Look-Ahead Reasoning on Learning Platforms

Haiqing Zhu
Australian National University
haiqing.zhu@anu.edu.au

Tijana Zrnic
Stanford University
tijana.zrnic@stanford.edu

Celestine Mendler-Dünner
ELLIS Institute Tübingen
MPI for Intelligent Systems, Tübingen
Tübingen AI Center
celestine@tue.ellis.eu

Abstract

Predictive models are often designed to minimize risk for the learner, yet their objectives do not always align with the interests of the users they affect. Thus, as a way to contest predictive systems, users might act strategically in order to achieve favorable outcomes. While past work has studied strategic user behavior on learning platforms, the focus has largely been on strategic responses to the deployed model, without considering the behavior of other users, or implications thereof for the deployed model. In contrast, *look-ahead reasoning* takes into account that user actions are coupled, and—at scale—impact future predictions. Within this framework, we first formalize level-k thinking, a concept from behavioral economics, where users aim to outsmart their peers by looking one step ahead. We show that, while convergence to an equilibrium is accelerated, the equilibrium remains the same, providing no benefit of higher-level reasoning for individuals in the long run. Then, we focus on collective reasoning, where users take coordinated actions by optimizing through their impact on the model. By contrasting collective with selfish behavior, we characterize the benefits and limits of coordination; a new notion of alignment between the learner's and the users' utilities emerges as a key concept. We discuss connections to several related mathematical frameworks, including strategic classification, performative prediction, and algorithmic collective action.

1 Introduction

Increasingly, digital platforms deploy learning algorithms that collect and analyze data about individuals to power services, personalize experiences, and allocate resources. As people come to understand how these systems make decisions, they often adapt strategically to improve their outcomes.

Prior research has largely modeled such strategic behavior as *unilateral*: each agent responds to the platform's decision rule by optimizing their own outcome while treating that rule as fixed. For example, a job applicant might rephrase their resume to include keywords that align with an automated screening system's preferences. This perspective neglects the fact that many others may be doing the same—thereby collectively shifting the data distribution from which the platform learns in the future.

In reality, there is ample empirical evidence that users frequently reason about one another's behavior. They may act in solidarity [Tassinari and Maccarrone, 2020], coordinate to amplify their collective influence [Chen, 2018], oftentimes facilitated by labor organizations,[1] or anticipate other

[1] https://www.drivers-united.org/

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

people's adaptations to gain an advantage [Kneeland, 2015]. On learning platforms in particular, such reasoning involves anticipating the behavior of other platform participants and how those behavioral changes will impact the learning algorithm in the future. We call this *look-ahead reasoning*. In the resume-screening example, look-ahead reasoning might surface as choosing to emphasize distinct keywords that others have abandoned, anticipating that popular buzzwords will lose predictive value as they become widespread.

1.1 Our contributions

We study the impact of look-ahead reasoning on learning platforms—user behavior that anticipates the actions of others in the population—by characterizing how it reshapes learning dynamics and equilibria. We begin with selfish agents, who act independently but strategically accounting for the other agents' responses. We then turn to coordinated behavior through collective action, where agents act jointly and strategize against a predictive model.

To capture agents who selfishly aim to outsmart their peers, we formalize the concept of *level-k* thinking [Nagel, 1995] from behavioral economics in the context of data-driven learning. Level-k thinking captures different depths of strategic thought: a level-k thinker acts assuming they are "one step ahead" of all other individuals in the population, who are level-$(k-1)$ thinkers. A level-1 thinker acts assuming everyone in the population is non-strategic, i.e., a level-0 thinker. Higher levels k are defined recursively. We study the dynamics of repeatedly retraining a model acting on a population of level-k thinkers. We show that "deeper" thinking achieved for larger k accelerates the learning dynamics, while resulting in the *same* equilibrium solution, no matter the depth of thinking.

