(i)^2 \phi(x_i))(S_0(i)^2 \phi(x_i))^\top}{t(i)^2 + \phi(x_i)^\top S_0(i)^2 \phi(x_i)}$ and $S(j) = S_0(j)^2 - \frac{(S_0(j)^2 \phi(x_j))(S_0(j)^2 \phi(x_j))^\top}{t(j)^2 + \phi(x_j)^\top S_0(j)^2 \phi(x_j)}$.

4: Depending on $\hat{t}_i$, the agent i's prediction score is

$$\mathrm{PS}_i(\mathbf{r}) = -\frac{(\hat{y}_j - \hat{\beta}_i^\top \phi(x_j))^2}{t(i)^2 + \phi(x_j) S(i) \phi(x_j)^\top} + \frac{(\hat{y}_j - b(i)^\top \phi(x_j))^2}{t(i)^2}, \tag{27}$$

and depends on $\hat{y}_i$, the information score is

$$\mathrm{IS}_i(\mathbf{r}) = -\frac{(\hat{y}_i - b(i)^\top \phi(x_i))^2}{t(i)^2} + \frac{(\hat{y}_i - \hat{\beta}_j^\top \phi(x_i))^2}{t(i)^2 + \phi(x_i)^\top S(j) \phi(x_i)}. \tag{28}$$

5: If there exists $k \in [n]$ such that (26) is singular, or (27) and (28) are not well-defined, the reward for agent i is

$$M_i(\mathbf{r}) = -100,$$

otherwise,

$$M_i(\mathbf{r}) = \mathrm{IS}_i(\mathbf{r}) + \mathrm{PS}_i(\mathbf{r}).$$

Prediction score is based on how *accurate* the reported prediction is. The first term, $-\frac{(\hat{y}_j-\hat{\beta}_i^\top\phi(x_j))^2}{t(i)^2+\phi(x_j)S(i)\phi(x_j)^\top}$, is larger when the agent's prediction for y_j, $\hat{\beta}_i^\top\phi(x_j)$, is closer to the reference agent's reported signal $\hat{y}_j$. Moreover, the value of the first term is essentially the log-likelihood of $\hat{\beta}_i^\top\phi(x_j)$ with respect to the Gaussian distribution $\mathcal{N}(\hat{y}_j, \phi(x_i)^\top\Sigma_i\phi(x_j)+\tau^2)$. Thus, it is maximized from the prospective of agent i when the reported prediction is the maximum likelihood estimator of the mean $\mathbb{E}[y_j \mid y_i] = \mathbb{E}[\beta^\top\phi(x_j) \mid y_i] = \mathbb{E}[\beta \mid y_i]^\top\phi(x_j)$.

Information score is based on how "surprisingly common" an agent's reported signal $\hat{y}_i$ is, compared to a reference agent prediction. Equation (28) has two terms. The first term $-\frac{(\hat{y}_i-b(i)^\top\phi(x_i))^2}{t(i)^2}$ is large when the reported signal $\hat{y}_i$ is close to the average prediction $b(i)^\top\phi(x_i)$, and the second term $\frac{(\hat{y}_i-\hat{\beta}_j^\top\phi(x_i))^2}{t(i)^2+\phi(x_i)^\top S(j)\phi(x_i)}$ is large if $\hat{y}_i$ is far from the reported prediction $\hat{\beta}_j^\top\phi(x_i)$. Therefore the first term can be interpreted as how *common* the reported signal is, and the second term is how *surprising* it is.

The normalization factor $t(i)^2$ and $t(i)^2+\phi(x_i)^\top S(j)\phi(x_i)$ are chosen carefully such that telling the truth maximizes the information score when other agents tell the truth.

The following Lemma shows the limits and convergence properties of the statistics in Mechanism 4, and the proof is moved to the appendix.

Lemma 6.5. *Under truth telling strategy profile $\theta_i(y_i) = (y_i, \mathbb{E}[\beta|y_i])$, as $n\to\infty$, for agent i*

$$b(i) \xrightarrow{p} \beta \text{ and } t(i) \xrightarrow{p} \tau^2 \tag{29}$$

$$S_0(i)^2 \xrightarrow{p} \sigma^2 \text{ and } b_0(i) \xrightarrow{p} \beta_0 \tag{30}$$

$$S(i) \xrightarrow{p} \sigma^2 - \frac{(\sigma^2\phi(x_i))(\sigma^2\phi(x_i))^\top}{\tau^2+\phi(x_i)^\top\sigma^2\phi(x_i)} = \Sigma_i \tag{31}$$

Now we are ready to prove the theorem

Proof of theorem 6.4. **Strictly truthful:** For the first part of the statement, we want to show each agent i plays the (pure) strategy

$$\theta_i(y_i) = (y_i, \mathbb{E}[\beta|y_i]) \tag{32}$$

forms a strict Bayesian Nash equilibrium.

Since this mechanism is decomposable, we can discuss the best response strategy on signal and prediction separately at Truth telling equilibrium. Note that for agent i the probability that (26) singular or (27), (28) not well-defined is 0, so the best response is maximizing its IS + PS.

Prediction score: Suppose every other agent use truth telling strategy defined in (32). For the prediction score in (27),

$$\arg\max_{\hat{\beta}_i} \mathbb{E}[\mathrm{PS}_i(\mathbf{r}) \mid y_i] = \arg\min_{\hat{\beta}_i} \mathbb{E}\left[-\frac{(\hat{y}_j-\hat{\beta}_i^\top\phi(x_j))^2}{t(i)^2+\phi(x_j)^\top S(i)\phi(x_j)} \mid y_i\right] \tag{33}$$

Now we study the quadratic form in (33). First, by (31) the denominator in (33) converges to

$$t(i)^2+\phi(x_j)^\top S(i)\phi(x_j) \xrightarrow{p} \tau^2 + \phi(x_j)^\top\left(\sigma^2 - \frac{(\sigma^2\phi(x_i))(\sigma^2\phi(x_i))^\top}{\tau^2+\phi(x_i)^\top\sigma^2\phi(x_i)}\right)\phi(x_j) \tag{34}$$

