

Pure Nash Equilibria under the Affine Mechanism: A Potential Game of Exaggeration

Jason Jisen Li, Young Wu, Yancheng Zhu, Jin-yi Cai, and Xiaojin Zhu

University of Wisconsin - Madison

Abstract. The mean mechanism is known to be non-incentive-compatible, namely, rational players are incentivized to misreport their values. Despite this game-theoretic issue, the mean mechanism is prevalent in practice due to its other desirable properties. We give a full characterization of pure Nash equilibria—how the players will misreport—for the affine mechanism, of which the mean is a special case. Furthermore, we characterize both complete-information and Bayesian games under the affine mechanism. Our results highlight the inevitability of extreme exaggeration in such games.

Keywords: Game Theory

1 Introduction

Consider the classic numerical aggregation problem: n players each report a number $x_1 \dots x_n$. A mechanism aggregates these into a single number $\hat{x}$, e.g., using the mean or the median function. Player i 's utility depends on how close x_i and $\hat{x}$ are (assuming the reported x_i is truthful). There has been a sharp divide in theory and practice:

- In theory, it is well understood that on single-peaked preference domains, the strategyproof rules are exactly the generalized median (phantom-voter) aggregation mechanisms ([Moulin, 1980]). In contrast, the mean mechanism is the textbook example of non-incentive-compatibility: players will have an incentive to misreport their numbers (Procaccia and Tennenholtz [2013], Chan et al. [2021]).
- In practice, the mean remains the dominant aggregation rule when participants submit numerical reports. Corporate and academic performance evaluations such as paper reviews or student evaluations of teaching are almost universally aggregated with the mean. In congressional committee legislative voting over continuous variables (e.g., deciding the specific dollar amount for a corporate tax rate), final compromises are routinely arrived at via split-the-difference arithmetic averages (Martin [2001]). Many expert judgment procedures, such as engineering risk assessment or technology forecasting, aggregate experts by equal-weight averaging (Cooke [1991]).

Using the mean is often not out of ignorance: it is a conscious choice to trade strategyproofness for better estimates under the L2 utilitarian social welfare

function (Gershkov et al. [2017]), for fairness when the policymaker believes extreme outliers deserve a voice even if it invites cheating (Maskin [2008]), and for interpretability, where explaining a median mechanism with phantom agents to the public is politically unpalatable (Walsh [2025]). When the stake is high, often practitioners patch up the mean mechanism via trimmed mean (as in skating judging) or weighted average, rather than completely abandoning the mean mechanism for the median. Therefore, despite being non-strategyproof, the mean mechanism is practiced ubiquitously.

While it was well-known that strategic players **will** misreport under the mean mechanism, it was less clear **how** they misreport. A better understanding is important for detecting strategic behaviors, and for making an informed decision when choosing a mechanism to balance welfare efficiency and control cheating. When players’ strategic misreporting is viewed as an attempt to “game the system”, it also has security implications. This paper gives a full characterization of how the players rationally misreport under affine mechanisms, including the mean mechanism. Our main contributions are:

1. We prove that pure Nash Equilibria (pure NEs) always exist under the mean mechanism. This is done via identifying its potential function.
2. In each pure NE, all but at most one player must exaggerate their misreporting to the maximum extent possible.
3. In decomposable high-dimensional action spaces, the pure NEs are also decomposable into individual 1D problems and admit closed-form solutions.
4. When players do not know other players’ types, we provide a similar Bayesian NE characterization.

2 Related Work

Our work is closely related to the facility location problem. Moulin [1980] and Procaccia and Tennenholtz [2013] introduce the mechanism design problem for facility location, where the mechanism designer selects a facility location as a function of the players’ reported preferences in a way that rational players will truthfully report their preferences. In particular, they note that direct averaging of players’ reports will incentivize them to misreport. Berga [2002] extends the problem to a multidimensional domain similar to our setting, but it focuses on an incentive-compatible mechanism called the generalized median voter scheme. Our paper is not to advocate for the mean mechanism. But given its practical relevance, we focus on characterizing its pure NEs. Schummer and Vohra [2002], Filimonov and Meir [2023] study the facility location problem on graphs and trees; Goel and Hann-Caruthers [2020], Zampetakis and Zhang [2023], Barak et al. [2024] study the approximation mechanisms in multidimensional Euclidean spaces, which use similar techniques and are related to our reduction of the multidimensional problem into multiple single-dimensional ones, but they also focus on decomposition of the generalized median type mechanisms instead of the mean mechanism.

An additional justification for the mean mechanism stems from a decision-theoretic perspective: the optimality of the Bayes estimator for the posterior mean. Under quadratic loss and Gaussian distributional assumptions, weighted averaging is not only commonly used but also optimal. This is empirically studied and commonly assumed in the social science literature, including models by DeGroot and Mortera [1991], Romeijn [2024], which argue that averaging, or what they call linear opinion pooling, is a form of optimal Bayesian updating based on other people’s opinions, assuming truthfulness. They interpret the weights in the averaging process as trust coefficients, which our model also uses. In the economics and forecasting literature, simple averaging often performs better than more complex methods across various economic forecasting tasks, as summarized in Clemen [1989], Wang et al. [2023].

One interesting line of research observed human exaggeration behaviors in the context of political polarization and media bias, including Jost et al. [2022], Kubin and Von Sikorski [2021], Van Bavel et al. [2021]. Although these studies do not directly link the behavior to a simple mean mechanism, they provide empirical evidence of exaggeration as an equilibrium behavior, and the non-incentive-compatibility of humans’ opinion aggregation mechanisms could be an important contributing factor.

3 Problem Setup

We consider the affine mechanism in an n -player simultaneous-move game. The players have a common continuous action space $\mathcal{X} \subset \mathbb{R}^d$ which is compact and convex. Let $\mathbf{x}_i \in \mathcal{X}$ be the action of the i th player for $i \in [n] := \{1, 2, \dots, n\}$. A joint action, or pure strategy profile, $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ denotes the simultaneous action choice of all players. As in standard literature, we also write $\mathbf{x} = (\mathbf{x}_i, \mathbf{x}_{-i})$ when we want to emphasize player i .

The mean mechanism aggregates to various degrees the inputs it receives from all players. In this paper, we consider a slight generalization of the mean, namely the affine mechanism of the form

$$\hat{\mathbf{x}} := w_0 \mathbf{x}_0 + \sum_{i=1}^n w_i \mathbf{x}_i, \quad (1)$$

where $w_i > 0, i \in [n]$ is a real-valued (not necessarily normalized) weight that signifies how much influence player i has on the mechanism. $\mathbf{x}_0 \in \mathcal{X}$ is a bias term that, together with $w_0 \in \mathbb{R}$, denotes a fixed, constant “background” influence that is beyond the control of the n players. The affine mechanism (1) is common knowledge to all players.

The n players each have a loss (negative utility) function ℓ_i defined by a target $\mathbf{t}_i \in \mathbb{R}^d$. Note the targets are not restricted to $\mathcal{X}$. The goal of player i is to drive the mechanism’s $\hat{\mathbf{x}}$ close to $\mathbf{t}_i$. We consider the squared 2-norm loss:

$$\ell_i(\mathbf{x}) := \|\hat{\mathbf{x}} - \mathbf{t}_i\|_2^2 = \|w_0 \mathbf{x}_0 + \sum_{j=1}^n w_j \mathbf{x}_j - \mathbf{t}_i\|_2^2. \quad (2)$$

As rational agents, the players want to selfishly minimize their own $\ell_i(\mathbf{x})$. The players compete with each other when their targets $\mathbf{t}_1, \dots, \mathbf{t}_n$ are distinct (though our setup allows some or all targets to overlap, too). The fact that the mechanism's $\hat{\mathbf{x}}$ is defined by the joint action $\mathbf{x}$ couples the players together in a general-sum game.

The above narrative can be abstracted into the following formal affine mechanism game,

Definition 1 (Affine Mechanism Game (AMG)). The Affine Mechanism Game is an n -player general-sum game $G = (n, \mathcal{X}, \{\ell_i\}_{i=1}^n)$, where n is the number of players, $\mathcal{X} \subset \mathbb{R}^d$ is a compact and convex action space, and $\ell_i : \mathcal{X}^n \mapsto \mathbb{R}$ taking the form of equation (2) is player i 's loss function.