Theorem 1 (Informal)**.** *For $k \geq 1$, let $\alpha_k \in (0,1)$ be the fraction of level k-thinkers in the population, $\sum_{k=1}^{\infty} \alpha_k = 1$. Assume the learner minimizes a loss function that is smooth and strongly convex, and suppose that the agent responses are sufficiently Lipschitz in the model parameters. Then, for some constant $c \in (0,1)$, repeated retraining converges to a unique stable point at rate*

$$O\left(\Big[\sum_{k=1}^{\infty} c^k \alpha_k\Big]^t\right).$$

Therefore, selfish behavior, even if it relies on higher levels of reasoning, does not improve the agents' utility in equilibrium.

Next, we show that agents can move past this obstacle if they coordinate. Look-ahead reasoning with coordination allows anticipating—and thus steering—model updates that result from population behavior. We show that the gap between coordination and lack thereof in terms of agent utility is governed by a notion of alignment between the objectives of the learning platform and the population. Below, we use $\ell(z,\theta)$ and $u(z,\theta)$ to denote the learner's loss and the agent utility for deploying model θ on instance z. Furthermore, $\langle a, b\rangle_M := a^\top M b$.

Theorem 2 (Informal)**.** *Let $\mathcal{D}^*$ and $\mathcal{D}^\sharp$ denote the population's data distributions at equilibrium under selfish reasoning and under collective reasoning, respectively. Then, the benefit of coordination, defined as the difference in population utility at the two equilibria, is bounded as*

$$B \leq \left(\langle \mathbb{E}_{z\sim\mathcal{D}^*}\left[\nabla_\theta u^*\right], \mathbb{E}_{z\sim\mathcal{D}^\sharp}\left[\nabla_\theta \ell^*\right]\rangle_{(\mathbf{H}^\star)^{-1}}\right)^2,$$

where $\mathbf{H}^\star = \mathbb{E}_{z\sim\mathcal{D}^*}\left[\nabla^2_\theta \ell(z,\theta^*)\right]$. We use the short-hand notation $\nabla_\theta u^* = u(z,\theta^*)$, $\nabla_\theta \ell^* = \ell(z,\theta^*)$, where θ^* denotes the equilibrium model under selfish reasoning.

Thus, if the average agent utility and the average loss of the learner are orthogonal, there is no benefit to coordination. However, when there is sufficient overlap between the objectives that the collective can exploit, coordination can lead to more favorable outcomes than selfish reasoning.

In additional results, we study heterogeneous populations comprised of selfish agents and collectives of varying size. The results shed light on the benefits and limitations of coordination; for example, bigger collectives do not always lead to a higher average utility for the collective. We also study the impact of selfish agents opting out from the collective on the learning dynamics, showing that broader participation in the collective implies faster convergence.

1.2 Background and related work

Strategic classification [Hardt et al., 2016, Brückner and Scheffer, 2011] introduces a model to study strategic behavior in learning systems based on assumptions of individual rationality. It describes a

population of agents best-responding to a decision rule by altering their features to achieve positive predictions, given a fixed decision rule. Several variations of this basic model have been studied [e.g., Dong et al., 2018, Chen et al., 2020, Bechavod et al., 2021, Ghalme et al., 2021, Jagadeesan et al., 2021]; see Podimata [2025] for a recent survey of this literature. All these works focus on studying how agents strategize against a fixed decision rule. Our work introduces a new dimension of reasoning to strategic classification, taking into account how individual agents' actions are coupled and how this influences the model the agents strategize against.

Performative prediction [Perdomo et al., 2020] introduces performative stability as an equilibrium notion that characterizes long-term outcomes in the interaction of a population with a learning system. Performative stability is a fixed point of repeated retraining by the learner in a dynamic environment. Prior work in performative prediction [Perdomo et al., 2020, Mendler-Dünner et al., 2020, Drusvyatskiy and Xiao, 2023, Brown et al., 2022, Narang et al., 2023] has studied the behavior of retraining and conditions that ensure its convergence to stability in different learning settings. We refer to Hardt and Mendler-Dünner [2025] for a more extensive overview of the performative prediction literature. A key concept in performative prediction is the "distribution map," which characterizes how different model deployments impact the population. This map is typically treated as a fixed unknown quantity. We study how different types of strategic reasoning impact the distribution map, thus also impacting the resulting convergence properties.

