Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 58-30 (2024) 314–319

On the Impact of Bounded Rationality in Strategic Data Gathering ⋆

Anju Anand * **Emrah Akyol** **

* *Binghamton University–SUNY at Binghamton, NY, 13902 USA (e-mail: aanand6@binghamton.edu).*
** *Binghamton University–SUNY at Binghamton, NY, 13902 USA (e-mail: eakyol@binghamton.edu).*

Abstract: This paper is concerned with the problem of debiasing survey data gathered from strategic and boundedly rational agents with heterogeneous objectives and partial information. Particularly, we explore a setting where there are K different types of survey responders with varying levels of available information, degree of strategic behavior, and cognitive hierarchy: i) a non-strategic agent with an honest response, ii) a $k-$th level $k \in [1 : K-1]$ strategic agent that believes the population is Poisson distributed over the lower cognitive types. We model each of these scenarios as a strategic classification of a 2-dimensional source (possibly correlated source and bias components) with quadratic distortion measures and provide a design method. We analyze the numerical results obtained via the proposed method whose implementation is available for research purposes at `https://github.com/strategic-quantization/bounded-rationality`.

Keywords: Game theory, strategic quantization, bounded rationality.

1. INTRODUCTION

Consider designing a survey to gauge public reception of a new plastic product, with responses influenced by respondents' attitudes toward climate change. Respondents' scores range from 1 ('will definitely not use') to 4 ('will definitely use'), and the survey needs to account for potential biases as well as varying levels of rationality among respondents. We model this problem using the hierarchical cognitive type model as studied by Camerer et al. (2004), considering three types of respondents:

- Type 0 (Honest-Nonstrategic Respondents): These respondents provide truthful information based on their actual opinions about the product, unaffected by their considerations of climate change or any desire to bias the survey.
- Type k, $k \in [0 : K-1]$: These respondents wish to influence the survey outcome correlated with their attitudes. They best respond to a mix of the lower types, assuming that the estimator (designer) is also only aware of the lower types.

The designer of the survey is aware of the existence of these types of respondents as well as their true statistics. The question explored in this paper is: What is the designer's optimal "de-biasing" procedure, i.e, optimally (in Bayesian sense) estimating the unbiased scores that reflect the true public reception of the plastic product?

We approach this problem via the recently introduced strategic quantization framework, see Akyol and Anand (2023), which is a special case of the information design problem in Economics.[1] This class of problems, notable studied by Kamenica and Gentzkow (2011); Rayo and Segal (2010) explore the use of information by an agent (sender) to influence the action taken by another agent (receiver), where the aforementioned action determines the payoffs for both agents. Our prior work explored strategic quantization problem settings where the sender and the receiver were assumed to be fully rational agents. In this paper, we extend our strategic quantization work to settings with boundedly-rational sender (quantizer), via employing the cognitive hierarchy model of Camerer et al. (2004).

Throughout this paper, we focus on the quadratic distortion measures. Particularly, the senders observe a two-dimensional source $(X, S) \sim f_{X,S}(\cdot, \cdot)$ with a known joint density function over X and S, where X and S can be interpreted as the state and bias variables. There is one honest sender type, and $K-1$ types of strategic senders, each trying to minimize $\mathbb{E}\{(X + S - Y)^2\}$, with different assumptions on the estimator (receiver), where Y is the action taken by the receiver upon observing the quantization index $Z = Q(X, S)$ sent by the sender.

The receiver's objective is to estimate the true state in the minimum mean squared error (MME) sense, i.e., the receiver minimizes $\eta(x, y) = (x - y)^2$ by choosing an action $\hat{X}$ which is the optimal MMSE estimate of x given the quantization index from the sender $z = Q(x, s)$, hence $\hat{X} = \mathbb{E}\{X|Z = z\}$. In sharp contrast with the conventional quantization problem where the sender chooses Q that minimizes $\mathbb{E}\{(X - \hat{X})^2\}$, in this setting

⋆ This research is supported by the NSF via grants CCF #1910715 and CAREER #2048042.

[1] Throughout the paper, we use the terms quantizer and classifier interchangeably.

10.1016/j.ifacol.2025.01.200

the sender's choice of quantization mapping Q minimizes a biased estimate, i.e., $\mathbb{E}\{(X+S-\hat{X})^2\}$. The objectives and the source distribution are common knowledge, available for all agents. We note that similar signaling problems with quadratic measures have been analyzed in the Economics literature, see e.g., Bénabou and Tirole (2006); Crawford and Sobel (1982); Fischer and Verrecchia (2000).