In the rest of this section, we assume that the targets (and hence loss functions) are common knowledge to all players. We show G is a potential game and study its pure strategy Nash equilibria. We then characterize the properties of these pure NEs. In section 4 we relax this assumption and study the resulting Bayesian game.

3.1 AMG is a Potential Game

Definition 2. A pure strategy Nash equilibrium of the game $G = (n, \mathcal{X}, \{\ell_i\}_{i=1}^n)$ is a strategy profile $\mathbf{x} \in \mathcal{X}^n$ satisfying $\ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) \leq \ell_i(\mathbf{y}, \mathbf{x}_{-i}), \forall \mathbf{y} \in \mathcal{X}, i \in [n]$.

Our AMG satisfies the Debreu-Glicksberg-Fan theorem and pure NEs exist. To find the pure NEs explicitly, we turn to potential games. Potential games are games with a special structure that allow strong results on pure Nash equilibria (Monderer and Shapley [1996]). We now show AMG is a potential game:

Theorem 3 (Potential Game). *The AMG G is a potential game with the potential function*

$$\phi(\mathbf{x}_1, \dots, \mathbf{x}_n) := \left\| \sum_{i=0}^n w_i \mathbf{x}_i \right\|_2^2 - 2 \sum_{i=1}^n w_i \mathbf{t}_i^\top \mathbf{x}_i. \quad (3)$$

As a potential game, AMG has pure NEs. Section 3.2 will be dedicated to the structure of these pure NEs. First, we remark on algorithms for finding pure NEs of AMG. Due to Proposition 5, any convex optimization algorithm that minimizes the convex function ϕ over the convex set $\mathcal{X}^n$ with strong guarantees can be utilized to find a pure NE (Boyd and Vandenberghe [2004]). Meanwhile, in the game theory community, the best-response dynamics is a traditional algorithm for finding a pure NE in potential games (Roughgarden [2010]):

Remark 4 (Best-Response Dynamics). Starting from an arbitrary $\mathbf{x}^{(0)} \in \mathcal{X}^n$, the best response sequence $\{\mathbf{x}^{(t)}\}_{t=1}^\infty$, under standard coordinate descent conditions (Tseng [2001]), every limit point of the sequence is a minimizer of ϕ , hence a pure Nash equilibrium of G , where $\mathbf{x}_i^{(t)} = \operatorname{argmin}_{\mathbf{x}_i} \ell_i(\mathbf{x}_i, \mathbf{x}_{-i}^{(t-1)})$ if $i = ((t-1) \bmod n) + 1$, and $\mathbf{x}_i^{(t-1)}$ otherwise.

One interesting observation that is relevant for AMG is that, under best response dynamics, no player needs to know other players' targets. Concretely, the players may carry out the best response dynamics as a learning dynamics in a distributed fashion over time, with no two players simultaneously updating their actions. When player i updates its own action $\mathbf{x}_i$, it observes other player's most recent actions $\mathbf{x}_{-i}$ but does not need to know their targets $\mathbf{t}_{-i}$. Therefore, the best response dynamics may offer a computational account of how players in the real world iteratively adjust their actions based on actions of other players without knowing the other players' true intentions, and still reach an equilibrium.

Note that minimizing the potential function ϕ along the $\mathbf{x}_i$ direction is equivalent to minimizing its own loss function ℓ_i : $\operatorname{argmin}_{\mathbf{x}_i} \phi(\mathbf{x}_i, \mathbf{x}_{-i}) = \operatorname{argmin}_{\mathbf{x}_i} \ell_i(\mathbf{x}_i, \mathbf{x}_{-i})$. As a result, best response dynamics also correspond to cyclic coordinate descent on ϕ (Wright and Recht [2022]).

3.2 Structure of Pure Nash Equilibria in AMG

We next show that $\mathbf{x} \in \mathcal{X}^n$ is a pure NE of G if and only if $\mathbf{x}$ is a minimum of ϕ restricted to the domain $\mathcal{X}^n$.

Proposition 5 (Pure NEs $\iff$ minima). *The set of pure Nash equilibria in G is $\operatorname{pureNE}(G) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}^n} \phi(\mathbf{x})$, and moreover, the set of correlated equilibria is the set of mixtures of pure Nash equilibria, that is, $\operatorname{CE}(G) = \Delta \operatorname{pureNE}(G)$.*

We remark that by definition $\mathcal{X}$ is compact and convex, thus $\mathcal{X}^n$ is bounded and closed. The potential function $\phi(\mathbf{x})$ may not have a global minimum on the extended domain $\mathbb{R}^{nd}$ (it could diverge to $-\infty$ there), but on $\mathcal{X}^n$ it will have at least one minimum (perhaps on the boundary). In fact, we have the following guarantee.

Corollary 6 (Cardinality of $\operatorname{pureNE}(G)$). *$\operatorname{pureNE}(G)$ is non-empty and convex, that is, G has either one pure NE or infinite pure NEs.*

In particular, the second part of Proposition 5 and this Corollary imply that in the case of a unique pure strategy Nash equilibrium $\mathbf{x}$ (for example, in the 1D two-player case with distinct targets, which we expand on in the next subsection), then $\mathbf{x}$ is the unique Nash equilibrium and the unique correlated equilibrium.

We provide a few illustrative examples of AMG in subsection 3.3. In AMG the pure NEs often seem to involve *all* players misrepresenting their true target to the mechanism (i.e. $\mathbf{x}_i \neq \mathbf{t}_i$). Furthermore, such misrepresentation often seems to take the form of *extreme exaggeration*, in the sense that a player's rational action $\mathbf{x}_i$ at any pure NE is often pushed to the boundary of the action space $\mathcal{X}$ so they cannot exaggerate the action further. These observations are almost true, but need a subtle correction: Given that the players' targets $\mathbf{t}_1 \dots \mathbf{t}_n$ are all distinct, **at any pure NE, extreme exaggeration is necessary for all (except perhaps one) players**. In other words, there are two possibilities: Either all players are at the boundary of $\mathcal{X}$, or one of them is in the interior of $\mathcal{X}$. In the latter case, that special player (say player i^*) will in general still need

to exaggerate its target $\mathbf{x}_{i^*} \neq \mathbf{t}_{i^*}$ to the mechanism; it is just that $\mathbf{x}_{i^*}$ is in the interior of $\mathcal{X}$ and not at the boundary. Furthermore, player i^* is the lucky winner in that the mechanism will end up at its target $\mathbf{t}_{i^*}$. Our next theorem precisely characterizes this phenomenon.

Theorem 7 (All-But-At-Most-One Extreme Exaggeration). *If $\{\mathbf{t}_i\}_{i=1}^n$ are all distinct, then every pure NE satisfies the property that $|\{i \in [n] : \mathbf{x}_i \in \text{int } \mathcal{X}\}| \leq 1$. Conversely, in a NE, if $\mathbf{x}_i, \mathbf{x}_j \in \text{int } \mathcal{X}$, then $\mathbf{t}_i = \mathbf{t}_j$. Furthermore, if $\mathbf{x}_{i^*} \in \text{int } \mathcal{X}$ for some $i^* \in [n]$, then $\hat{\mathbf{x}} = \mathbf{t}_{i^*}$.*

3.3 Computing Pure Nash Equilibria

We now show how to find pure NEs in high dimensional action space by decomposing the problem. The base case is when $\mathcal{X}$ is in a one-dimensional space, under the assumption that it is compact and convex, we can write $\mathcal{X} = [L, U]$, and in this case, we can fully characterize the set of all Nash equilibria as follows.