A more recent literature on algorithmic collective action [Hardt et al., 2023] studies coordinated agent efforts with the goal of steering learning systems; see [Baumann and Mendler-Dünner, 2024, Ben-Dov et al., 2024, Gauthier et al., 2025, Sigg et al., 2025] for recent developments in this area, as well as related discussions of data leverage [Vincent et al., 2021]. From the perspective of our work, collective action is a type of look-ahead reasoning: agents plan through model updates under the assumption that they coordinate with other agents. We study the tradeoffs and implications of coordinated reasoning to population utility. Relatedly, Hardt et al. [2022] discuss how platforms can reduce risk by actively steering a population. Collective action reverses this perspective and shows how the population can improve its utility by steering the learner. This perspective is related to [Zrnic et al., 2021], who also deviate from the classical model of strategic classification and instead model the population as the leader in the Stackelberg game against the learning platform.

Finally, at a technical level, our work leverages ideas from game theory. Balduzzi et al. [2018] proposed the decomposition of differentiable games into the "Hamiltonian" part and the "potential" part through the decomposition of the Hessian of the game. Our work also utilizes the Hessian to describe the alignment of utilities between the learner and the collective.

2 Setup

We consider a population of individuals interacting with a learning platform. We assume the platform trains a predictive model on the population's data, and individuals strategically alter their data to achieve favorable outcomes. We elaborate below.

Learning platform. Upon observing data about the population, the learner optimizes the parameters $\theta \in \Theta$ of their predictive model f_θ. We work with the following optimality assumption on the learning algorithm: given a loss function ℓ, the learner's response $\mathcal{A}(\mathcal{D})$ to a data distribution $\mathcal{D}$ is given by *risk minimization*, defined as

$$\mathcal{A}(\mathcal{D}) := \operatorname*{arg\,min}_{\theta \in \Theta} \ \mathbb{E}_{z \sim \mathcal{D}} \left[\ell(z, \theta)\right].$$

Strategic agents. We assume individuals are described by data points $z \in \mathcal{Z}$ sampled from a base distribution $\mathcal{D}_0$. Typically, $z = (x, y) \in \mathcal{X} \times \mathcal{Y}$ are feature–label pairs. Individuals implement a data modification strategy $h_\theta : \mathcal{Z} \to \mathcal{Z}$ that maps an individual's data point z to a modified data point $h_\theta(z)$; the strategy can depend on the learning platform's currently deployed model θ. We will sometimes omit the subscript θ if the strategy is independent of the current model, i.e., $h_\theta \equiv h_{\theta'}$ for all θ, θ'. We use $\mathcal{D}_{h_\theta}$ to denote the distribution of $h_\theta(z)$ for $z \sim \mathcal{D}_0$; in other words, this is the distribution of data points after applying strategy h_θ to all base data points. Following the terminology of Perdomo et al. [2020], we call $\mathcal{D}_{h_\theta}$ a *distribution map*. Note that different strategies h_θ correspond to different distribution maps. When the strategy is clear from the context, we will write $\mathcal{D}_{h_\theta} \equiv \mathcal{D}(\theta)$. The notation $\mathcal{D}_h$ equally applies to model-independent strategies h.

Equilibria and learning dynamics. We study the long-term behavior of the learner repeatedly optimizing their model. Formally, we study the learning dynamics of *repeated risk minimization*:

$$\theta_{t+1} = \mathcal{A}(\mathcal{D}_{h_{\theta_t}}). \tag{1}$$

The natural equilibrium of these dynamics is called *performative stability* [Perdomo et al., 2020]. We say a model θ^* is performatively stable with respect to a strategy h_θ if

$$\theta^* = \mathcal{A}(\mathcal{D}_{h_{\theta^*}}).$$

In words, there is no reason for the learner to deviate from the current model given the data distribution it implies.

Population utility. Different strategies h_θ lead to different equilibria. Rather than just focusing on the learner's loss, we evaluate equilibria in terms of the utility they imply for the population. We denote the population's utility after implementing strategy h_θ by

$$U(h_\theta) = \mathbb{E}_{z\sim\mathcal{D}_{h_{\theta^*}}}\left[u(z,\theta^*)\right],$$

where $u(z,\theta)$ is the utility of an individual with data point z when the deployed model is θ, and θ^* denotes the equilibrium model under strategy h_θ, i.e., the performatively stable point. Again, we will sometimes omit the subscript when denoting the strategy if it is independent of the deployed model.