This paper is organized as follows: In Section II we state the notation. In Section III we present the problem formulation. In Section IV, we present a design algorithm to compute the classifier implemented by the boundedly rational agent. We provide numerical results in Section V, and conclude in Section VI.

2. PRELIMINARIES

2.1 Notation

In this paper, random variables are denoted using capital letters (say X), their sample values with respective lowercase letters (x), and their alphabet with respective calligraphic letters ($\mathcal{X}$). Vectors are denoted in bold font. The set of real numbers is denoted by $\mathbb{R}$. The 2-dimensional jointly Gaussian probability density function with mean $[t_1\ \ t_2]'$ and respective variances σ_1^2, σ_2^2 with a correlation ρ is denoted by $\mathcal{N}\left(\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}\right)$, $t_1, t_2 \in \mathbb{R}$. The expectation operator is written as $\mathbb{E}\{\cdot\}$. The operator $|\cdot|$ denotes the cardinality with the argument as a set.

3. PROBLEM FORMULATION

Consider the following classification problem: K classifiers, sender $k, k \in [0 : K-1]$ observe realizations of the two sources $X \in \mathcal{X} \subseteq [a_X, b_X], S \in \mathcal{S} \subseteq [a_S, b_S]$, $a_X, b_X, a_S, b_S \in \mathbb{R}$ with joint probability density $(X,S) \sim f_{X,S}(\cdot,\cdot)$. One of the classifier's, sender k, is chosen with probability p_k. The chosen sender k maps (X,S) to a message $Z \in \mathcal{Z}$, where $\mathcal{Z}$ is a set of discrete messages with a cardinality constraint $|\mathcal{Z}| \le M$ using a non-injective mapping $Q^{(k)} : (\mathcal{X} \times \mathcal{S}) \to \mathcal{Z}$, $Q^{(k)} \in \mathcal{Q}$, where $\mathcal{Q}$ is the set of possible classifiers with the cardinality constraint $|\mathcal{Z}| \le M$. After receiving the message Z, the receiver applies a mapping $\phi : \mathcal{Z} \to \mathcal{Y}$ on the message Z and takes an action $Y = \phi(Z)$.

The set $\mathcal{X}$ is divided into mutually exclusive and exhaustive sets by each classifier k as $\mathcal{V}_1^{(k)}, \mathcal{V}_2^{(k)}, \ldots, \mathcal{V}_M^{(k)}$.

The probability p_k of sender k being chosen follows a normalized Poisson distribution

$$p_k = \frac{e^{\lambda}\frac{\lambda^k}{k!}}{\sum_{i=0}^{K-1} e^{\lambda}\frac{\lambda^i}{i!}}, \tag{1}$$

where the parameter $\lambda \in \mathbb{R}_+$ indicates the cognitive levels of the overall population. As λ increases, the population consists of higher cognitive levels.

The belief of sender k about the probability distribution over sender t, $t \in [0 : k-1]$ is represented as $\mu_t^{(k)}$,

$$\mu_t^{(k)} = \frac{e^{\lambda}\frac{\lambda^t}{t!}}{\sum_{i=0}^{k-1} e^{\lambda}\frac{\lambda^i}{i!}}, \quad t \in [0 : k-1],$$

$\mu_t^{(k)} = 0$ for $t \ge k$ and $\mu_0^{(0)} = 1$.

Remark 1. Note that this belief $\boldsymbol{\mu}^{(k)}$ is not the true statistics of the population, $\boldsymbol{\mu}^{(k)} \neq \mathbf{p}$. A level-$k$ sender can only estimate the relative proportions of the lower levels accurately, i.e., it can estimate $\mu_{t_1}^{(k)}/\mu_{t_2}^{(k)} = p_{t_1}/p_{t_2} = \frac{\lambda^{t_1}/t_1!}{\lambda^{t_2}/t_2!}$, $t_1, t_2 \in [0 : k-1]$.

The sender k's distortion measure is $\eta_E^{(k)}(x,s,y)$,

$$\eta_E^{(k)}(x,s,y) = \begin{cases} (x-y)^2, & \text{if } k = 0 \\ (x+s-y)^2, & \text{otherwise.} \end{cases}$$

We take the receiver's distortion measure as $\eta_D(x,y) = (x-y)^2$.