Lemma 8 (Nash Equilibria in 1D). *When $\mathcal{X} = [L, U]$, then there exists some $t^* \in \mathbb{R}$, such that in every Nash equilibrium $(x_1, x_2, \dots, x_n)$ of AMG, we have*

$$x : \begin{cases} x_i = L & \text{if } t_i < t^* \\ x_i : \sum_{i:t_i=t^*} w_i x_i = t^* - w_0 x_0 - \sum_{i:t_i < t^*} w_i L - \sum_{i:t_i > t^*} w_i U & \text{if } t_i = t^* \\ x_i = U & \text{if } t_i > t^* \end{cases} \quad (4)$$

In particular, when targets are sorted $t_1 \leq t_2 \leq \dots \leq t_n$, (4) is equivalent to the existence of $\underline{i}$ and $\bar{i}$ such that in every NE $x_1 = x_2 = \dots = x_{\underline{i}} = L$, $x_{\bar{i}} = x_{\bar{i}+1} = \dots = x_n = U$, and if there are players with the target $t_i = t^*$ with $\underline{i} < i < \bar{i}$, then it is NE as long as their x_i results in $\hat{x} = t^*$.

Remark 9. A simple algorithm to find an NE is to iterate over the following set, $t^* \in \{t_1, t_2, \dots, t_n\} \cup \left\{ \frac{1}{2}(t_1 + t_2), \frac{1}{2}(t_2 + t_3), \dots, \frac{1}{2}(t_{n-1} + t_n) \right\} \cup \{L, U\}$, and for each candidate t^* , compare the value of the potential function $\phi(x)$ with x satisfying (4), and the x that minimizes ϕ is a NE.

In general, it is difficult to find a closed form solution in higher dimensions, but for special $\mathcal{X}$ that is a Cartesian product of intervals (i.e. a hyper-rectangle),

that is $\mathcal{X} = \prod_{k=1}^d \mathcal{X}_k$ with $\mathcal{X}_k = [L_k, U_k]$, we provide the following decomposition result, which implies an algorithm to solve d one-dimensional problems using Lemma 8 to obtain the NE coordinate-wise.

Since the loss function is coordinate-wise-separable, $\ell_i(\mathbf{x}) = \sum_{k=1}^d \ell_{ik}(\mathbf{x}_k)$, where $\ell_{ik}(\mathbf{x}_k) := (\hat{\mathbf{x}}_k - \mathbf{t}_{ik})^2$. We can define for each $k \in [d]$, G_k with the action space

$\mathcal{X}_k = [L_k, U_k]$, loss ℓ_{ik} , and mechanism $\hat{\mathbf{x}}_k = w_0 \mathbf{x}_{0k} + \sum_{i=1}^n w_i \mathbf{x}_{ik}$, for any player $i \in [n]$. Then, the pure Nash Equilibrium of the multi-dimensional game is simply a cross-product of all d pureNE(G_k) sets. More formally, a conversion between the Nash equilibrium in G and the 1D games G_k can be established through the following proposition:

Proposition 10 (Decomposition of Multi-dimensional AMG). *Assume*

$\mathcal{X} = \prod_{k=1}^d \mathcal{X}_k \subseteq \mathbb{R}^d$, *then the set of pure Nash equilibria of G can be characterized*

by $\text{pureNE}(G) = \prod_{k=1}^d \text{pureNE}(G_k)$, *where* $G_k = (n, \mathcal{X}_k, \ell_{ik})$.

Given this proposition, it is sufficient to form and solve the d one-dimensional games G_k independently, and take the Cartesian product of their pure Nash equilibrium sets to obtain $\text{pureNE}(G)$ for multi-dimensional AMG G .

Example 11 (NEs for Two-Player AMG in 1D and 2D). In a one-dimensional two-player AMG G with targets t_1 and t_2 , pure actions $x_1, x_2 \in \mathcal{X} = [L, U]$, and mechanism $\hat{x} = w_0 x_0 + w_1 x_1 + w_2 x_2$, the pure NE set only depends on the relative positions of t_1 and t_2 with respect to the weighted middle point m , where $m = w_0 x_0 + w_1 L + w_2 U$ if $t_1 \leq t_2$, and $m = w_0 x_0 + w_1 U + w_2 L$ if $t_1 > t_2$. The closed form solution of pure NE can be divided into three cases:

1. If $t_1 = t_2 = t$: $\text{pureNE}(G) = \begin{cases} \{(L, L)\}, & \text{if } t < w_0 x_0 + L(w_1 + w_2), \\ \{(U, U)\}, & \text{if } t > w_0 x_0 + U(w_1 + w_2), \\ \{(x_1, x_2) : \hat{x} = t\}, & \text{otherwise.} \end{cases}$
2. If $t_1 \neq t_2$ and $t_1 < m < t_2$ or $t_2 < m < t_1$: $\text{pureNE}(G) = \begin{cases} \{(L, U)\}, & \text{if } t_1 < t_2, \\ \{(U, L)\}, & \text{if } t_2 < t_1. \end{cases}$
3. If $t_1 \neq t_2$ and $t_1, t_2 \leq m$ or $t_1, t_2 \geq m$:

$$\text{pureNE}(G) = \begin{cases} \{(L, \max(L, \frac{1}{w_2}(t_2 - w_0 x_0 - w_1 L)))\}, & \text{if } t_1 < t_2 \leq m, \\ \{(\max(L, \frac{1}{w_1}(t_1 - w_0 x_0 - w_2 L)), L)\}, & \text{if } t_2 < t_1 \leq m, \\ \{(U, \min(U, \frac{1}{w_2}(t_2 - w_0 x_0 - w_1 U)))\}, & \text{if } m \leq t_2 < t_1, \\ \{(\min(U, \frac{1}{w_1}(t_1 - w_0 x_0 - w_2 U)), U)\}, & \text{if } m \leq t_1 < t_2. \end{cases}$$

In case 1, the two targets coincide. NE happens when $\hat{x}$ is in the location closest to t . This is also the case when there can be infinite many pure NEs (the third branch). In case 2, the targets lie on different sides of m . Both players must exaggerate their actions to the maximum extent possible. In case 3, both targets lie on the same side of m . One player will exaggerate to the maximum extent, while the other either “wins the game” by forcing the mechanism to land on its own target or also exaggerates to the maximum when winning is impossible due to a large $w_0 x_0$.

We illustrate in Figure 1 the three NE cases for 1-dimensional two player AMG examples with domain $\mathcal{X} = [L, U]$, equal weights $w_1 = w_2 = \frac{1}{2}$, and bias term $w_0 x_0 = 0$. In case 1, x_1 and x_2 can be anywhere in the gray box, as long as $\hat{x}$ sits on $t_1 = t_2$. Thus, there are infinite pure NE. Case 2 and 3 both have a unique pure NE. In case 2 x_1 and x_2 both exaggerate to the extreme action, L or U , that is closest to their respective targets. In case 3, player 1 plays an interior action and “wins” as $\hat{x}$ lands on t_1 , while player 2 chooses an extreme action.

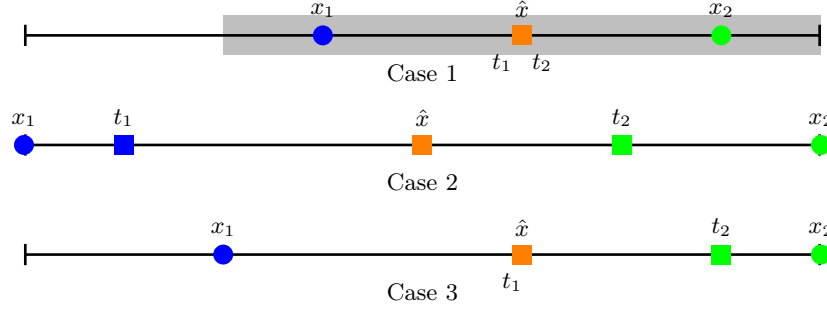


Fig. 1: Pure NE of two-player 1D AMG with Equal Weights

Now, for multi-dimensional games we can combine the 1D pure NEs. In particular, for a 2-dimensional game, the pure NE falls into six categories formed by combining the three 1-D characterizations, which yields $3 \times 3 = 9$ cases, then eliminating the three mirroring cases. In Figure 2 we present these six characterizations of a 2-dimensional AMG with $\mathcal{X} = [L, U]^2$, uniform weights $w_1 = w_2 = \frac{1}{2}$, and $w_0 x_0 = 0$. These derive directly from combining the three 1-dimensional NE cases in Figure 1. In case 1 ($\mathbf{t}_1 = \mathbf{t}_2$), $\mathbf{x}_1, \mathbf{x}_2$ can be anywhere in the gray box so long as they average to $\mathbf{t}_1$. In cases 2 and 3, $\mathbf{t}_1, \mathbf{t}_2$ agree on one coordinate, and the infinite many pure NEs are chosen from the blue line and the green line, respectively. Cases 4,5,6 have a unique pure NE. Case 6 is where player 1 is the “winner” and has an interior action.