To prove equation (34) is positive, we define an equivalent kernel $K(i,j) \triangleq \phi(x_i)^\top\sigma^2\phi(x_j)$ to simplify the notions and have

$$\begin{aligned}
(34) &= \left(\tau^2+\phi(x_j)^\top\sigma^2\phi(x_j)\right) - \frac{(\phi(x_j)^\top\sigma^2\phi(x_i))^2}{\tau^2+\phi(x_i)^\top\sigma^2\phi(x_i)}\\
&= \frac{1}{\tau^2+K(i,i)}\left(\left(\tau^2+K(j,j)\right)\left(\tau^2+K(i,i)\right) - (K(i,j))^2\right)\\
&= \frac{1}{\tau^2+K(i,i)}\left(\tau^4+\tau^2(K(j,j)+K(i,i))\right) && (K(i,i)K(j,j)=K(i,j)^2)\\
&= \frac{\tau^2+K(j,j)+K(i,i)}{\tau^2+K(i,i)}\cdot\tau^2 > 0
\end{aligned} \tag{35}$$

Therefore to maximize (33), it is sufficient to minimize

$$\begin{aligned}
&\mathbb{E}\left[(y_j-\hat\beta_i^\top\phi(x_j))^2\mid y_i\right]=\mathbb{E}\left[(y_j-\beta_i^\top\phi(x_j)+\beta_i^\top\phi(x_j)-\hat\beta_i^\top\phi(x_j))^2\mid y_i\right]\\
=&\mathbb{E}\left[(y_j-\beta_i^\top\phi(x_j))^2\mid y_i\right]+2\mathbb{E}_j\left[\mathbb{E}_{y_j}[(y_j-\beta_i^\top\phi(x_j))]((\beta_i-\hat\beta_i)^\top\phi(x_j))\mid y_i\right]+\mathbb{E}\left[((\beta_i-\hat\beta_i)^\top\phi(x_j))^2\mid y_i\right]\\
=&\mathbb{E}\left[(y_j-\beta_i^\top\phi(x_j))^2\mid y_i\right]+\mathbb{E}\left[((\beta_i-\hat\beta_i)^\top\phi(x_j))^2\mid y_i\right] \qquad (\mathbb{E}_{y_j}[(y_j-\beta_i^\top\phi(x_j))]=0)\\
=&\mathbb{E}\left[(y_j-\beta_i^\top\phi(x_j))^2\mid y_i\right]+\mathrm{Tr}\left((\beta_i-\hat\beta_i)(\beta_i-\hat\beta_i)^\top\mathrm{Cov}(\phi(x_j))\right)+\left((\beta_i-\hat\beta_i)^\top\mathbb{E}[\phi(x_j)]\right)^2.\\
&\text{(by Proposition 2.14)}
\end{aligned}$$

Using similar arguments in (17), we have

$$\Pr[\hat\beta_i=\beta_i]=1$$

Information score: Similarly to the theorem 3.1, it is sufficient to consider deterministic signal strategy, and the signal best-response against truth-telling problem reduces to a unconstrained optimization problem

$$\max_y \mathbb{E}[\mathrm{IS}((y,\beta_i),r_{-i})\mid y_i]$$

and its first derivatives

$$\nabla\mathbb{E}[\mathrm{IS}((y,\beta_i),r_{-i})|y_i]=2\mathbb{E}\left[-\frac{y-b(i)^\top\phi(x_i)}{t(i)^2}+\frac{y-\hat\beta_j^\top\phi(x_i)}{t(i)^2+\phi(x_i)^\top S(j)\phi(x_i)}\mid y_i\right]$$

and the second derivatives is

$$\nabla^2\mathbb{E}[\mathrm{IS}((y,t_i),r_{-i})|y_i]=2\mathbb{E}\left[-\frac{1}{t(i)^2}+\frac{1}{t(i)^2+\phi(x_i)^\top S(j)\phi(x_i)}\mid y_i\right]$$

which is negative by (35) as $n\to\infty$. Therefore, the information score is strictly concave and y^* maximizes the information score if and only if $\nabla\mathbb{E}[\mathrm{IS}((y,\beta_i),r_{-i})|y_i]\big|_{y=y^*}=0$, and take $n\to\infty$ we have $y^*=y_i$ which completes the proof.

Truth-telling has the maximum welfare: It is identical to the proof of theorem 3.1

Informed truthful: Finally, we show all oblivious equilibrium has strictly less welfare. Suppose there is an oblivious equilibrium with strategy profile of signal $\boldsymbol{\theta}^{(y)}$, by definition of oblivious $\theta_i^{(y)}(y_i)$ is independent to s_i, so $\{\hat y_i\}$ are independent samples from random variables $\{\theta_i(\Phi)\}$ respectively. Suppose all $\{S_0(i)\}_{i\in A}$ are invertible, without loss of generality, there exists $i\in A$ such that $\hat\beta_i\neq\mathbb{E}_{j\in B}[\theta_j(\Phi)]$, it is not an equilibrium because there exists $y_i^*\triangleq(\hat y_i+\mathbb{E}_{j\in B}[\theta_j(\Phi)])/2$ which has larger expected payment by (17). Therefore, there exist i such that $S_0(i)(i)$ is singular and the average welfare is -100. ☐

6.2 Disagreement Mechanism for Gaussian Linear Prior

The disagreement mechanism for the mean estimation can be adapted to the linear regression setting as well. The adapted mechanism is informed truthful for a small number of agents, but is non-minimal. We present it in Algorithm 5, and defer the proof of Theorem 6.6 to the appendix.

Theorem 6.6. *For $n\geq 3d+6$, under the Gaussian prior $G(n,\beta_0,\sigma^2,\tau^2)$ with public attribute profile Φ sampled from a non-degenerate distribution over $\mathbb{R}^{d+1}$ with bounded fourth moment such that $\mathrm{Cov}(\phi(X))\in\mathcal{S}_{++}^{d+1}$, the disagreement mechanism $\mathcal{M}$ in Algorithm 5 is an optimal estimator, and is, in addition, informed truthful.*

7 Future Work

There are many attractive avenues available for future work. Within our model, the question of whether there exists a minimal non-revelation mechanism that applies in the small group setting remains open.