We consider senders with K hierarchical cognitive types and define the senders' and their respective perceived receiver distortions as $D_E^{(k)}, D_D^{(k)}, k \in [0 : K-1]$ below:

$$D_E^{(k)} = \sum_{m=1}^{M} \int_{(x,s)\in\mathcal{V}_m^{(k)}} \eta_E^{(k)}(x,s,y_m^{(k)}) f_{X,S}(x,s)\mathrm{d}x\mathrm{d}s,$$

$$D_D^{(k)} = \sum_{i=0}^{k-1} \mu_i^{(k)} \sum_{m=1}^{M} \int_{(x,s)\in\mathcal{V}_m^{(k)}} \eta_D(x,y_m^{(k)}) f_{X,S}(x,s)\mathrm{d}x\mathrm{d}s,$$

where $\mathbf{y}^{(k)}$ are the estimates that sender k assumes are optimized by the receiver with respect to the respective perceived receiver distortion $D_D^{(k)}$, obtained by enforcing KKT conditions of optimality, $\partial D_D^{(k)}/\partial y_m^{(k)} = 0, m \in [1 : M]$. [2]

Remark 2. The perceived receiver distortion $D_D^{(k)}$ is only a function of the quantizers of the lower cognitive levels $\{\mathcal{V}_m^{(i)}, i \in [0 : k-1], m \in [1 : M]\}$, i.e., the perceived receiver action $\mathbf{y}^{(k)}$ does not depend on the $k-$th level quantizer.

Our problem simplifies to the following optimization at each sender of cognitive level k:

$$Q^{(k)} = \underset{Q\in\mathcal{Q}}{\arg\min}\, \mathbb{E}\{\eta_E^{(k)}(X,S,Y(Q^{(0)},\ldots,Q^{(k-1)}))\},$$

where Y is a function of $\{Q^{(t)}, t \in [0 : k-1]\}$, and is hence independent of changes in $Q^{(k)}$.

The different cognitive level type senders, and their classifiers are as follows:

(1) Non-strategic sender 0: similar to level L_0 cognitive type, the sender assumes all senders are of level 0, and that the receiver assumes all senders are of level 0. Sender 0 considers the receiver's distortion as the same as the sender's, $D_E^{(0)} = D_D^{(0)}$ (provides the information required by the receiver honestly)

$$D_D^{(0)} = \sum_{m=1}^{M} \int_{(x,s)\in\mathcal{V}_m^{(0)}} \eta_D(x,s,y_m^{(0)}) f_{X,S}(x,s)\mathrm{d}x\mathrm{d}s.$$

(2) Level-k strategic sender k: similar to level L_k cognitive type, the sender assumes all other senders are distributed over the lower cognitive types $t, t \in [0 : k-1]$

[2] Note that $\partial D_D^{(k)}/\partial y_m^{(k)}$ here is the first order derivative of sender k's distortion with respect to $y_m^{(k)}$, and not the $k-$th order derivative.

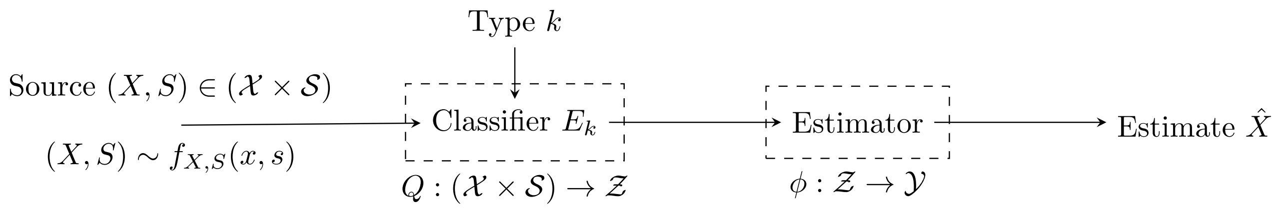


Fig. 1. Communication diagram: with probability p_k, sender type k sends a message Z which is a function of the source (X, S) over a noiseless channel.

with probability $\mu_t^{(k)}$ and that it is uniquely of type k. The sender assumes the receiver thinks that all sender types are of types $t, t \in [0 : k-1]$ as well. This results in the estimates perceived by sender k being a function of only $\{Q^{(t)}, t \in [0 : k-1]\}$, i.e., $\mathbf{y}^{(k)}$ does not change as $Q^{(k)}$ changes. The sender k assumes a fixed $\mathbf{y}^{(k)}$ as a function of $\{Q^{(t)}, t \in [0 : k-1]\}$, and the resulting quantizer is found by enforcing KKT optimality conditions.

The receiver's distortion is given by

$$D_D^* = \min_{\mathbf{y}} \sum_{i=0}^{K-1} p_i \sum_{m=1}^{M} \int_{(x,s)\in\mathcal{V}_m^{(i)}} \eta_t^{(i)}(x, s, y_m) f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s,$$

and $\mathbf{y}$ that minimizes the above expression is the actual receiver's action, $\mathbf{y}^*$.