4 Bayesian Games under the Affine Mechanism

In the previous formulation of AMG, we assumed that each player knows the target location $\mathbf{t}_1, \dots, \mathbf{t}_n$ of all players, and hence all loss functions are common knowledge. However, it is also interesting to consider the scenario in which each player (say player i) only knows its own target $\mathbf{t}_i$, but not any of the other players’ target $\mathbf{t}_{-i}$. For such games with incomplete information, we adopt an interim Bayesian game approach in which the player knows only the prior distributions of the other targets.

Concretely, the Bayesian Affine Mechanism Game has the same formulation as the standard AMG, except that any target $\mathbf{t}_i$ shall be chosen from a distribution

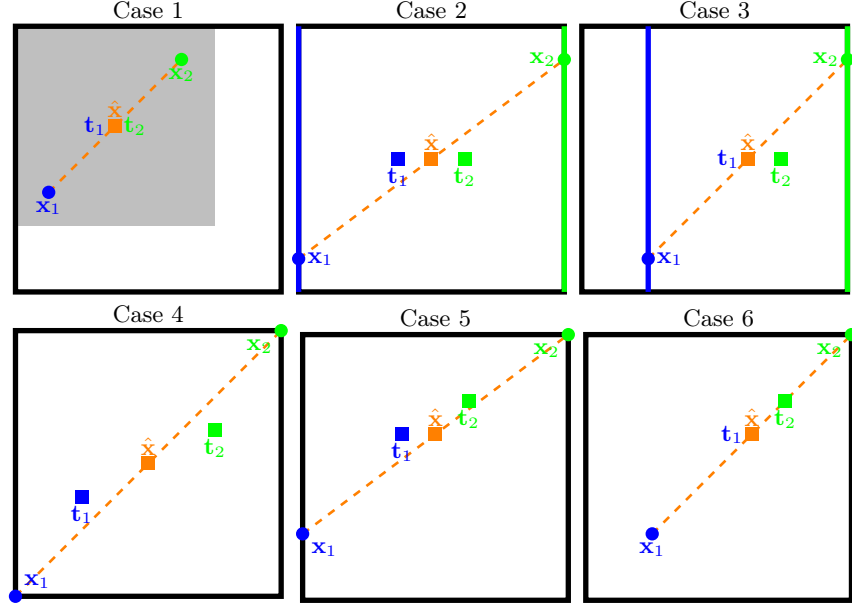


Fig. 2: Pure NE of two-player 2D AMG with Equal Weights

T_i , for all $i \in [n]$. Each player i knows the exact location of their own target $\mathbf{t}_i$. They also know the distributions of T_j , $\forall j \in [n]$, the AMG mechanism, and that other players also know this information. The players then make the simultaneous action $\mathbf{x}_i \in \mathcal{X}$ from the common continuous action space $\mathcal{X} \subset \mathbb{R}^d$ that is compact and convex. All other settings remain the same. We have the following definition:

$$\ell_i(\mathbf{x}_i, \mathbf{x}_{-i} | \mathbf{t}_i) := \|\hat{\mathbf{x}} - \mathbf{t}_i\|_2^2 = \left\| w_0 \mathbf{x}_0 + \sum_{j=1}^n w_j \mathbf{x}_j - \mathbf{t}_i \right\|_2^2 \quad (5)$$

Definition 12 (Bayesian AMG). The Bayesian Affine Mechanism Game is an n -player general-sum Bayesian game $G = (n, \mathcal{X}, \{T_i\}_{i=1}^n, \{\ell_i\}_{i=1}^n)$, where n is the number of players, $\mathcal{X}$ is the compact and convex action space, T_i is the independent prior type distribution of player i 's target $\mathbf{t}_i \in \mathbb{R}^d$ and $\ell_i : \mathcal{X}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$ is given by (5).

We denote a player's strategy by $\mathbf{x}_i = \pi_i(\mathbf{t}_i)$ where $\pi_i : \mathbb{R}^d \rightarrow \mathcal{X}$ maps their target position to an action $\mathbf{x}_i \in \mathcal{X}$. We again approach the problem with the assumption of a product space domain $\mathcal{X} = \prod_{k=1}^d [L_k, U_k]$, and the following result characterizes the best responses in the Bayesian game.

Proposition 13 (Bayesian Best Responses). *The best response function for each player i in the Bayesian AMG game with $\mathcal{X} = \prod_{k=1}^d [L_k, U_k]$ has the ramp*

function form,

$$br_i(\pi_{-i}|t_i) = \mathbf{x}_i, \text{ where for every } k \in [d]$$

$$x_{ik} = \min \left\{ U_k, \max \left\{ L_k, \frac{1}{w_i} t_{ik} - \frac{1}{w_i} \left(w_0 x_{0k} + \sum_{j \neq i} w_j \mathbb{E}_{T_j} [\pi_{jk}(t_j)] \right) \right\} \right\}. \quad (6)$$

In particular, (6) implies that in all Bayesian Nash equilibria, all players use a strategy with this ramp function form in all coordinates.

To be specific, in a BNE, the ramp function can be rewritten as, for every

$$\text{player } i, \pi_i(t_i) = \mathbf{x}_i, \text{ where } x_{ik} = \begin{cases} L_k & \text{if } t_{ik} \leq a_{ik} \\ \frac{1}{w_i} t_{ik} + c_{ik} & \text{if } a_{ik} < t_{ik} < b_{ik} \quad \forall k \in [d], \\ U_k & \text{if } t_{ik} \geq b_{ik} \end{cases}$$

where $c_{ik} = -\frac{1}{w_i} \left(w_0 x_{0k} + \sum_{j \neq i} w_j \mathbb{E}_{T_j} [\pi_{jk}(t_j)] \right)$, $a_{ik} = w_i (L_k - c_{ik})$, and $b_{ik} = a_{ik} + w_i (U_k - L_k) = w_i (U_k - c_{ik})$. This set of equations in each dimension k characterizes $\{c_{ik}\}_{i=1}^n$ thus π_i . Notice that the equation set in each dimension can be calculated independently.

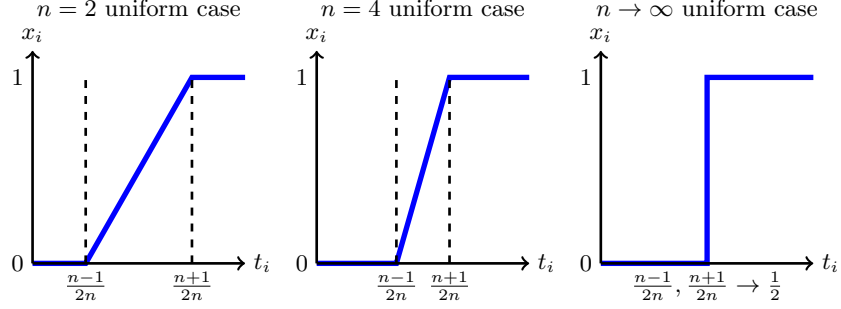
This result can be interpreted as follows: **in any dimension, the exaggeration factor is always $\frac{1}{w_i}$, that is, the less influence the player has, the more the player will exaggerate.**

For arbitrary prior type distributions, $\{c_{ik}\}_{i=1}^n \quad \forall k \in [d]$ can be solved numerically, but as an example, we can solve for a closed-form solution under the assumption of uniformly distributed targets for all players.