Algorithm 5 The disagreement mechanism for linear regression $\mathcal{M}(\Omega_R, \pi_R, \Phi, M_1, \ldots, M_n)$

1: Each agent i with attribute x_i reports a signal $\hat{y}_i \in \mathbb{R}$, the mean $\hat{\beta}_i$ of his posterior belief $\mathbf{Q}_i$, and an *amount of untruthfulness* $\xi_i \in \mathbb{R}_{\geq 0}$ where agent i confesses how untruthful his report is, with $\xi_i = 0$ being completely truthful. That is, $\Omega_R = \mathbb{R} \times \mathbb{R}^{d+1} \times \mathbb{R}_{\geq 0}$ and $\pi_R(\mathbf{P}, y_i) = (y_i, \beta_i, 0)$.

2: Partition the agents into three groups A, B and C of sizes at least $d+2$. Let $\{a_0, a_1, \ldots, a_{d+1}\}$, $\{b_0, b_1, \ldots, b_{d+1}\}$ and $\{c_0, c_1, \ldots, c_{d+1}\}$ be the first $d+2$ agents in the three groups respectively.

3: Each agent i is assigned a *prediction score* defined as

$$\mathrm{PS}_i(\mathbf{r}) = \begin{cases} -\sum_{j=0}^{d} \left\| \hat{y}_{b_j} - \hat{\beta}_i^\top \phi\left(x_{b_j}\right) \right\|^2 - \xi_i & \text{if } i \in A \\ -\sum_{j=0}^{d} \left\| \hat{y}_{c_j} - \hat{\beta}_i^\top \phi\left(x_{b_j}\right) \right\|^2 - \xi_i & \text{if } i \in B \\ -\sum_{j=0}^{d} \left\| \hat{y}_{a_j} - \hat{\beta}_i^\top \phi\left(x_{b_j}\right) \right\|^2 - \xi_i & \text{if } i \in C \end{cases}.$$

4: Each agent i is also assigned an *inconsistency score* computed as follows:

- For each agent i with feature $\phi(x_i)$ and report $r_i = (\hat{y}_i, \hat{\beta}_i, \xi_i)$, denote $\delta_i = \hat{y}_i \phi(x_i) - \phi(x_i)\phi(x_i)^\top \hat{\beta}_i$.
- for each $j \in \{1, \ldots, d+1\}$, let $t_{a_j} = \hat{\beta}_{a_j} - \hat{\beta}_{a_0}$ and $T_a = [t_{a_1} t_{a_2} \cdots t_{a_{d+1}}]$; define T_b, T_c similarly;
- for each $j \in \{1, \ldots, d+1\}$, let $\nu_{a_j} = \delta_{a_j} - \delta_{a_0}$ and $U_a = [\nu_{a_1} \nu_{a_2} \cdots \nu_{a_{d+1}}]$; define U_b, U_c similarly;
- the inconsistency score for each $i \in A$ is

$$\mathrm{IS}_i(\mathbf{r}) = \begin{cases} -\left\| U_b T_b^{-1}(\hat{\beta}_i - \hat{\beta}_{b_0}) - (\delta_i - \delta_{b_0}) \right\| & \text{if } T_b \text{ is invertible and } \xi_i = 0 \\ 0 & \text{otherwise} \end{cases};$$

define the inconsistency score for agents in group B and C similarly.

5: The *payment before normalization* for agent i is

$$\bar{M}_i(\mathbf{r}) = \mathrm{PS}_i(\mathbf{r}) + \mathrm{IS}_i(\mathbf{r}).$$

6: Normalize the payments for each agent i as follows:

$$M_i(\mathbf{r}) = \begin{cases} \bar{M}_i(\mathbf{r}) - \frac{1}{|A|} \sum_{j \in B} \bar{M}_j(\mathbf{r}) & \text{if } i \in A \\ \bar{M}_i(\mathbf{r}) - \frac{1}{|B|} \sum_{j \in C} \bar{M}_j(\mathbf{r}) & \text{if } i \in B \\ \bar{M}_i(\mathbf{r}) - \frac{1}{|C|} \sum_{j \in A} \bar{M}_j(\mathbf{r}) & \text{if } i \in C \end{cases}.$$

7: If all the agents in group A report the same posterior mean $\hat{\beta}_i$, then update the score for each $i \in A$

$$M_i(\mathbf{r}) \leftarrow M_i(\mathbf{r}) - 100.$$

Do the same for agents in group B and C.

Moving beyond our model, one could study priors that are more general than Gaussian distributions (e.g. exponential families), or learning problems beyond linear regression (e.g. SVMs). Additionally, one could study how to efficiently (with minimum cost) elicit data in a setting similar to ours. A final direction in full pipe-line learning, is to relax the assumptions on the agents being fully rational. For example, a fraction of the agents could be random, malicious, or naturally truth-telling.

References

[1] Yang Cai, Constantinos Daskalakis, and Christos H Papadimitriou. Optimum statistical estimation with strategic data sources. *arXiv preprint arXiv:1408.2539*, 2014.

[2] Yiling Chen, Nicole Immorlica, Brendan Lucier, Vasilis Syrgkanis, and Juba Ziani. Optimal data acquisition for statistical estimation. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 27–44. ACM, 2018.

[3] Yiling Chen, Chara Podimata, Ariel D Procaccia, and Nisarg Shah. Strategyproof linear regression in high dimensions. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 9–26. ACM, 2018.

[4] Anirban Dasgupta and Arpita Ghosh. Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd international conference on World Wide Web*, pages 319–330. International World Wide Web Conferences Steering Committee, 2013.

[5] Naman Goel and Boi Faltings. Personalized peer truth serum for crowdsourcing multi-attribute personal data.

[6] Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 453–461, 1946.