3 Level-k reasoning

Strategic classification [Hardt et al., 2016] assumes that each individual selfishly best-responds to a deployed model θ. As explained earlier, this model does not take into account the agents' awareness that they as a whole determine the deployed model. To account for this dimension of reasoning, we build on the cognitive hierarchy framework from behavioral economics [Nagel, 1995] and generalize strategic classification to allow individuals to reason through the other individuals' responses. In particular, we formalize *level-k* thinking, which categorizes players by the "depth" of their strategic thought. Intuitively, an individual reasoning at level k assumes a level of cognitive reasoning for the rest of the population and tries to "outsmart" them. In other words, they are always one step ahead: a level-k thinker best-responds to the model that would result from a population of level-$(k-1)$ thinkers. The basic level-k model starts with an explicit assumption about how individuals at level 0 behave. It then defines higher levels of thinking recursively.

Suppose that agents at level 0 are non-strategic and implement $h_\theta^{(0)}(z) = z$ in response to all θ. Then, for every higher level of thinking $k \geq 1$ we define the strategy for level-k thinkers recursively as

$$h_\theta^{(k)}(z) := \mathrm{argmax}_{z'} u(z', \mathcal{A}(\mathcal{D}_{k-1}(\theta))), \tag{2}$$

where $\mathcal{D}_{k-1}(\theta)$ is the distribution obtained by applying the strategy $h_\theta^{(k-1)}$ in response to a deployed model θ to every $z \sim \mathcal{D}_0$. At level $k = 1$, we recover the standard microfoundation model of strategic classification [Hardt et al., 2016], where individuals best-respond to a fixed model. For larger k, the agents anticipate the actions of other agents and best-respond to the hypothetical model resulting from the shifted distribution.

Different individuals in the population might implement different levels of reasoning. To reflect this we deviate from a homogeneous population and let the population consists of level-k thinkers at different levels k. In particular, we assume that α_k-fraction of the population has cognitive level k, for $k = 1, 2, \ldots$ and $\sum_{k=1}^{\infty} \alpha_k = 1$. If $\alpha_k = 1$ for some k, then all individuals in the population have the same level of reasoning. This model results in the distribution map:

$$\mathcal{D}(\theta) := \sum_{k=1}^{\infty} \alpha_k \mathcal{D}_k(\theta). \tag{3}$$

We characterize the learning dynamics for different levels of thinking. We use the following Lipschitzness assumption on the induced distribution at level $k = 1$:

$$\mathcal{W}(\mathcal{D}_1(\theta), \mathcal{D}_1(\theta')) \leq \epsilon \left\|\theta - \theta'\right\|_2, \quad \forall \theta, \theta' \in \Theta,$$

where $\mathcal{W}$ denotes the Wasserstein-1 distance. This condition is known as ϵ-sensitivity [Perdomo et al., 2020].

Theorem 3 (Retraining with level-k thinkers)**.** *Suppose ℓ is γ-strongly convex and β-smooth in z, and that the distribution map $\mathcal{D}_1(\theta)$ is ϵ-sensitive. Then, as long as $\epsilon < \frac{\gamma}{\beta}$, there is a unique stable point θ^* such that for any $(\alpha_k)_{k=1}^{\infty}$ retraining on the mixed population* (3) *converges as*

$$\|\theta_t - \theta^*\|_2 \leq \left(\sum_{k=1}^{\infty} \left(\frac{\epsilon\beta}{\gamma} \right)^k \alpha_k \right)^t \|\theta_0 - \theta^*\|_2 . \tag{4}$$

The core technical step in the proof is to show how the sensitivity of the distribution map $\mathcal{D}_k(\theta)$ changes recursively with k. In particular, the distribution map $\mathcal{D}(\theta)$ in (3) has sensitivity $\sum_{k=1}^{\infty} \alpha_k \left(\epsilon\beta/\gamma\right)^{k-1} \epsilon$. We refer to Appendix A for the full poof.