Each sender type $k, k \in [1 : K-1]$ optimizes its classifier $Q^{(k)}$ to minimize $D_E^{(k)}$, assuming the receiver is aware of only $i, i < k$ sender types. Sender $k, k \in [0 : K-1]$ designs $Q^{(k)}$ *ex-ante*, i.e., without the knowledge of the realization of (X, S), using only the objectives $D_E^{(k)}$ and $D_D^{(k)}$, and the statistics of the source $f_{X,S}(\cdot, \cdot)$.

The receiver is fully rational and has full information about the classification setup. The shared prior ($f_{X,S}$), the probability mass function over the sender types ($\mathbf{p} = [p_0, p_1, \dots, p_{K-1}]$) and the mappings ($\mathbf{Q} = \{Q^{(k)}, k \in [0 : K-1]\}$) are known to the receiver. The problem, as depicted in Fig. 1, is to design the classifiers $\mathbf{Q}$ for the equilibrium, i.e., each sender type k minimizes its own objective, assuming that the receiver minimizes its corresponding perceived objective $D_D^{(k)}$. Since the senders choose the classifiers $\mathbf{Q}$ first, followed by the receiver choosing the perceived estimates ($\mathbf{y}^{(k)}, k \in [0 : K-1]$), we look for a Stackelberg equilibrium.

The classifier design involves computing classifiers for each realization of S by classifier i as $\mathcal{U}_{s,m}^{(i)}, s \in \mathcal{S}$, where $\bigcup_{s\in\mathcal{S}} \mathcal{U}_{s,m}^{(i)} = \mathcal{V}_m^{(i)}$. Throughout this paper, we make the following assumption on the sets $\{\mathcal{U}_{s,m}^{(i)}\}$.

Assumption 1. $\mathcal{U}_{s,m}^{(i)}$ is convex for all $m \in [1 : M], s \in \mathcal{S}, i \in [0 : 2]$.

Then, $\mathcal{U}_{s,m}^{(i)} = [q_{s,m}^{(i)}, q_{s,m+1}^{(i)}], q_{s,m}^{(i)} < q_{s,m+1}^{(i)}$, and $Q^{(i)} = \{\mathbf{q}_s^{(i)}, s \in \mathcal{S}\}$, where $\mathbf{q}_s^{(i)} = [q_{s,0}^{(i)}, \dots, q_{s,M+1}^{(i)}]$.

Since sender 0's distortion function $\eta_E^{(0)}(x, s, y) = (x-y)^2$ is not a function of s, $Q^{(0)} : (\mathcal{X} \times \mathcal{S}) \to \mathcal{Z}$ simplifies to $Q^{(0)} : \mathcal{X} \to \mathcal{Z}$. Let the marginal probability density function of X be $f_X(x)$. Sender 0 responds honestly, $D_E^{(0)} = D_D^{(0)}$ (equivalent to the non-strategic classification setting), hence its classifier $\mathbf{q}_s^{(0)}$ and perceived estimates $\mathbf{y}^{(0)}$ are,

$$\mathbf{q}_s^{(0)} = \arg\min_{\mathbf{n}\in\mathcal{Q}} \sum_{m=1}^{M} \int_{n_{m-1}}^{n_m} (x - y_m^{(0)})^2 f_X(x)\mathrm{d}x, \quad \forall s \in \mathcal{S}$$

$$y_m^{(0)} = \arg\min_{y\in\mathcal{Y}} \int_{n_{m-1}}^{n_m} (x-y)^2 f_X(x)\mathrm{d}x = \mathbb{E}\{X | x \in \mathcal{V}_m^0\}.$$

Sender $k, k > 0$ assumes the other agents are of levels $i, i \in [0 : k-1]$ with probability mass function $\boldsymbol{\mu}^{(k)}$, and that the receiver also perceives the agents as the same, of levels $i, i \in [0 : k-1]$ with a probability mass function $\boldsymbol{\mu}^{(k)}$, with receiver distortion

$$D_D^{(k)} = \sum_{i=0}^{k-1} \mu_i^{(k)} \sum_{m=1}^{M} \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(i)}}^{q_{s,m}^{(i)}} (x - y_m^{(k)})^2 f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s,$$

resulting in sender k's perceived estimates $y_m^{(k)}$,

$$y_m^{(k)} = \frac{\sum_{i=0}^{k-1} \mu_i^{(k)} \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(i)}}^{q_{s,m}^{(i)}} x f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s}{\sum_{i=0}^{k-1} \mu_i^{(k)} \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(i)}}^{q_{s,m}^{(i)}} f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s}. \tag{2}$$