[7] Y. Kong and G. Schoenebeck. A Framework For Designing Information Elicitation Mechanisms That Reward Truth-telling. *ArXiv e-prints*, May 2016.

[8] Yuqing Kong and Grant Schoenebeck. Equilibrium selection in information elicitation without verification via information monotonicity. In *9th Innovations in Theoretical Computer Science Conference*, 2018.

[9] Yuqing Kong and Grant Schoenebeck. Water from two rocks: Maximizing the mutual information. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 177–194. ACM, 2018.

[10] Yuqing Kong, Katrina Ligett, and Grant Schoenebeck. Putting peer prediction under the micro (economic) scope and making truth-telling focal. In *International Conference on Web and Internet Economics*, pages 251–264. Springer, 2016.

[11] Yang Liu and Yiling Chen. Surrogate scoring rules and a dominant truth serum for information elicitation. *CoRR*, abs/1802.09158, 2018. URL `http://arxiv.org/abs/1802.09158`.

[12] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, pages 1359–1373, 2005.

[13] D. Prelec. A Bayesian Truth Serum for subjective data. *Science*, 306(5695):462–466, 2004.

[14] Goran Radanovic and Boi Faltings. A robust bayesian truth serum for non-binary signals. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI" 13)*, number EPFL-CONF-197486, pages 833–839, 2013.

[15] Goran Radanovic and Boi Faltings. Incentives for truthful information elicitation of continuous signals. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI" 14)*, number EPFL-CONF-215878, pages 770–776, 2014.

[16] Aaron Roth and Grant Schoenebeck. Conducting truthful surveys, cheaply. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 826–843. ACM, 2012.

[17] Victor Shnayder, Arpit Agarwal, Rafael Frongillo, and David C. Parkes. Informed truthfulness in multi-task peer prediction. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, pages 179–196, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3936-0. doi: 10.1145/2940716.2940790. URL `http://doi.acm.org/10.1145/2940716.2940790`.

[18] Jens Witkowski and David C Parkes. Peer prediction without a common prior. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 964–981. ACM, 2012.

[19] Shuran Zheng, Bo Waggoner, Yang Liu, and Yiling Chen. Active information acquisition for linear optimization. *arXiv preprint arXiv:1709.10061*, 2017.

A Proofs in Sect. 2

Proof of Proposition 2.15. We prove a stronger statement: if x is drawn from some distribution D_X with randomness independent to $s \sim D$ ($x \perp\!\!\!\perp s$) and minimizes the expected value $\mathbb{E}_x[\phi(x)]$, $X = \mathbb{E}[s]$ almost surely.

Formally, let x is sampled from some distribution D_X with finite covariance $\mathrm{Cov}(X)$ independent of the randomness of $s \sim D$.

$$\begin{aligned}
\mathbb{E}_x[\phi(x)] &= \mathbb{E}_{x,s}\left[(x-s)^\top A(x-s)\right] \\
&= \mathbb{E}_{x,s}\left[\mathrm{Tr}(A\,\mathrm{Cov}(x-s))\right] + \mathbb{E}[x-s]^\top A\,\mathbb{E}[x-s] && \text{(by proposition 2.14)} \\
&= \mathbb{E}_x\left[\mathrm{Tr}(A\,\mathrm{Cov}(x))\right] + \mathbb{E}_s\left[\mathrm{Tr}(A\,\mathrm{Cov}(s))\right] + \mathbb{E}[x-s]^\top A\,\mathbb{E}[x-s] && (s \perp\!\!\!\perp x) \\
&\geq \mathbb{E}_x\left[\mathrm{Tr}(A\,\mathrm{Cov}(x))\right] + \mathbb{E}_s\left[\mathrm{Tr}(A\,\mathrm{Cov}(s))\right] && (A \in \mathcal{S}_{++}) \\
&\geq \mathbb{E}_s\left[\mathrm{Tr}(A\,\mathrm{Cov}(s))\right] = \mathrm{Tr}(A\,\mathrm{Cov}(s))
\end{aligned}$$

Therefore $\mathbb{E}[\phi(x)]$ is lower bounded by $\mathrm{Tr}(A\,\mathbb{E}_s\left[\mathrm{Cov}(s)\right])$ and the equality holds when the two inequalities are tight. The first one is equal when

$$\begin{aligned}
&\mathbb{E}[x-s]^\top A\,\mathbb{E}[x-s] = 0 \\
\Leftrightarrow\ &\mathbb{E}[x-s] = 0 && (A \text{ is positive definite}) \\
\Leftrightarrow\ &\mathbb{E}[x] = \mathbb{E}[s]
\end{aligned}$$

The second one is equality when

$$\begin{aligned}
&[\mathrm{Tr}(A\,\mathrm{Cov}(x))] = 0 \\
\Leftrightarrow\ &\mathrm{Cov}(x) = 0 && (A \text{ is positive definite})
\end{aligned}$$

Therefore if $x \sim D_X$ is a minimizer of $\mathbb{E}[\phi(x)]$, $x = \mathbb{E}[s]$ almost surely.

Finally, if for the function $\phi(x) : \mathbb{R}^d \mapsto \mathbb{R}$ by the second line of above argument,

$$\phi(x) = \mathbb{E}_s\left[\mathrm{Tr}(A\,\mathrm{Cov}(s))\right] + (x - \mathbb{E}[s])^\top A\,(x - \mathbb{E}[s])\,,$$

and ϕ is a quadratic function, so it is continuous. ☐

B Proofs in Sect. 3

Proof of Lemma 3.3. For equations (8) to (11), because under the truth-telling equilibrium agent j's report is s_j and $\mathbb{E}[s_k|s_j]$ defined in definition 2.2 where $k \neq j$ and they are independent between agents, $\mu_s(i)$ and $\mu_t(i)$, are the sample mean of X_j and $\mathbb{E}[X_k|X_j]$ respectively which are consistent estimators, so as n increases (8) and (9) hold. Equation (10) and (11) hold in the same manner.