For the case where $\alpha_1 = 1$ and thus all agents reason at level $k = 1$, we recover the retraining result of Perdomo et al. [2020]. There are two interesting implications of the generalization in Theorem 3. First, we observe that for populations with higher levels of thinking k, the rate of convergence increases (although the condition for convergence, $\epsilon < \gamma/\beta$, remains the same). This can be interpreted as saying that performative distribution shifts are mitigated when the population has a deeper level of strategic thought. The second implication is that, as long as agents act selfishly, they cannot benefit from higher levels of reasoning at stability.

Corollary 1. *Under the assumptions of Theorem 3, it holds that*

$$U(h_\theta^{(1)}) = U(h_\theta^{(k)}),\ \forall k \geq 1.$$

Moreover, the utility at stability remains unaltered for any mixed population consisting of level-k thinkers regardless of $(\alpha_k)_{k=1}^{\infty}$.

This corollary follows from the observation that the stable point θ^* is the same for any mixed population of level-k thinkers. Another consequence of this fact is that the equilibrium strategies are identical for every k.

In the following sections we will denote this unique *optimal selfish strategy* by $h^* = h_{\theta^*}^{(k)}$ and the implied data distribution by $\mathcal{D}^*$.

4 Collective reasoning

So far we considered individuals who make use of higher levels of thinking to reason through the actions of others, allowing them to anticipate model changes implied by the population's actions. We saw that higher levels of reasoning do not improve their utility at equilibrium. The fundamental reason is that individually they cannot steer the trajectory of the learning algorithm; they can merely anticipate it. In the following we show how individuals can achieve more favorable outcomes by joining forces and making decisions *collectively*; this gives them steering power.

We denote by $h^\sharp$ the *optimal collective strategy*:

$$h^\sharp = \arg\max_h U(h) = \arg\max_h \mathbb{E}_{z\sim\mathcal{D}_h} \left[u(z, \mathcal{A}(\mathcal{D}_h)) \right].$$

Notice the difference compared to (2). In (2), the optimization variable z' does not enter the model training $\mathcal{A}$, while above h directly determines the subsequently deployed model. The optimal collective strategy is a Stackelberg equilibrium: the population acts as the Stackelberg leader.

To contrast the optimal collective strategy with the optimal selfish strategy, we define the benefit of coordination.

Definition 1 (Benefit of coordination)**.** *Let h^* be the optimal selfish strategy and $h^\sharp$ the optimal collective strategy. We define the benefit of coordination as*

$$\mathrm{B} = U(h^\sharp) - U(h^*).$$

Since $h^\sharp$ is globally optimal, it holds that $\mathrm{B} \geq 0$. How large B is depends on the goals pursued by the learner and the population, as characterized by ℓ and u, respectively. Through coordinated data modifications, the population can steer the model towards a common target. But to do so they have to deviate from their individually optimal strategy. Thus, what governs the benefit of coordination is the tradeoff between the loss experienced by taking locally suboptimal actions and the gain achieved by steering the model.

We start with a simple case where the benefit of coordination is zero.

Proposition 4. *Suppose $u = c \cdot \ell$ for some $c \neq 0$. Then, it holds that $B = 0$.*

In an adversarial setting where $c > 0$, the game between the learner and the collective is a zero-sum game. In this case the benefit of coordination is zero, as the cost of steering is equal to its return. When $c < 0$, the platform and the agents pursue the same goal and the game becomes a potential game between the two. In this case selfish actions are simultaneously optimal for the collective and B is again zero.

To further understand the benefit of coordination beyond this special case, we consider linear distribution maps. Additional results can be found in Appendix B.

Assumption 1 (Linearity)**.** *Let each strategy $h \in \mathcal{H}$ be represented by a parameter vector $\eta(h) \in \mathbb{R}^d$ through a parameterization $\eta : \mathcal{H} \to \mathbb{R}^d$, where $\mathcal{H}$ denotes the strategy space. Define the induced distribution map $\tilde{\mathcal{D}} : \mathbb{R}^d \to \Delta(\mathcal{Z})$ by composition, $\tilde{\mathcal{D}}(\eta(h)) := \mathcal{D}_h$, where $\Delta(\mathcal{Z})$ denotes the space of probability distributions over the support $\mathcal{Z}$. We say that the distribution map is* linear with respect to the parameterization *if*

$$\tilde{\mathcal{D}}\big(\alpha\,\eta(h) + (1-\alpha)\,\eta(h')\big) = \alpha\,\tilde{\mathcal{D}}\big(\eta(h)\big) + (1-\alpha)\,\tilde{\mathcal{D}}\big(\eta(h')\big), \qquad \forall\,\alpha \in [0,1].$$