The classifier distortions for sender $k, k \in [1 : K-1]$

$$D_E^{(k)} = \sum_{m=1}^{M} \int_{a_S}^{b_S} \int_{q_{s,m-1}^{k}}^{q_{s,m}^{k}} (x + s - y_m^{(k)})^2 f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s.$$

The receiver's distortion and estimates are

$$D_D^* = \sum_{i=0}^{K-1} p_i \sum_{m=1}^{M} \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(i)}}^{q_{s,m}^{(i)}} (x - y_m^*)^2 f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s, \tag{3}$$

$$y_m^* = \frac{\sum_{i=0}^{K-1} p_i \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(i)}}^{q_{s,m}^{(i)}} x f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s}{\sum_{i=0}^{K-1} p_i \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(i)}}^{q_{s,m}^{(i)}} f_{X,S}(x, s)\mathrm{d}x\mathrm{d}s}. \tag{4}$$

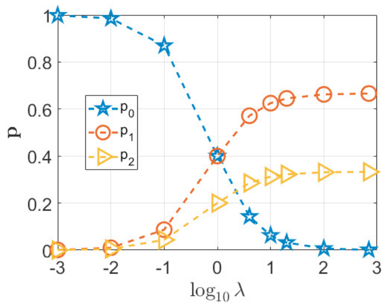


Fig. 2. Probability mass function $\mathbf{p}$ with respect to λ.

4. MAIN RESULTS

In this section, we present our design algorithm for the computation of $Q^{(k)}, k \in [0:K-1]$.

In Anand and Akyol (2024), we proposed a gradient-descent based algorithm to solve the problem of quantization of a 2-dimensional source (X, S) for fully rational sender and receiver with full information by extending our algorithm in Akyol and Anand (2023) for a scalar source to the 2-dimensional setting by a simple method of computing quantizers for each value of $s \in \mathcal{S}$.

Here, as in Anand and Akyol (2024) we compute classifiers for each realization of S. However, because $D_D^{(k)}$ does not depend on $Q^{(k)}$, the optimization simplifies to a nearest neighbor classification as shown in the Appendix A.

A sketch of the proposed method is summarized in the Algorithm. The algorithm takes the non-strategic quantizer $Q^{(0)}$ as an input, which can be computed by any classical quantization method like lloyd-max, gradient descent, etc. The MATLAB codes are provided at `https://github.com/strategic-quantization/bounded-rationality` for research purposes.

The classifers implemented by senders $k, k \in [0:K-1]$ are as follows:

(1) Sender 0 implements a non-strategic (classical) classifier for the source X.
(2) Sender $k, k \in [1:K-1]$ implements a nearest neighbor classifier for $X + s$ for each $s \in \mathcal{S}$ with respect to $\mathbf{y}^{(k)}$ as shown in the Appendix A.

This analysis can be extended to a general distortion measure, $\{\eta_E^{(k)}(x,s,y), \eta_D(x,y)\}$ where $\eta_E^{(0)}(x,s,y) = \eta_D(x,y)$ as shown in Appendix B.

Algorithm 1 Proposed strategic quantizer design

Parameters: λ
Input: $f_{X,S}(\cdot,\cdot), \mathcal{X}, \mathcal{S}, M, \{\eta_E^{(k)}, \eta_D, k \in [0:K-1]\}, Q^{(0)}, \mathbf{y}^{(0)}, \{\boldsymbol{\mu}^{(t)}, t \in [0:K-1]\}, \mathbf{p}$
Output: $\{Q^{(k)}, \mathbf{y}^{(k)}, k \in [1:K-1]\}, \mathbf{y}^*, D_D^*$
for $k \in [1:K-1]$ **do**
  Compute $\mathbf{y}^{(k)}$ from (2)
  Compute $Q^{(k)}$ from (A.1)
Compute D_D^* from (3) with $\mathbf{y}^*$ from (4).

5. NUMERICAL RESULTS

We consider a jointly Gaussian 2-dimensional source [3]

$$(X, S) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \sigma_S\rho\\ \sigma_S\rho & \sigma_S^2\end{bmatrix}\right)$$

and present results for different settings with parameters source variance $\sigma_X^2 = 1$, bias variance $\sigma_S^2 \in \{0.1, 1, 1.5\}$, correlation $\rho \in \{0.1, 0.5, 0.7\}$, and probability distribution of senders $\mathbf{p}$ following (1) with $\lambda \in [0.001, 700]$ for an $M \in \{1,2,4,8,16\}$ classifier with $K \in \{3,5,10,20\}$ highest cognitive level of senders. We consider discretized S due to its computational feasibility. The probability mass function over the sender types $\mathbf{p}$ for different values of λ is plotted in Fig. 2.