Because σ, τ are positive definite, let $\sigma_t = (\mathrm{Cov}_{s_j}[\mathbb{E}[s_k|s_j]])^{1/2}$ be the positive-definite square root which is unique, $(\sigma^{-2} + \tau^{-2})^{-1} = \tau(\tau^{-1}\sigma_t^2\tau^{-1})^{1/2}\tau$, and $\kappa = \tau(\tau^{-1}\sigma_t^2\tau^{-1})^{1/2}\tau^{-1}$.

Finally, because Σ_s, Σ_t converge to τ, σ_t in probability and K can be seen as a continuous function from Σ_s, Σ_t to $\mathbb{R}$, by the Continuous Mapping Theorem which shows continuous functions are limit-preserving, we prove $K = \Sigma_s(\Sigma_s^{-1}\Sigma_t^2\Sigma_s^{-1})^{1/2}\Sigma_s^{-1}$ converges to $\tau(\tau^{-1}\sigma_t^2\tau^{-1})^{1/2}\tau^{-1} = \kappa$ in probability which proves (13). The proof of (14) is the same. ☐

C Proofs in Sect. 5

C.1 Proof of Theorem 5.1

Proof. Suppose for the sake of contradiction that such a mechanism $\mathcal{M}$ exists. We consider an one-dimensional case where each $s_i \in \mathbb{R}$. Recall that $M_i : \mathbb{R}^n \mapsto \mathbb{R}^+$ is the payment function for agent i.

We consider the following two different possible priors (with the same standard deviation for the conjugate and the likelihood): $\mathbf{P}^{(1)} = (n, m_0^{(1)} = 0, \sigma^2 = 1, \tau^2 = 1)$ and $\mathbf{P}^{(2)} = (n, m_0^{(2)} = 0.5, \sigma^2 = 1, \tau^2 = 1)$. We consider the two scenarios: $\mathbf{P}^{(1)}$ is the common prior and agent 1 receives signal $s_1^{(1)} = 1$, and $\mathbf{P}^{(2)}$ is the common prior and agent 1 receives signal $s_1^{(2)} = 0.5$. Applying the results in Proposition 2.3 in one-dimensional case, we can see that the posterior distributions for agent 1 are the same in both scenarios:

$$\mathbf{Q}_{-1} = G\left(n-1, \frac{1^2 \cdot 1 + 1^2 \cdot 0}{1^2 + 1^2}, \frac{1^2 \cdot 1^2}{1^2 + 1^2}, 1\right) = G\left(n-1, \frac{1}{2}, \frac{1}{2}, 1\right).$$

Assuming the remaining $n-1$ agents report their private signals truthfully, agent 1's expected payment (in both scenarios) by reporting $x \in \mathbb{R}$ can be represented by a function $\phi : \mathbb{R} \mapsto \mathbb{R}^+$ defined as

$$\phi(x) = \mathop{\mathbb{E}}_{(s_2,\ldots,s_n)\sim\mathbf{Q}_{-1}} \left[M_1(x, s_2, \ldots, s_n)\right].$$

Since $\mathcal{M}$ has truth-telling as the strict Bayes Nash equilibrium, reporting $x = s_1^{(1)} = 1$ should be the unique maximizer for ϕ in the first scenario: $\phi(1) > \phi(x)$ for any $x \neq 1$. This, in particular, implies $\phi(1) > \phi(0.5)$. Thus, agent 1 will receive higher expected payment if he misreports his signal to be 1 instead of the true signal $s_1^{(2)} = 0.5$ in the second scenario. $\square$

C.2 Proof of Theorem 5.2

Proof. Suppose for the sake of contradiction that such a mechanism $\mathcal{M}$ exists. We consider an one-dimensional case where each $s_i \in \mathbb{R}$. Without loss of generality, assume $c_1 > 0$. Fix any $m_0 \in \mathbb{R}$, and consider the common prior $\mathbf{P}^{(m_0)} = (n, m_0, 1, 1)$. Consider a class of scenarios characterized by two arbitrary numbers $m_0, \bar{m}_0 \in \mathbb{R}$: $\mathbf{P}^{(m_0)}$ is the common prior and agent 1 receives signal $s_1 = 2\bar{m}_0 - m_0$. Applying the results in Proposition 2.3 in the one-dimensional case, for any scenarios in this class, agent 1 will have the same posterior $\mathbf{Q}_{-1} = (n-1, \bar{m}_0, \frac{1}{2}, \frac{1}{2})$. For each $\bar{m}_0 \in \mathbb{R}$, define function $\phi_{\bar{m}_0} : \mathbb{R} \mapsto \mathbb{R}^+$ as

$$\phi_{\bar{m}_0}(x) = \mathop{\mathbb{E}}_{(s_2,\ldots,s_n)\sim\mathbf{Q}_{-1}=(n-1,\bar{m}_0,\frac{1}{2},\frac{1}{2})} \left[M_1(x, s_2, \ldots, s_n)\right].$$

Since truth-telling is a Bayes Nash equilibrium, for any $\bar{m}_0$ and any $x \in \mathbb{R}$, reporting x should be the best response for agent 1 in the scenario characterized by $m_0 = 2\bar{m}_0 - x$ and $\bar{m}_0$, implying $\phi_{\bar{m}_0}$ reaches global maximum at x. Since x is arbitrary, $\phi_{\bar{m}_0}$ is a constant function for any $\bar{m}_0$, and let $P_{\bar{m}_0} = \phi_{\bar{m}_0}(x)$. Therefore, if the common prior is $\mathbf{P}^{(0)}$ and the signal is s_1, we are in the scenario in the above class which is characterized by $m_0 = 0$ and $\bar{m}_0 = 0.5s_1$, and agent 1 always receives the same payment $P_{0.5s_1}$ regardless of what he reports. As a result, in a best response, agent 1 will not bother to learn the value of s_1 and will just report an arbitrary real number to the mechanism, which saves him the cost c_1 from obtaining the signal. Notice that, in particular, agent 1's payment depends on the signal s_1 he receives, but not on what agent 1 reports to the mechanism. Thus, we have identify a scenario (when the common prior is $\mathbf{P}^{(0)} = (n, 0, 1, 1)$) in which for agent 1, when other agents are truth-telling, truthful strategy $\theta_1(\mathbf{P}, s_1) = s_1$ is strictly dominated. $\square$