Intuitively, linearity means that the population's data distribution is the same whether agents linearly interpolate between two strategies h and h', or they split up in two subgroups and each implements one of the two strategies. Under this assumption, the following result provides a bound on the benefit of coordination.

Theorem 5 (Bound on the benefit of coordination)**.** *Let Assumption 1 hold. Let $U(h)$ be γ-strongly concave in the parameterisation $\eta(h)$ and $U(\alpha h + (1-\alpha)h')$ be differentiable with respect to α. Then, we have*

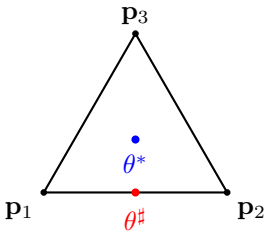
$$\mathrm{B} \leq \frac{1}{2\gamma}\Phi^2,$$

where

$$\Phi := \left\langle \mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta u(z,\theta^*)\right], \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[\nabla_\theta \ell(z,\theta^*)\right]\right\rangle_{(\mathbf{H}^\star)^{-1}} \text{ and } \mathbf{H}^\star = \mathbb{E}_{z\in\mathcal{D}_{h^*}}\left[\nabla^2_{\theta,\theta}\ell(z,\theta^*)\right].$$

This result shows how the benefit of coordination is governed by the alignment between the utility u of the population and the loss ℓ of the learner, quantified by the inner product of their gradients at equilibrium. Moreover, the inverse-Hessian weighting in Φ shows that what really matters is not a raw gradient alignment but one filtered through the local curvature of the loss landscape. Indeed, the directions that the learner finds "flat" (small Hessian eigenvalues) allow for more influence on the model through small data modifications, and thus they offer more leverage. In line with Proposition 4, when $u = c \cdot \ell$ for some constant c, we have $\Phi = 0$ because the gradients in the inner product become zero at stability. In addition, $\Phi = 0$ for the case where the gradients are orthogonal and the two functions are unrelated. To show that Φ can be positive in general, we present a simple case with a close-form expression for Φ in the following example.

Example. Consider a toy setting where the learner estimates the centroid θ of a distribution supported on three anchor points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$—the corners of an equilateral triangle. The collective applies a strategy h that moves the data point to one of the three anchors with probability (w_1, w_2, w_3), forming a distribution $\mathcal{D}_h = \sum_i w_i \delta_{\mathbf{p}_i}$ with $\sum_i w_i = 1$ and $w_i \geq 0$. Thus, the strategy is parameterized by $\mathbf{w}$. The learner minimizes the squared loss $\ell(z,\theta) = \|z-\theta\|^2$ over the distribution $\mathcal{D}_h$. The collective, on the other hand, prefers the centroid to lie between $\mathbf{p}_1$ and $\mathbf{p}_2$, and maximizes



$$u(z,\theta) = -\|\mathbf{p}_1 - \theta\|_2^2 - \|\mathbf{p}_2 - \theta\|_2^2 - \|z-\theta\|_2^2.$$

Here, it can be seen that $\mathcal{A}(\mathcal{D}_h) = \sum_i w_i x_i$ and hence $U(w) = -\|P\mathbf{w}\|^2 + 2(\mathbf{p}_1+\mathbf{p}_2)^\top P\mathbf{w} - \sum_{i=1}^3 w_i \|\mathbf{p}_i\|_2^2 + \text{const}$ where $P = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3]$.

When the collective distributes mass uniformly, i.e., $h^* = (1/3, 1/3, 1/3)$, the resulting centroid $\theta^* = \frac{1}{3}\sum_{i=1}^3 \mathbf{p}_i$ is a performatively stable point. There is no incentive for either party to change their strategy since they are both best-responding to the current state. However, a look-ahead collective

would prefer $h^\sharp = (1/2, 1/2, 0)$, which can be calculated by directly maximizing $U(h)$. This leads to a look-ahead optimal point $\theta^\sharp = \frac{1}{2}(\mathbf{p}_1 + \mathbf{p}_2)$, which deviates from the performative stability.