5.1 Varying the cognitive parameter λ

From Fig. 2 we note that as $\lambda \to 0$, the population mostly consists of level-0 cognitive level. As λ increases, the population shifts towards higher cognitive types, and we expect the receiver distortion to increase with λ, as we observe in Fig. 3. For $\lambda \to 0$, the receiver distortion does not change significantly with varying bias variance σ_S^2 since the population is mostly of level-0 type, and they respond honestly. For $\lambda > 100$, the statistics of the population remain fairly constant, and hence the receiver distortion varies negligibly.

5.2 Varying the bias variance σ_S^2

For a given correlation ρ, we observe in Fig. 3 that as σ_S^2 decreases, the receiver distortion decreases. As the variance of the bias σ_S^2 decreases, the objectives of the sender and the receiver become more aligned. If S is deterministic, the sender is honest or non-strategic for this setting, as shown in Akyol and Anand (2024).

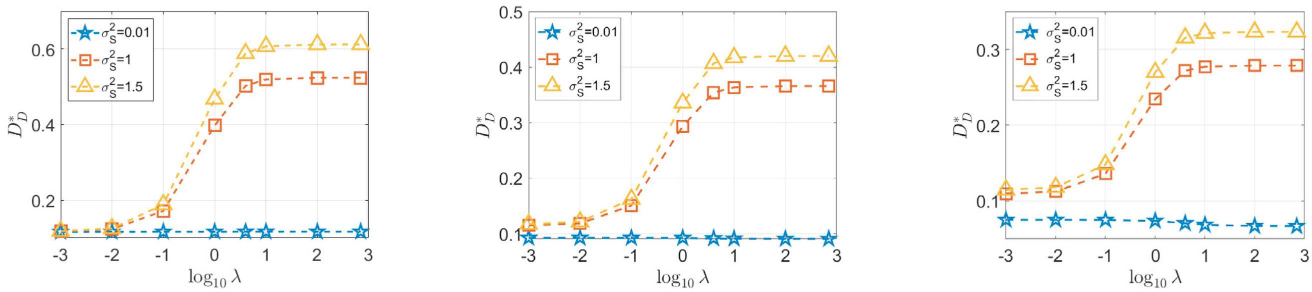
For small bias ($\sigma_S \to 0$), the objectives of all the senders are similar, resulting in a small change in the receiver distortion with λ, which we observe in Fig. 3 for $\sigma_S^2 = 0.01$.

5.3 Varying the types of senders

In Fig. 4, the receiver distortion for the following four different types of senders is plotted for a specific setting with $\sigma_X^2 = \sigma_S^2 = 1, \rho = 0.5$:

(1) non-strategic (S_n): All agents are non-strategic and send their honest reply (senders are of cognitive level 0). The distortion in this case is solely from the quantization aspect.
(2) full information (S_s): All agents are fully rational and have full information. The classifier here is that in Anand and Akyol (2024).
(3) bounded rational (S_b): The agents follow the setting described in this paper.
(4) level-1 strategic (S_{L_1}): All agents minimize $\mathbb{E}\{(X + S - Y)^2\}$, but they assume the receiver is not strategic and hence implements a naive estimator, $\mathbf{y}^{(0)}$. The classifier is the same as that for sender 1.

[3] Although we consider a jointly Gaussian source for numerical results here, this algorithm can be applied to any source distribution.



(a) $\rho = 0.1$ (b) $\rho = 0.5$ (c) $\rho = 0.7$

Fig. 3. Receiver distortion D_D^* for $M = 4$ classification of $(X, S) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \sigma_S\rho\\ \sigma_S\rho & \sigma_S^2\end{bmatrix}\right)$ for a given correlation ρ with respect to bias variance σ_S^2.

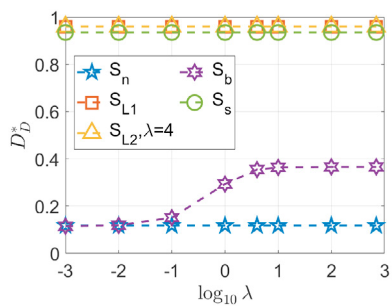


Fig. 4. Receiver distortion D_D^* for $M = 4$ classification of $(X, S) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0.5\\ 0.5 & 1\end{bmatrix}\right)$ for a full information fully rational estimator with five different types of senders.