Example 14 (BNE for Uniformly Distributed Targets). A one-dimensional Bayesian AMG with $\mathcal{X} = [0, 1]$, targets $T_i \sim \text{Unif}[0, 1]$, bias term $w_0 x_0 = 0$, and $\sum_{i=1}^n w_i = 1$ for all players has the unique BNE strategy profile $(\pi_1, \dots, \pi_n)$ given by

$$\pi_i^*(t_i) = \begin{cases} 0, & t_i \leq \frac{1-w_i}{2}, \\ \frac{1}{w_i} t_i - \frac{1-w_i}{2w_i}, & \frac{1-w_i}{2} < t_i < \frac{1+w_i}{2}, \\ 1, & t_i \geq \frac{1+w_i}{2}. \end{cases} \quad (7)$$

(7) follows from Proposition 13 by solving the system of n equations in the form of (6). More details of the derivation can be found in the appendix. With $w_0 = 0$ and $w_i = \frac{1}{n}$ for every i , (7) can be further simplified and plotted for various values of n in Figure 3. In this example, $1 - (b_i - a_i) = 1 - w_i$ fraction of the types of player i will report one of two extreme types 0 or 1, and the remaining w_i fraction will exaggerate by a factor of $\frac{1}{w_i}$.

Fig. 3: BNE of Games with n Players and Uniform Type Distributions

5 Extensions of the Affine Mechanism Game

5.1 Extension 1: Negative Inner Product Loss

Up to now player i 's loss function (2) is based on the Euclidean distance between its target point $\mathbf{t}_i \in \mathbb{R}^d$ and the mechanism $\hat{\mathbf{x}}$. In some applications, the following *negative inner product loss* can be more appropriate:

$$\ell_i(\mathbf{x}) := -\mathbf{t}_i^\top \hat{\mathbf{x}}. \quad (8)$$

That is, player i has a small loss if the mechanism $\hat{\mathbf{x}}$ has a large projection onto the direction of target direction $\mathbf{t}_i$.

AMG with this negative inner product loss (8) has an even stronger guarantee: the game has a *Weakly Dominant Strategy Equilibrium*.

Definition 15 (Weakly Dominant Strategy Equilibrium (wDSE)). An action profile $\mathbf{x} = (\mathbf{x}_1 \dots \mathbf{x}_n) \in \mathcal{X}^n$ is a wDSE if for every player i , $\ell_i(\mathbf{x}_i, \mathbf{x}'_{-i}) \leq \ell_i(\mathbf{y}, \mathbf{x}'_{-i})$, $\forall \mathbf{y} \in \mathcal{X}, \mathbf{x}'_{-i} \in \mathcal{X}^{n-1}$.

Remark 16. The term weakly dominant strategy equilibrium is used in Chapter 4.5 of Tadelis [2013], and it is also called dominant strategy equilibrium in Chapter 10.3 of Osborne [1994], and dominant strategy solution in Chapter 1.3.1 of Roughgarden [2010]. As noted in Osborne [1994], an action in a wDSE is not required to weakly dominate (with at least one strict inequality) all other actions for a player. wDSE is also a weaker solution concept compared to (strictly) dominant strategy equilibrium, which requires strict inequality everywhere, $\ell_i(\mathbf{x}_i, \mathbf{x}'_{-i}) < \ell_i(\mathbf{y}, \mathbf{x}'_{-i})$, $\forall \mathbf{y} \in \mathcal{X}, \mathbf{y} \neq \mathbf{x}_i, \mathbf{x}'_{-i} \in \mathcal{X}^{n-1}$.

Proposition 17 (Existence of wDSE). Under loss (8), game G has a wDSE $\mathbf{x}_i^*$ satisfying $\mathbf{x}_i^* \in \operatorname{argmax}_{\mathbf{x}_i \in \mathcal{X}} \mathbf{t}_i^\top \mathbf{x}_i$, $\forall i \in [n]$.

One significant benefit of this wDSE is that each player can compute their $\mathbf{x}_i^*$ on their own, without even the knowledge of other players' actions $\mathbf{x}_j, \forall j \neq i$. Contrast this with the best-response dynamics in section 3.1. We remark that

$\mathbf{x}_i^*$ may not be exactly along the direction of $\mathbf{t}_i$ since it depends on the domain $\mathcal{X}$. Nonetheless, computing $\mathbf{x}_i^*$ is a convex optimization problem—maximizing a linear function over a convex set—and thus efficient.

5.2 Extension 2: Finite Action Space

So far, we have assumed that the players’ action space $\mathcal{X}$ is a compact and convex (hence infinite unless singleton) subset of $\mathbb{R}^d$. In some applications, the players are restricted to picking their actions from a finite $\mathcal{X}$ instead. For example, $\mathcal{X}$ may be the collection of news articles (represented by embedding vectors in $\mathbb{R}^d$) published by all professional news agencies within the past 24 hours, and each player may select a handful of such news articles to place on a social media user (the mechanism)’s timeline. This motivates the extension to finite action space:

Definition 18 (AMG with finite action space). Affine Mechanism Game with finite action space is an n -player general-sum game $F = (n, \{\mathcal{P}_{k_i}(\mathcal{X})\}_{i=1}^n, \{\ell_i\}_{i=1}^n)$, where the action space of player i is $\mathcal{P}_{k_i}(\mathcal{X})$, and $\mathcal{P}_{k_i}(\mathcal{X})$ is the set of all subsets containing k_i elements from a finite $\mathcal{X} \subset \mathbb{R}^d$. The loss function of player i is given by $\ell_i(\mathbf{x}) = \|\hat{\mathbf{x}} - \mathbf{t}_i\|_2^2$ with $\hat{\mathbf{x}} = w_0 \mathbf{x}_0 + \sum_{i=1}^n w_i \sum_{j=1}^{k_i} \mathbf{x}_i^{(j)}$, where $\mathbf{x}_i^{(j)}$ is the j th element in player i ’s chosen subset.

In the original AMG (Definition 1) where $\mathcal{X}$ is convex, allowing the players to choose multiple items would not affect the results since choosing multiple items is equivalent to choosing the mean of these items, which is still in $\mathcal{X}$. In the new game, the average of items in $\mathcal{X}$ may not be in $\mathcal{X}$. However, the new game can be interpreted as each player picks one “meta item” $\mathbf{x}_i'$ instead of k_i items, with

$$w_i' = w_i k_i \text{ and } \mathbf{x}_i' = \frac{1}{k_i} \sum_{j=1}^{k_i} \mathbf{x}_i^{(j)}.$$

Corollary 19. F is also a potential game, with a potential function

$$\phi \left(\left\{ \left\{ \mathbf{x}_i^{(j)} \right\}_{j=1}^{k_i} \right\}_{i=1}^n \right) := \left\| w_0 \mathbf{x}_0 + \sum_{i=1}^n w_i' \mathbf{x}_i' \right\|_2^2 - 2 \sum_{i=1}^n w_i' \mathbf{t}_i^\top \mathbf{x}_i'. \quad (9)$$

Therefore, the new game F also has at least one pure NE. Since $\mathcal{X}$ is finite and the number of players is finite, F is a finite game. Although a pure NE can be similarly found through best-response dynamics, unlike the continuous case, each iteration of best-response dynamics can be costly to compute for large values of k_i since it involves solving a variant of the subset sum problem. Since F is finite, it clearly cannot have infinite pure NEs (contrast with corollary 6), but it can still have many pure NEs.

6 Conclusion and Future Work

We introduced the Affine Mechanism Game and its Bayesian variant, and characterized the structures of pure Nash and Bayesian Nash equilibria, including closed-form solutions under simplifying assumptions. In particular, we showed the players' extreme-exaggeration equilibrium behavior in the complete-information case and quantified the exaggeration ratio as a function of the players' influence (weight).

The presented work has several limitations that can be enhanced by future work, including broadening the mechanism to non-affine functions (such as trimmed mean) and characterizing the full set of NEs, and allowing uncertainty in the players' beliefs about the mechanism (such as the weights w_i 's).