In this example, Assumption 1 is satisfied. $\Phi^2 > 0$ since $\Phi = -2r^2$ where $r := \|\mathbf{p}_3\|_2$. The benefit of coordination is $B = U(h^\sharp) - U(h^*) = \frac{3}{4}r^2$ is strictly positive as long as the anchor points are appropriately spaced. One can check that U is $2r^2$-strongly concave, and Theorem 5 can be verified since $B = \frac{3}{4}r^2 \leq \frac{\Phi^2}{2\gamma} = r^2$ which is tight up to a factor 1/4 coming from the slack in the strong concavity assumption.

5 Heterogeneous populations

A perfectly coordinated population, or one that implements the *optimal* collective strategy, is unlikely to emerge in practice. In the following we consider some plausible deviations from the idealized collective studied in the previous section and discuss how this impacts agent utilities and outcomes. Unless stated otherwise, we assume the collective implements any fixed strategy h (independent of θ), which could be a simpler alternative to a potentially hard-to-implement optimal strategy $h^\sharp$.

5.1 Learning dynamics in the presence of selfish agents

The learning dynamics of repeated risk minimization under idealized collective reasoning converge in a single step, since the strategy is fixed and independent of θ. However, this changes as soon as some agents deviate from the collective strategy. To reflect this scenario we consider the following mixture model:

$$\mathcal{D}^\alpha(\theta) = \alpha\mathcal{D}_h + (1-\alpha)\cdot\mathcal{D}(\theta), \tag{5}$$

where an α-fraction of the population implements the collective strategy h and the remaining $(1-\alpha)$-fraction deviates from the collective strategy. We assume this remainder of the population acts selfishly, and their behavior can be characterized by $\mathcal{D}(\theta)$. The actions of these agents can depend on the deployed model, such as in level-k reasoning discussed in Section 3.

We characterize the rate of convergence of repeated risk minimization under this model.

Proposition 6. *Consider the heterogeneous population model* (5)*. Suppose ℓ is γ-strongly convex and β-smooth in z, and that the distribution map $\mathcal{D}(\theta)$ is ϵ-sensitive. Then, as long as $\epsilon < \frac{\gamma}{\beta}$, repeated risk minimization is guaranteed to converge to a unique stable point θ^*_α at rate*

$$\|\theta_t - \theta^*_\alpha\|_2 \leq \left(\frac{\epsilon\beta(1-\alpha)}{\gamma}\right)^t \|\theta_0 - \theta^*_\alpha\|_2 . \tag{6}$$

This result shows how the sensitivity of the non-participating agents to changes in the deployed model, together with the fraction of these agents, determines the rate of convergence to stability. The smaller α the slower the rate of convergence. Thus, larger collectives have the advantage of stabilizing the learning dynamics.

5.2 Limits of coordination in the presence of non-strategic agents

Next, we study conditions under which it is worth scaling up a strategy h, meaning when a larger collective implies a higher utility for the collective. To study how the collective's utility changes with its size, we study the following mixture model for the population

$$\mathcal{D}^\alpha = \alpha\mathcal{D}_h + (1-\alpha)\cdot\mathcal{D}_0, \tag{7}$$

where the $(1-\alpha)$-fraction of agents deviating from the collective strategy are non-strategic. The collective strategy h can be any fixed strategy, not necessarily the optimal one. In the following, we are interested in the average utility for agents participating in the collective, denoted as $U_\alpha := \mathbb{E}_{z\sim\mathcal{D}_h}[u(z;\theta^*_\alpha)]$, where θ^*_α denotes the equilibrium under the mixture model (7).