C.3 Proof of Theorem 5.3

Proof. **Strictly truthful:** We first analyze the best response for agent i at truth-telling strategy profile $\boldsymbol{\theta}$. Note that if everyone tell the truth $\hat{s}_j = s_j$ which is a two-step Gaussian distribution, and the probability of $s_j = s_i$ for any $k \neq i$ is 0. So, a strategy θ_i is a best response if it minimize the quadratic form

$$\mathbb{E}[\|\hat{s}_j - \hat{s}_i\|_2^2 | s_i] = \mathbb{E}[(s_j - \hat{s}_i)^\top \mathbb{I}(s_j - \hat{s}_i) | s_i]$$

Because identity matrix $\mathbb{I}$ is positive definite, and s_j and $\hat{s}_i$ are independent conditioned on s_i, by proposition 2.15, it is minimized when

$$\mathop{\mathbb{E}}_{j\in B} [s_j - \hat{s}_i | s_i] = 0 \tag{36}$$

$$\mathbb{E}\left[\mathrm{Tr}(\mathrm{Cov}(\hat{s}_i)) | s_i\right] = 0, \tag{37}$$

so $\mathbb{E}[\theta_i(s_i)] = \mathbb{E}_{j\in B}[s_j|s_i] = s_i$ due to the Jeffreys prior. As a result, truth-telling is a strict Bayesian Nash equilibrium.

Truth-telling has the maximum welfare: We want to show the social welfare of the truth-telling equilibrium is greater (or equal) to that of all other non-oblivious strategy Bayesian Nash equilibrium.

Given any fixed report profile $\hat{\mathbf{s}}$ such that $\forall i \neq k\ \hat{s}_i \neq \hat{s}_k$, we consider the expected social welfare with respect to the randomness over the choice of reference agents j, k: with even n

$$\begin{aligned}
\sum_i M_i(\mathbf{s}) &= \sum \mathop{\mathbb{E}}_{j,k}\left[\|\hat{s}_j - \hat{s}_k\|_2^2 - \|\hat{s}_j - \hat{s}_i\|_2^2\right] \\
&= 2\sum_{j\in A, k\in B}\left(\left(\sum_{i\in A\setminus\{j\}} \frac{4\|\hat{s}_j - \hat{s}_k\|_2^2}{n(n-2)}\right) - \frac{2\|\hat{s}_j - \hat{s}_k\|_2^2}{n} - \frac{2\|\hat{s}_k - \hat{s}_j\|_2^2}{n}\right) \\
&= 0
\end{aligned}$$

Informed truthful: Finally, we show all oblivious equilibra have strictly less welfare. Suppose there is an oblivious equilibrium with strategy profile θ, by definition of oblivious $\theta_i(s_i)$ is independent to s_i, so $\hat{s}_i$ are independent samples from random variables θ_i. Suppose $\exists i, k \in A$, $\hat{s}_i \neq \hat{s}_k$ and, without loss of generality, $\hat{s}_i \neq \mathbb{E}_{j\in B}[\theta_j]$, it is not an equilibrium because there exists $s_i^* \triangleq (\hat{s}_i + \mathbb{E}_{j\in B}[\theta_j])/2$ which has larger expected payment by (36). Therefore $\forall i, k \in A$, $\hat{s}_i = \hat{s}_k$ and the average welfare is -100. $\square$

D Proofs in Sect. 6

Proof of lemma 6.5. The first one is the maximize likelihood estimator β_{MLE} which is a consistent estimator for state parameter β because the distribution of $\phi(X)$ is nondegerated,

$$b(i) = (\Phi_{-i}^\top \Phi_{-i})^{-1}\Phi_{-i}^\top y_{-i} \xrightarrow{p} \beta.$$

Moreover since $b(i) \to \beta$, equation (25) is a consistent estimator for variance τ^2.

$$t(i) = \frac{1}{|A|-2}\sum_{k\in A\setminus\{i\}} (y_k - b(i)^\top\phi(x_k))^2 \xrightarrow{p} \tau^2.$$

Formally, we can use triangle inequality, and use $(y_k - b(i)^\top\phi(x_k))^2 = (y_k - \beta^\top\phi(x_k))^2 + 2(y_k - \beta^\top\phi(x_k))(\beta^\top\phi(x_k) - b(i)^\top\phi(x_k)) + (\beta^\top\phi(x_k) - b(i)^\top\phi(x_k))^2$.

$$\begin{aligned}
&\left|\frac{1}{|A|-2}\sum_{k\in A\setminus\{i\}} (y_k - b(i)^\top\phi(x_k))^2 - \tau^2\right| \\
\leq &\left|\frac{1}{|A|-2}\sum_{k\in A\setminus\{i\}} (y_k - \beta^\top\phi(x_k))^2 - \tau^2\right| \qquad (38)\\
+ &\left|\frac{1}{|A|-2}\sum_{k\in A\setminus\{i\}} 2(y_k - \beta^\top\phi(x_k))(\beta^\top\phi(x_k) - b(i)^\top\phi(x_k))\right| \qquad (39)\\
+ &\left|\frac{1}{|A|-2}\sum_{k\in A\setminus\{i\}} (\beta^\top\phi(x_k) - b(i)^\top\phi(x_k))^2\right|. \qquad (40)
\end{aligned}$$