7 Technical Appendices and Supplementary Material

7.1 Proof of Theorem 3

Proof. We need to show if any player i deviates from action $\mathbf{x}_i$ to any action $\mathbf{y} \in \mathcal{X}$, we have $\ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) - \ell_i(\mathbf{y}, \mathbf{x}_{-i}) = \phi(\mathbf{x}_i, \mathbf{x}_{-i}) - \phi(\mathbf{y}, \mathbf{x}_{-i})$. To this end, define an auxiliary variable $\mathbf{z}$ that does not depend on $\mathbf{x}_i$ or $\mathbf{y}$:

$$\mathbf{z} := w_0 \mathbf{x}_0 + \sum_{j \neq i} w_j \mathbf{x}_j.$$

Then $\ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) = \|w_i \mathbf{x}_i + (\mathbf{z} - \mathbf{t}_i)\|^2 = \|w_i \mathbf{x}_i\|^2 + 2w_i \mathbf{x}_i^\top (\mathbf{z} - \mathbf{t}_i) + \|\mathbf{z} - \mathbf{t}_i\|^2$, and

$$\ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) - \ell_i(\mathbf{y}, \mathbf{x}_{-i}) = \|w_i \mathbf{x}_i\|^2 + 2w_i \mathbf{x}_i^\top (\mathbf{z} - \mathbf{t}_i) - \|w_i \mathbf{y}\|^2 - 2w_i \mathbf{y}^\top (\mathbf{z} - \mathbf{t}_i).$$

On the other hand,

$$\begin{aligned} \phi(\mathbf{x}_i, \mathbf{x}_{-i}) &= \|w_i \mathbf{x}_i + \mathbf{z}\|^2 - 2w_i \mathbf{t}_i^\top \mathbf{x}_i - 2 \sum_{j \neq i} w_j \mathbf{t}_j^\top \mathbf{x}_j \\ &= \|w_i \mathbf{x}_i\|^2 + 2w_i (\mathbf{z} - \mathbf{t}_i)^\top \mathbf{x}_i + \|\mathbf{z}\|^2 - 2 \sum_{j \neq i} w_j \mathbf{t}_j^\top \mathbf{x}_j. \end{aligned}$$

The last two terms do not depend on $\mathbf{x}_i$. Hence

$$\begin{aligned} &\phi(\mathbf{x}_i, \mathbf{x}_{-i}) - \phi(\mathbf{y}, \mathbf{x}_{-i}) \\ &= \|w_i \mathbf{x}_i\|^2 + 2w_i (\mathbf{z} - \mathbf{t}_i)^\top \mathbf{x}_i - \|w_i \mathbf{y}\|^2 - 2w_i (\mathbf{z} - \mathbf{t}_i)^\top \mathbf{y} \\ &= \ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) - \ell_i(\mathbf{y}, \mathbf{x}_{-i}). \end{aligned}$$

7.2 Proof of Proposition 5

Proof. Our potential function ϕ is the sum of two convex functions in $\mathbf{x}$ and hence convex. Using utilities $u_i = -\ell_i$, the game has concave potential $-\phi$, and since the domain $\mathcal{X}^n$ is convex and ϕ is smooth and convex, by Theorem 1 of Neyman [1997] and its corollary, the set of pure Nash equilibria coincides with the minima of the potential function on $\mathcal{X}^n$.

7.3 Proof of Corollary 6

Proof. Since the set of minima of the convex potential function on a compact domain $\mathcal{X}^n$ is non-empty, there is at least one pure Nash equilibrium. Since the set of minima of the convex potential function on a convex domain $\mathcal{X}^n$ is convex (Corollary in Neyman [1997] below Theorem 1), any convex combination of two distinct pure Nash equilibria is another pure Nash equilibrium.

7.4 Proof of Theorem 7

Proof. For any $i \in [n]$,

$$\nabla_{\mathbf{x}_i} \phi = 2w_i \sum_{k=0}^n w_k \mathbf{x}_k - 2w_i \mathbf{t}_i. \quad (10)$$

Suppose $\exists i, j \in [n], i \neq j : \mathbf{x}_i, \mathbf{x}_j \in \text{int } \mathcal{X}$. Then

$$\nabla_{\mathbf{x}_i} \phi = \mathbf{0} = \nabla_{\mathbf{x}_j} \phi \Rightarrow \mathbf{t}_i = \sum_{k=0}^n w_k \mathbf{x}_k = \mathbf{t}_j, \quad (11)$$

a contradiction. The equation $\hat{\mathbf{x}} = \mathbf{t}_{i^*}$ follows from $\nabla_{\mathbf{x}_{i^*}} \phi = \mathbf{0}$ and the definition of the mechanism (1).

7.5 Proof of Lemma 8

First we prove the following lemma:

Lemma 20 (Sorted Pure Nash Equilibrium). *For a joint action x , if there exist $i < j$ such that $x_j < x_i$ and $t_i < t_j$, then $x \notin \text{pureNE}(G)$.*

Proof. By the assumption in the lemma, at least one of the following must be true:

Case 1: $\hat{x} < t_j$. Consider the unilateral deviation of player j to $x_j + \epsilon$, for $\epsilon > 0$. Let $\hat{x}'$ be the new mechanism, where $\hat{x}' = \hat{x} + w_j \epsilon$. Choose ϵ small enough such that $\hat{x} < \hat{x}' < t_j$. This will always decrease the loss of player j .

Case 2: $\hat{x} > t_i$. Similarly, player i unilaterally deviate to $x_i - \epsilon$, for $\epsilon > 0$. The new mechanism is $\hat{x}' = \hat{x} - w_i \epsilon$. Choose ϵ small enough such that $\hat{x} > \hat{x}' > t_i$. This will always decrease the loss of player i .

In either case, some player has a unilateral deviation that strictly decreases their loss, so $x \notin \text{pureNE}(G)$.

Now we actually prove Lemma 8

Proof. Assume that the players are sorted in ascending order by their targets. By the Lemma 20, we know that in any pure NE, the players' actions must be sorted by their targets. This shows that $\exists \underline{i}, \bar{i} \in \{0, 1, \dots, n+1\}$ s.t. $x_i = L$ for $i \leq \underline{i}$, $x_i = U$ for $i \geq \bar{i}$, because pure NE strategies of the players are sorted by

their targets, and the action space is not unbounded. In other words, there could be a set of players playing L in the start of the queue of players sorted by their targets, and a potential set of players playing U in the end of that queue. Note that when $\underline{i} < 1$, we define that there will be no player playing L , and when $\bar{i} > n$, there will be no player playing U .

In addition, by Theorem 7, $\forall i$ s.t. $\underline{i} < i < \bar{i}$ must satisfy $t_i = t^*$ for some $t^* \in [L, U]$. Therefore, $x_i = L \ \forall i$ s.t. $t_i < t^*$, and $x_i = U \ \forall i$ s.t. $t_i > t^*$.

By Theorem 7, we have $\hat{x} = t^*$, so there is

$$\hat{x} = \sum_{i=0}^n w_i x_i = t^* \implies \sum_{i:t_i=t^*} w_i x_i = t^* - w_0 x_0 - \sum_{i:t_i < t^*} w_i L - \sum_{i:t_i > t^*} w_i U$$

7.6 Proof of Proposition 10

Proof. The loss for any player $i \in [n]$ in G can be separated this way:

$$\ell_i(\mathbf{x}) = \|w_0 \mathbf{x}_0 + \sum_{j=1}^n w_j \mathbf{x}_j - \mathbf{t}_i\|_2^2 = \sum_{k=1}^d (w_0 x_{0k} + \sum_{j=1}^n w_j x_{jk} - t_{ik})^2 = \sum_{k=1}^d \ell_{ik}(\mathbf{x}_k).$$

We argue that $\mathbf{x}$ is a pure Nash equilibrium if and only if, for each coordinate $k \in [d]$, the profile $(\mathbf{x}_{1k}, \mathbf{x}_{2k}, \dots, \mathbf{x}_{nk})$ is a pure Nash equilibrium of G_k .