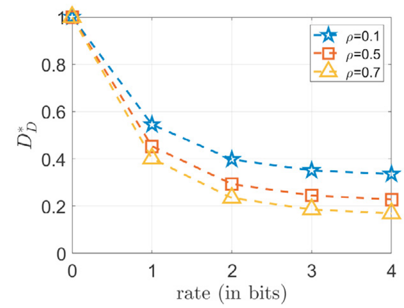


Fig. 6. Receiver distortion D_D^* for $M \in \{1, 2, 4, 8, 16\}$ classification of $(X, S) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right), \rho \in \{0.1, 0.5, 0.7\}$ for a full information fully rational estimator with $K = 3$ highest cognitive level of senders and $\lambda = 1$.

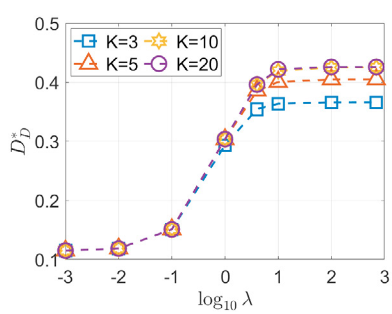


Fig. 5. Receiver distortion D_D^* for $M = 4$ classification of $(X, S) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0.5\\ 0.5 & 1\end{bmatrix}\right)$ for a full information fully rational estimator with $K \in \{3, 5, 10, 20\}$ highest cognitive levels of senders.

(5) level-2 strategic (S_{L_2}): All agents are level-2 strategic and assume the lower levels are normalized Poisson distributed, $\boldsymbol{\mu}^{(2)}$ with parameter λ. The classifier is the same as that for sender 2.

The receiver is fully rational with full information about the types of sender, the source distribution, and sender and receiver objectives.

As expected, the non-strategic sender results in the lowest receiver distortion. For small λ, the population is mostly of level-0 cognitive type, as mentioned before. Since they respond honestly, S_b is closer to S_n as $\lambda \to 0$.

Although we expect that S_s results in the maximum receiver distortion among the above four senders, we observe from Fig. 4 that the receiver may prefer a fully rational sender with full information to other types of partially strategic senders, S_{L_1} and S_{L_2}. We also observe that the receiver benefits from a boundedly rational setup compared to the setting where all senders are partially or fully strategic (S_{L_1}, S_{L_2}, or S_s).

5.4 Varying the highest cognitive level K

In Fig. 5 the receiver distortion is plotted for $K \in \{3, 5, 10, 20\}$ for the specific setting of (X, S) jointly Gaussian with $\sigma_X^2 = \sigma_S^2 = 1, \rho = 0.5$. As expected, the receiver distortion increases as K increases. We observe that for negligible λ, change in K does not change D_D^* much since most of the population is composed of lower types, but as λ increases, the receiver distortion changes with K. We also observe that D_D^* does not vary much for a high value of K. This is due to the convergence of beliefs $\boldsymbol{\mu}^{(k)}$ at high values of k as shown in Chong et al. (2005), which results in a convergence of classifiers.

5.5 Varying the rate M

In Fig. 6, the receiver distortion is plotted for $K = 3$ for a specific setting with $\sigma_X^2 = \sigma_S^2 = 1, \rho \in \{0.1, 0.5, 0.7\}$, $\lambda = 1$ and $M \in \{1, 2, 4, 8, 16\}$. We observe that while the distortion decreases with the rate, it does not vanish at high rates, as expected due to the strategic aspect of the problem (see e.g., Akyol et al. (2015)). We also observe that the decoder distortion decreases with correlation ρ. These observations are similar to our results in Anand and Akyol (2024) on strategic quantization of a 2-dimensional source with full information fully rational sender, where it is shown that the optimal sender is non-revealing for $\rho = 0$ and fully revealing at $\rho = 1$, i.e., the sender's ability to persuade the receiver decreases with ρ.

6. CONCLUSIONS

In this paper, we have analyzed the problem of strategic classification of a 2-dimensional source (X, S) with K types of senders with hierarchical cognitive types: level-0 non-strategic (honest), level-k strategic, $k \in \{1 : K - 1\}$. In conjunction with the quadratic objectives for the sender and the receiver, we have presented a method of computing the optimal classifiers for this hierarchical cognitive model of bounded-rationality. Our future work includes a theoretical analysis of the problems that we studied numerically in this paper.