To show (29), it is sufficient to prove (38), (39), and (40) all converge to 0 in probability as $|A| \to \infty$. Because $\{y_k - \beta^\top \phi(x_k)\}_k$ are independent Gaussian, (38) converges to 0 by law of large numbers. For (40),

$$\begin{aligned}
(40) &\le \frac{1}{|A|-2} \sum_{k \in A \setminus \{i\}} \left| (\beta^\top \phi(x_k) - b(i)^\top \phi(x_k))^2 \right| \\
&\le \frac{1}{|A|-2} \sum_{k \in A \setminus \{i\}} \|\beta - b(i)\|^2 \|\phi(x_k)\|^2 \\
&= \left(\frac{1}{|A|-2} \sum_{k \in A \setminus \{i\}} \|\phi(x_k)\|^2 \right) \|\beta - b(i)\|^2 \\
&\xrightarrow{p} \mathbb{E}\left[\|\phi(x_k)\|^2\right] \cdot 0. && \text{(the forth moment of } \phi(X) \text{ is bounded)}
\end{aligned}$$

For (39),

$$\begin{aligned}
(39) &\le \frac{1}{|A|-2} \sum_{k \in A \setminus \{i\}} 2 \left| (y_k - \beta^\top \phi(x_k))(\beta^\top \phi(x_k) - b(i)^\top \phi(x_k)) \right| \\
&\le \left(\frac{1}{|A|-2} \sum_{k \in A \setminus \{i\}} |y_k - \beta^\top \phi(x_k)| \cdot \|\phi(x_k))\| \right) 2\|\beta - b(i)\| \\
&\xrightarrow{p} \mathbb{E}\left[|y_k - \beta^\top \phi(x_k)| \cdot \|\phi(x_k))\|\right] \cdot 0.
\end{aligned}$$

Finally for (30), when agent i tells the truth by Proposition 6.2 $\beta_i = \mathbb{E}[\beta | y_i] = S_i \left(\sigma^{-2}\beta_0 + \frac{y_i}{\tau^2}\phi(x_i)\right)$ where $S_i^{-1} = \sigma^{-2} + \frac{1}{\tau^2}\phi(x_i)\phi(x_i)^\top$. Apply Sherman Morrison formula in proposition 2.13, $S_i = \sigma^2 - \frac{\sigma^2 \phi(x_i)\phi(x_i)^\top \sigma^2}{\tau^2 + \phi(x_i)^\top \sigma^2 \phi(x_i)}$, and

$$\begin{aligned}
\beta_i &= S_i \left(\sigma^{-2}\beta_0 + \frac{y_i}{\tau^2}\phi(x_i)\right) \\
&= \left(\sigma^2 - \frac{\sigma^2 \phi(x_i)\phi(x_i)^\top \sigma^2}{\tau^2 + \phi(x_i)^\top \sigma^2 \phi(x_i)}\right) \left(\sigma^{-2}\beta_0 + \frac{y_i}{\tau^2}\phi(x_i)\right) \\
&= \left(\mathbb{I} - \frac{w_i \phi(x_i)^\top}{\tau^2 + \phi(x_i)^\top \sigma^2 \phi(x_i)}\right) \left(\beta_0 + \frac{y_i}{\tau^2} w_i\right) && (\text{set } w_i \triangleq \sigma^2 \phi(x_i)) \\
&= \beta_0 + \left(\frac{y_i}{\tau^2}\left(1 - \frac{w_i^\top \phi(x_i)}{\tau^2 + \phi(x_i)^\top \sigma^2 \phi(x_i)}\right) - \frac{\beta_0^\top \phi(x_i)}{\tau^2 + \phi(x_i)^\top \sigma^2 \phi(x_i)}\right) w_i \\
&= \beta_0 + \left(\frac{y_i - \beta_0^\top \phi(x_i)}{\tau^2 + \phi(x_i)^\top \sigma^2 \phi(x_i)}\right) \sigma^2 \phi(x_i)
\end{aligned}$$

which shows the posterior β_i is an affine transformation of β_0 and y_i.

On the other hand, if we multiply both sides with $\phi(x_i)^\top$ and let $z_i \triangleq \phi(x_i)^\top \sigma^2 \phi(x_i)$,

$$\begin{aligned}
\phi(x_i)^\top \beta_i &= \phi(x_i)^\top \beta_0 + \left(\frac{y_i - \beta_0^\top \phi(x_i)}{\tau^2 + \phi(x_i)^\top \sigma^2 \phi(x_i)}\right) \phi(x_i)^\top \sigma^2 \phi(x_i) \\
\beta_i^\top \phi(x_i) &= \beta_0^\top \phi(x_i) + \left(\frac{y_i - \beta_0^\top \phi(x_i)}{\tau^2 + z_i}\right) z_i \\
z_i &= \frac{\beta_i^\top \phi(x_i) - \beta_0^\top \phi(x_i)}{y_i - \beta_i^\top \phi(x_i)} \cdot \tau^2.
\end{aligned}$$

Therefore, combine these two equality we have for all i,

$$\sigma^2 \phi(x_i) = \frac{\tau^2}{y_i - \beta_i^\top \phi(x_i)} (\beta_i - \beta_0) \tag{41}$$

which are equal to (26) as $n \to \infty$. Therefore σ^2 and β_0 is a solution of $S_0(i)$ and $b_0(i)$ as $n \to \infty$. Since $\phi(X)$ is non-degenerated, the system of equation (41) has unique solution with probability 1 which finish the proof. ☐

Proof of Theorem 6.6. The most part of this proof is the same as the proof of Theorem 4.1. We will show one-by-one that $\mathcal{M}$ is strictly truthful, truth-telling has the maximum welfare, and informed truthful. Fix an arbitrary Gaussian linear prior $G(n, \beta_0, \sigma^2, \tau^2)$.

To show the strict truthfulness, fix an arbitrary agent in an arbitrary group, say agent i in group A, and assume the remaining $n-1$ agents play the truthful strategy. We need to show that truth-telling is the unique best response for agent i.