Assuming that the action profile $\mathbf{x}$ is a pure NE. Then $\forall i \in [n]$, $\forall \mathbf{x}'_i \neq \mathbf{x}_i$, $\ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) \leq \ell_i(\mathbf{x}'_i, \mathbf{x}_{-i})$. Suppose that $\exists k \in [d]$, $i \in [n]$, $\mathbf{x}'_{ik} \neq \mathbf{x}_{ik}$ s.t. $\ell_{ik}(\mathbf{x}_{ik}, \mathbf{x}_{-ik}) > \ell_{ik}(\mathbf{x}'_{ik}, \mathbf{x}_{-ik})$. Define a full-dimensional deviation $\mathbf{x}'$ by changing only the k th coordinate:

$$x'_i = (x_{i1}, \dots, x_{i,k-1}, x'_{ik}, x_{i,k+1}, \dots, x_{id}).$$

then the pure NE assumption is violated:

$$\ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) = \sum_{k=1}^d \ell_{ik}(\mathbf{x}_{ik}, \mathbf{x}_{-ik}) > \ell_{i1}(\mathbf{x}_1) + \dots + \ell_{ik}(\mathbf{x}'_{ik}, \mathbf{x}_{-ik}) + \dots + \ell_{id}(\mathbf{x}_d) = \ell_i(\mathbf{x}'_i, \mathbf{x}_{-i}).$$

Hence, if $\mathbf{x}$ is a pure NE, $\mathbf{x}_k$ is a pure NE for G_k .

Conversely, assume that for all $k \in [d]$, $\mathbf{x}_k = (\mathbf{x}_{ik}, \mathbf{x}_{-ik})$ is a pure NE of G_k . Then $\forall i, k, \forall \mathbf{x}'_{ik} \in \mathcal{X}_k$, $\ell_{ik}(\mathbf{x}_{ik}, \mathbf{x}_{-ik}) \leq \ell_{ik}(\mathbf{x}'_{ik}, \mathbf{x}_{-ik})$. Summing the LHS and RHS of the inequalities gives:

$$\ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) = \sum_{k=1}^d \ell_{ik}(\mathbf{x}_{ik}, \mathbf{x}_{-ik}) \leq \sum_{k=1}^d \ell_{ik}(\mathbf{x}'_{ik}, \mathbf{x}_{-ik}) = \ell_i(\mathbf{x}'_i, \mathbf{x}_{-i}) \quad \forall \mathbf{x}'_i \neq \mathbf{x}_i.$$

This is exactly the definition of pure NE for G . Hence, if $\mathbf{x}_k$ is a pure NE for G_k , $\mathbf{x}$, which is created by stacking up all $\mathbf{x}_k$, is a pure NE for G .

Hence we have $\text{pureNE}(G) = \prod_{k=1}^d \text{pureNE}(G_k)$.

7.7 Proof of Proposition 13

Proof. First, the expected loss of player i is separable:

$$\begin{aligned}\mathbb{E}_{T_{-i}}[\ell_i(\mathbf{x}_i, \pi_{-i}; \mathbf{t}_i)] &= \mathbb{E}_{T_{-i}} \left[\sum_{k=1}^d \left(w_0 x_{0k} + w_i x_{ik} + \sum_{j \neq i} w_j \pi_{jk}(\mathbf{t}_j) - t_{ik} \right)^2 \mid \mathbf{t}_i \right] \\ &= \sum_{k=1}^d \mathbb{E}_{T_{-i}} \left[\left(w_0 x_{0k} + w_i x_{ik} + \sum_{j \neq i} w_j \pi_{jk}(\mathbf{t}_j) - t_{ik} \right)^2 \mid \mathbf{t}_i \right].\end{aligned}$$

Fixing any $\pi_{-i}(\mathbf{t}_{-i})$, the best response of player i given target $\mathbf{t}_i$ is:

$$\mathbf{x}_i^* = \arg \min_{\mathbf{x}_i \in \mathcal{X}} \sum_{k=1}^d \mathbb{E}_{T_{-i}} \left[\left(w_0 x_{0k} + w_i x_{ik} + \sum_{j \neq i} w_j \pi_{jk}(\mathbf{t}_j) - t_{ik} \right)^2 \mid \mathbf{t}_i \right] \quad (12)$$

When minimizing $\mathbb{E}_{T_{-i}}[\ell_i(\mathbf{x}_i, \pi_{-i}; \mathbf{t}_i)]$, term k of (12) depends on player i 's action only through x_{ik} , i.e. any change on $x_{ik'}$, $k' \neq k$ will not change term k . In addition, the domain is a product-space $\mathcal{X} = \prod_{k=1}^d \mathcal{X}_k$, so the choice of x_{ik} is unaffected by the choice in any other coordinate. Thus, minimizing $\mathbb{E}_{T_{-i}}[\ell_i(\mathbf{x}_i, \pi_{-i}; \mathbf{t}_i)]$ is equivalent to minimizing each of its d terms.

The distribution of targets $T_1, \dots, T_n$ are independent across players, so the expected action for each opponent is independent of $\mathbf{t}_i$, which means $\mathbb{E}_{T_j}[\pi_{jk}(\mathbf{t}_j) \mid \mathbf{t}_i] = \mathbb{E}_{T_j}[\pi_{jk}(\mathbf{t}_j)] \quad \forall j \neq i$, given i and k . We will use this to evaluate the derivative below.

Since term k is strictly convex in x_{ik} , its minimum on the domain $[L_k, U_k]$ occurs at $x_{ik} = L_k$ if and only if it is monotonically increasing on $[L_k, U_k]$, which happens when

$$\begin{aligned}\frac{d}{dx_{ik}} \mathbb{E}_{T_{-i}} \left[\left(w_0 x_{0k} + w_i x_{ik} + \sum_{j \neq i} w_j \pi_{jk}(\mathbf{t}_j) - t_{ik} \right)^2 \right] \Big|_{x_{ik}=L_k} &\geq 0 \\ \implies t_{ik} &\leq w_0 x_{0k} + w_i L_k + \sum_{j \neq i} w_j \mathbb{E}_{T_j}[\pi_{jk}(\mathbf{t}_j)].\end{aligned}$$

Similarly, the minimum on $[L_k, U_k]$ occurs at $x_{ik} = U_k$ if and only if term k is monotonically decreasing on $[L_k, U_k]$, which happens when

$$\begin{aligned}\frac{d}{dx_{ik}} \mathbb{E}_{T_{-i}} \left[\left(w_0 x_{0k} + w_i x_{ik} + \sum_{j \neq i} w_j \pi_{jk}(\mathbf{t}_j) - t_{ik} \right)^2 \right] \Big|_{x_{ik}=U_k} &\leq 0 \\ \implies t_{ik} &\geq w_0 x_{0k} + w_i U_k + \sum_{j \neq i} w_j \mathbb{E}_{T_j}[\pi_{jk}(\mathbf{t}_j)].\end{aligned}$$

The best response is an interior point of $[L_k, U_k]$ when neither condition above holds, i.e. when term k of (12) is non-monotonic on $[L_k, U_k]$. It happens when its unconstrained global minimizer x_{ik}^* lies strictly within (L_k, U_k) , i.e. $L_k < x_{ik}^* < U_k$.

The unconstrained best response x_{ik}^* is defined as

$$x_{ik}^* = \arg \min_{x_{ik} \in \mathbb{R}} \mathbb{E}_{T_{-i}} \left[\left(w_0 x_{0k} + w_i x_{ik} + \sum_{j \neq i} w_j \pi_{jk}(\mathbf{t}_j) - t_{ik} \right)^2 \right].$$

To find x_{ik}^* , we set the derivative of that expectation to 0, and we have:

$$\begin{aligned} \frac{d}{dx_{ik}} \mathbb{E}_{T_{-i}} \left[\left(w_0 x_{0k} + w_i x_{ik} + \sum_{j \neq i} w_j \pi_{jk}(\mathbf{t}_j) - t_{ik} \right)^2 \right] \Big|_{x_{ik}=x_{ik}^*} &= 0 \\ \implies x_{ik}^* &= \frac{1}{w_i} t_{ik} - \frac{1}{w_i} \left(w_0 x_{0k} + \sum_{j \neq i} w_j \mathbb{E}_{T_j} [\pi_{jk}(\mathbf{t}_j)] \right). \end{aligned}$$

Therefore, term k is minimized at

$$\pi_{ik}^*(\mathbf{t}_i) = x_{ik}^* = \begin{cases} L_k, & t_{ik} \leq a_{ik}, \\ \frac{1}{w_i} t_{ik} + c_{ik}, & a_{ik} < t_{ik} < b_{ik}, \\ U_k, & t_{ik} \geq b_{ik}, \end{cases}$$

where $c_{ik} = -\frac{1}{w_i} \left(w_0 x_{0k} + \sum_{j \neq i} w_j \mathbb{E}_{T_j} [\pi_{jk}(t_j)] \right)$, $a_{ik} = w_i (L_k - c_{ik})$, and $b_{ik} = a_{ik} + w_i (U_k - L_k) = w_i (U_k - c_{ik})$.