Appendix A. COMPUTATION OF $Q^{(k)}$

We present the classifier computation implemented by sender k as follows. The sender k distortion

$$D_E^{(k)} = \sum_{m=1}^{M} \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(k)}}^{q_{s,m}^{(k)}} (x + s - y_m^{(k)})^2 f_{X,S}(x, s) \mathrm{d}x\mathrm{d}s,$$

where $y_m^{(k)}$ are from (2). We obtain $q_{s,m}^{(k)}, s \in \mathcal{S}, m \in [1 : M]$ by enforcing the optimality conditions:

$$\begin{aligned}\frac{\partial D_E^{(k)}}{\partial q_{s,m}^{(k)}} &= f_{X,S}(q_{s,m}^{(k)}, s)(q_{s,m}^{(k)} + s - y_m^{(k)})^2 \\ &\quad - f_{X,S}(q_{s,m}^{(k)}, s)(q_{s,m}^{(k)} + s - y_{m+1}^{(k)})^2, \\ q_{s,m}^{(k)} &= \frac{y_m^{(k)} + y_{m+1}^{(k)}}{2} - s. \qquad \text{(A.1)}\end{aligned}$$

Appendix B. COMPUTATION OF $Q^{(k)}$ FOR GENERAL DISTORTION MEASURES

Consider sender and receiver distortions $\eta_E^{(k)}(x, s, y)$, $\eta_D(x, y)$, where $\eta_E^{(0)}(x, s, y) = \eta_D(x, y)$. The distortion to sender k,

$$D_E^{(k)} = \sum_{m=1}^{M} \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(k)}}^{q_{s,m}^{(k)}} \eta_E^{(k)}(x, s, y_m^{(k)}) f_{X,S}(x, s) \mathrm{d}x\mathrm{d}s,$$

where $y_m^{(k)}$ are from minimizing

$$y_m^{(k)} = \arg\min_{y \in \mathcal{Y}} \sum_{i=0}^{k-1} \mu_i^{(k)} \int_{a_S}^{b_S} \int_{q_{s,m-1}^{(i)}}^{q_{s,m}^{(i)}} \eta_D(x, y_m^{(k)}) f_{X,S}(x, s) \mathrm{d}x\mathrm{d}s.$$

Enforcing KKT optimality conditions,

$$\begin{aligned}\frac{\partial D_E^{(k)}}{\partial q_{s,m}^{(k)}} &= f_{X,S}(q_{s,m}^{(k)}, s)\eta_E^{(k)}(q_{s,m}^{(k)}, s, y_m^{(k)}) \\ &\quad - f_{X,S}(q_{s,m}^{(k)}, s)\eta_E^{(k)}(q_{s,m}^{(k)}, s, y_{m+1}^{(k)}),\end{aligned}$$

we obtain $q_{s,m}^{(k)}$ as the solution to

$$\eta_E^{(k)}(q_{s,m}^{(k)}, s, y_m^{(k)}) = \eta_E^{(k)}(q_{s,m}^{(k)}, s, y_{m+1}^{(k)}).$$

REFERENCES

Akyol, E., Langbort, C., and Başar, T. (2015). Strategic compression and transmission of information. In *Proceedings of the IEEE Information Theory Workshop, Korea*, volume 2015.

Akyol, E. and Anand, A. (2023). Strategic Quantization. In *IEEE International Symposium on Information Theory (ISIT)*, 543–548.

Akyol, E. and Anand, A. (2024). Strategic Quantization. *Working paper, available at arXiv.*

Anand, A. and Akyol, E. (2024). Strategic Quantization with Quadratic Distortion Measures. In *2024 60th Annual Allerton Conference on Communication, Control, and Computing (Allerton).*

Bénabou, R. and Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5), 1652–1678.

Camerer, C.F., Ho, T.H., and Chong, J.K. (2004). A Cognitive Hierarchy Model of Games. *The Quarterly Journal of Economics*, 119(3), 861–898.

Chong, J.K., Camerer, C.F., and Ho, T.H. (2005). Cognitive Hierarchy: A Limited Thinking Theory in Games. In *Experimental Business Research: Marketing, Accounting and Cognitive Perspectives Volume III*, 203–228. Springer.

Crawford, V.P. and Sobel, J. (1982). Strategic Information Transmission. *Econometrica: Journal of the Econometric Society*, 1431–1451.

Fischer, P.E. and Verrecchia, R.E. (2000). Reporting Bias. *The Accounting Review*, 75(2), 229–245.

Kamenica, E. and Gentzkow, M. (2011). Bayesian Persuasion. *American Economic Review*, 101(6), 2590–2615.

Rayo, L. and Segal, I. (2010). Optimal Information Disclosure. *Journal of Political Economy*, 118(5), 949–987.