Upon receiving the private signal y_i, agent i believes the signal received by each other agent j, including agents $b_0, b_1, \ldots, b_d$, is from the Gaussian distribution $\mathcal{N}(\beta_i^\top \phi(x_j), \phi(x_i)^\top S_i \phi(x_j) + \tau^2)$ where $S_i^{-1} = \sigma^{-2} + \frac{1}{\tau^2}\phi(x_i)\phi(x_i)^\top$ and $\beta_i = S_i\left(\sigma^{-2}\beta_0 + \frac{y_i}{\tau^2}\phi(x_i)\right)$. (see Proposition 6.2). Firstly, we show that $(\hat{\beta}_i, \xi_i) = (\beta_i, 0)$ is the unique maximizer to the prediction score. It is obvious that agent i should report $\xi_i = 0$ to avoid the $-\xi_i$ punishment. The expected prediction score (that agent i believes) for i is then

$$\sum_{j=0}^{d} \mathop{\mathbb{E}}_{y_{b_j} \sim \mathcal{N}(\beta_i^\top \phi(x_{b_j}), \phi(x_i)^\top S_i \phi(x_{b_j}) + \tau^2)} \left[- \left\| y_{b_j} - \hat{\beta}_i^\top \phi(x_{b_j}) \right\|^2 \right],$$

which, by Proposition 2.15 (taking A as the identity matrix), is maximized when $\hat{\beta}_i^\top \phi(x_{b_j}) = \beta_i^\top \phi(x_{b_j})$ for each $j = 0, 1, \ldots, d$. Since we have assumed that Φ sampled from some non-degenerate distribution over $\mathbb{R}^{d+1}$, with probability 1 that $\phi(x_{b_0}), \phi(x_{b_1}), \ldots, \phi(x_{b_d})$ are linearly independent, so the system of linear equations $\{\hat{\beta}_i^\top \phi(x_{b_j}) = \beta_i^\top \phi(x_{b_j})\}_{j=0,1,\ldots,d}$ has unique solution $\hat{\beta}_i = \beta_i$.

Secondly, we show that, fixing $(\hat{\beta}_i, \xi_i) = (\beta_i, 0)$, reporting $\hat{y}_i = y_i$ is the only way to make the inconsistency score 0. Since agents $b_0, b_1, \ldots, b_{d+1}$ are truth-telling, by Proposition 6.2, we have

$$\forall j \in \{0, 1, \ldots, d+1\}: \qquad S_{b_j}^{-1} \hat{\beta}_{b_j} = \sigma^{-2}\beta_0 + \frac{1}{\tau^2} \hat{y}_{b_j} \phi\left(x_{b_j}\right).$$

Substituting $S_{b_j}^{-1} = \sigma^{-2} + \frac{1}{\tau^2}\phi(x_{b_j})\phi(x_{b_j})^\top$ into above and by rearranging terms, we have

$$\forall j \in \{0, 1, \ldots, d+1\}: \qquad \tau^2\sigma^{-2}\hat{\beta}_{b_j} = \tau^2\sigma^{-2}\beta_0 + \delta_{b_j}, \tag{42}$$

which implies

$$\forall j \in \{1, \ldots, d+1\}: \qquad \tau^2\sigma^{-2}(\hat{\beta}_{b_j} - \hat{\beta}_{b_0}) = (\delta_j - \delta_0),$$

which further implies $\tau^2\sigma^{-2}T_b = U_b$. With probability 1, T_b is non-degenerate, and $\tau^2\sigma^{-2} = U_b T_b^{-1}$. The inconsistency score for i is

$$\begin{aligned}
\mathrm{IS}_i(\mathbf{r}) &= -\left\| U_b T_b^{-1}(\hat{\beta}_i - \hat{\beta}_{b_0}) - (\delta_i - \delta_{b_0}) \right\| \\
&= -\left\| \tau^2\sigma^{-2}\hat{\beta}_i + \left(\delta_{b_0} - \tau^2\sigma^{-2}\hat{\beta}_{b_0}\right) - \delta_i \right\| \\
&= -\left\| \tau^2\sigma^{-2}\hat{\beta}_i - \tau^2\sigma^{-2}\beta_0 - \delta_i \right\| && \text{(by (42))} \\
&= -\left\| \tau^2\sigma^{-2}\beta_i - \tau^2\sigma^{-2}\beta_0 - \left(\hat{y}_i\phi(x_i) - \phi(x_i)\phi(x_i)^\top\beta_i\right) \right\| && \text{(substituting } \delta_i \text{ and } \hat{\beta}_i = \beta_i) \\
&= -\tau^2 \left\| S_i^{-1}\beta_i - \sigma^{-2}\beta_0 - \frac{1}{\tau^2}\hat{y}_i\phi(x_i) \right\|. && \text{(substituting } S_i^{-1} = \sigma^{-2} + \tfrac{1}{\tau^2}\phi(x_i)\phi(x_i)^\top)
\end{aligned}$$

We know from Proposition 6.2 that $S_i^{-1}\beta_i - \sigma^{-2}\beta_0 - \frac{1}{\tau^2}y_i\phi(x_i) = 0$. Since $\frac{1}{\tau^2}\phi(x_i)$ is nonzero with probability 1, $\hat{y}_i = y_i$ is the only value making the inconsistency score 0. Thus, fixing $(\hat{\beta}_i, \xi_i) = (\beta_i, 0)$, agent i's unique best strategy for signal is $\hat{y}_i = y_i$. Combining the arguments above, fixing the other $n-1$ agents' truthful report $\mathbf{r}_{-i}$, reporting $r_i = (y_i, \beta_i, 0)$ is the unique maximizer to $\bar{M}_i(\cdot, \mathbf{r}_{-i})$.

The remaining part of the proof is identical to the proof of Theorem 4.1: to conclude strict truthfulness, we can show that $M_i(\mathbf{r}) - \bar{M}_i(\mathbf{r})$ does not depend on i's report, so $r_i = (y_i, \beta_i, 0)$ is the unique best response if the other $n-1$ agents are truth-telling; Step 6 and 7 make sure the maximum welfare is 0, so the truth-telling profile has the maximum welfare; the same arguments in the proof of Theorem 4.1 can show that the game has no oblivious Bayes Nash equilibrium (if all agents in a group report the same posterior mean, one of them can perturb the reported mean a little bit to avoid the punishment -100; if not, some agent i can improve his prediction score by reporting some β_i' which are nearly optimal and confessing some very small $\xi_i > 0$ to avoid inconsistency punishment). $\square$