7.8 Proof of Example 14

Proof. By Proposition 13, the BNE set consists of ramp profiles parameterized by $(c_1, \dots, c_n)$ that satisfy the consistency system on our only dimension. We will solve for $c_i \forall i \in [n]$.

First, since $0 \leq \mathbb{E}_{T_j} [\pi_j(t_j)] \leq 1 \forall j, \pi_j$, for any BNE strategy profile $\{\pi_j^*(t_j)\}_{j=1}^n$, we have the following properties for a_i, b_i of player i :

$$\begin{aligned} a_i &= \sum_{j \neq i} w_j \mathbb{E}_{T_j} [\pi_j^*(t_j)] \geq \sum_{j \neq i} w_j \cdot 0 = 0 \\ b_i &= w_i + \sum_{j \neq i} w_j \mathbb{E}_{T_j} [\pi_j^*(t_j)] \leq w_i + \sum_{j \neq i} w_j = 1 \end{aligned}$$

This means for strategy $\pi_i^*(t_i)$ in any BNE, it must satisfy $0 \leq a_i < b_i \leq 1$. We are not limiting the action space of the game, because this constraint is directly

derived from existing constraints, including $\mathcal{X} = [0, 1]$ and the characterizations of a_i and b_i .

Therefore we can compute $\mathbb{E}_{T_j} [\pi_j^*(T_j)]$ under the uniform prior by assuming that $0 \leq a_i < b_i \leq 1$ holds for this BNE strategy:

$$\mathbb{E}_{T_j} [\pi_j^*(t_j)] = 0 \cdot a_j + 1 \cdot (1 - b_j) + \int_{a_j}^{a_j + w_j} \left(\frac{1}{w_j} t_j - \frac{a_j}{w_j} \right) dt_j = 1 + w_j c_j - \frac{w_j}{2}.$$

Substituting into the consistency system $c_i = -\frac{1}{w_i} \sum_{j \neq i} w_j \mathbb{E}_{T_j} [\pi_j^*(t_j)]$ yields the linear system

$$w_i c_i + \sum_{j \neq i} w_j^2 c_j = -\sum_{j \neq i} \left(w_j - \frac{w_j^2}{2} \right), \quad \forall i \in [n].$$

Let the matrix representing the LHS weights be M , with $M_{ij} = \begin{cases} w_i & \text{if } j = i \\ w_j^2 & \text{if } j \neq i \end{cases}$

subtracting row 1 from all other rows, and then applying row $1 - \frac{w_i^2}{w_i - w_i^2} \cdot \text{row } i$, $\forall i \in \{2, \dots, n\}$ makes a lower triangular matrix with non-zero diagonal values. This means that M is full rank and there is a unique solution to our system of equations. $c_i = -\frac{1-w_i}{2w_i} \forall i \in [n]$ is the unique solution here, with verification omitted.

If we plug $\{c_i\}_{i=1}^n$ back into the ramp function, the unique BNE strategy has $a_i = \frac{1-w_i}{2}$, $b_i = \frac{1+w_i}{2}$, and $c_i = -\frac{1-w_i}{2w_i}$, which matches our claim in Example 14.

7.9 Proof of Proposition 17

Proof. Given an arbitrary joint action from other players $\mathbf{x}_{-i}$, player i 's best response is

$$\begin{aligned} BR(\mathbf{x}_{-i}) &:= \operatorname{argmin}_{\mathbf{x}_i \in \mathcal{X}} \ell_i(\mathbf{x}_i, \mathbf{x}_{-i}) = \operatorname{argmin}_{\mathbf{x}_i \in \mathcal{X}} -\mathbf{t}_i^\top (w_0 \mathbf{x}_0 + \sum_{j=1}^n w_j \mathbf{x}_j) \\ &= \operatorname{argmin}_{\mathbf{x}_i \in \mathcal{X}} -w_i \mathbf{t}_i^\top \mathbf{x}_i + \text{const} = \mathbf{x}_i^*. \end{aligned}$$

The best response is independent of $\mathbf{x}_{-i}$.

7.10 Proof of Corollary 19

Proof. The proof is identical to the one for Theorem 3, since the loss functions are identical, only on a different domain.

Bibliography

- Zohar Barak, Anupam Gupta, and Inbal Talgam-Cohen. Mac advice for facility location mechanism design. *Advances in Neural Information Processing Systems*, 37:129564–129604, 2024.
- Dolors Berga. Single-peakedness and strategy-proofness of generalized median voter schemes. *Social Choice and Welfare*, 19(1):175–192, 2002.
- Stephen Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, 2004.
- Hau Chan, Aris Filos-Ratsikas, Bo Li, Minming Li, and Chenhao Wang. Mechanism design for facility location problem: A survey. In *The 30th International Joint Conference on Artificial Intelligence (IJCAI 2021)*, pages 1–17, 2021.
- Robert T Clemen. Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583, 1989.
- Roger Cooke. *Experts in uncertainty: opinion and subjective probability in science*. Oxford university press, 1991.
- Morris H DeGroot and Julia Mortera. Optimal linear opinion pools. *Management Science*, 37(5):546–558, 1991.
- Alina Filimonov and Reshef Meir. Strategyproof facility location mechanisms on discrete trees. *Autonomous Agents and Multi-Agent Systems*, 37(1):10, 2023.
- Alex Gershkov, Benny Moldovanu, and Xianwen Shi. Optimal voting rules. *The Review of Economic Studies*, 84(2):688–717, 2017.
- Sumit Goel and Wade Hann-Caruthers. Optimality of the coordinate-wise median mechanism for strategyproof facility location in two dimensions. *arXiv preprint arXiv:2007.00903*, 2020.
- John T Jost, Delia S Baldassarri, and James N Druckman. Cognitive-motivational mechanisms of political polarization in social-communicative contexts. *Nature reviews psychology*, 1(10):560–576, 2022.
- Emily Kubin and Christian Von Sikorski. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association*, 45(3):188–206, 2021.
- Andrew D Martin. Congressional decision making and the separation of powers. *American Political Science Review*, 95(2):361–378, 2001.
- Eric S Maskin. Mechanism design: How to implement social goals. *American Economic Review*, 98(3):567–576, 2008.
- Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.
- Hervé Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.
- Abraham Neyman. Correlated equilibrium and potential games. *International Journal of Game Theory*, 26(2):223–227, 1997.
- Martin J Osborne. *A course in game theory*. MIT Press, 1994.
- Ariel D Procaccia and Moshe Tennenholtz. Approximate mechanism design without money. *ACM Transactions on Economics and Computation (TEAC)*, 1(4):1–26, 2013.

- Jan-Willem Romeijn. An interpretation of weights in linear opinion pooling. *Episteme*, 21(1):19–33, 2024.
- Tim Roughgarden. Algorithmic game theory. *Communications of the ACM*, 53(7):78–86, 2010.
- James Schummer and Rakesh V Vohra. Strategy-proof location on a network. *Journal of Economic Theory*, 104(2):405–428, 2002.
- Steven Tadelis. *Game Theory: An Introduction*. Princeton University Press, 2013.
- Paul Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications*, 109(3):475–494, 2001.
- Jay J Van Bavel, Steve Rathje, Elizabeth Harris, Claire Robertson, and Anni Sternisko. How social media shapes polarization. *Trends in cognitive sciences*, 25(11):913–916, 2021.
- Toby Walsh. Equitable mechanism design for facility location. In *Third IJCAI Workshop on Computational Fair Division*, 2025.
- Xiaoqian Wang, Rob J Hyndman, Feng Li, and Yanfei Kang. Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4):1518–1547, 2023.
- Stephen J Wright and Benjamin Recht. *Optimization for data analysis*. Cambridge University Press, 2022.
- Emmanouil Zampetakis and Fred Zhang. Bayesian strategy-proof facility location via robust estimation. In *International Conference on Artificial Intelligence and Statistics*, pages 4196–4208. PMLR, 2023.