Proposition 7 (Benefit of scaling up a strategy)**.** *Consider the mixture model in* (7)*, fix a strategy h, and denote the resulting equilibrium by θ^*_α. Then, the benefit of scaling up h at size α is positive, i.e.,*

$$\frac{\partial U_\alpha}{\partial\alpha} > 0, \quad \textit{if and only if} \quad \langle \mathbb{E}_{z\sim\mathcal{D}_h}[\nabla_\theta u(z;\theta^*_\alpha)], \mathbb{E}_{z\sim\mathcal{D}_h}[\nabla_\theta \ell(z;\theta^*_\alpha)]\rangle_{\mathbf{H}^{-1}} < 0,$$

where

$$\mathbf{H} := \nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\mathcal{D}^\alpha}[\ell(z;\theta^*_\alpha)] .$$

The result is a direct consequence of the envelope theorem. It provides a condition for whether a strategy is worth scaling up or not. The condition is again linked to a notion of alignment described by the inner product between the loss and utility gradients. Note that the result holds for any fixed strategy h. Suppose $\mathbf{H}$ is positive semi-definite; then, if $u = \ell$, $\frac{\partial U}{\partial \alpha} \leq 0$, and if $u = -\ell$, $\frac{\partial U}{\partial \alpha} \geq 0$.

Finally, we aim to understand what collectives can achieve if they are aware of partial participation and optimize their strategy accordingly. We define the optimal size-aware collective strategy for size α as:

$$h^\sharp_\alpha = \arg\max_h \mathbb{E}_{z\sim\mathcal{D}_h} \left[u(z; \mathcal{A}(\alpha\mathcal{D}_h + (1-\alpha)\mathcal{D}_0)\right]. \tag{8}$$

The global Stackelberg solution $h^\sharp$ corresponds to the case where $\alpha = 1$ and the collective utility corresponds to the population utility. In the case where $\alpha < 1$, the collective optimizes the utility of participants, rather than the full population. Agents are informed of their collective size and will choose the best strategy $h^\sharp_\alpha$ accordingly. In the following proposition, we characterize the utility of agents participating in a collective that deploys a size-aware strategy. We use U^*_α to denote the utility of a population of size α implementing the optimal size-aware strategy $h^\sharp_\alpha$.

Proposition 8 (Benefit of larger collectives)**.** *Consider the mixture model* (7) *with* $h = h^\sharp_\alpha$*. Then, the utility* U^*_α *achieved by implementing the optimal size-aware strategy* $h^\sharp_\alpha$ *satisifies*

$$\frac{\partial U^*_\alpha}{\partial \alpha} \geq 0 \quad \textit{if and only if} \quad \left.\frac{\partial U_\alpha}{\partial \alpha}\right|_{h=h^\sharp_\alpha} \geq 0.$$

Note that the derivative in the first term takes into account the dependence of the strategy on α. Thus, the result says that reoptimizing a strategy as a function of collective size does not change whether scaling up is worth it or not. The argument involves considering how the equilibrium changes after reoptimizing the strategy and evaluating this change against the overall change in the population.

6 Simulations

We validate our theoretical findings empirically. We adapt the credit-scoring simulator from Perdomo et al. [2020] that models how a lending institution classifies loan applicants by creditworthiness.[2]

6.1 Retraining dynamics under level-k thinking

We consider a strategic classification setup with a logistic regression classifier $\theta \in \mathbb{R}^{10}$. Let S be the set of the strategic features that the agents can manipulate. We choose this to be: remaining credit card balance, open credit lines, and number of real estate loans. Given some $\epsilon > 0$, the utility of the agents is given by

$$u_\epsilon(z, \theta_S) = -\langle \theta, z\rangle - \frac{1}{2\epsilon}\|z_0 - z\|_2^2, \tag{9}$$

where z_0 is their feature value under $\mathcal{D}_0$. Assuming the agents can only manipulate coordinates of z corresponding to strategic features, the best response of the agents for these coordinates would be $z^*_S = z_S - \epsilon\theta_S$, corresponding to the strategy for agents thinking at level-1. It is not hard to see that the resulting distribution map $\mathcal{D}_1(\theta)$ is ϵ-sensitive. Under this model we simulate the repeated retraining dynamics for different populations of level-k thinkers, with $(\alpha_1, \alpha_2) \in \{(0.9, 0.1), (0.5, 0.5), (0.1, 0.9)\}$, where α_k is the fraction of the agents who are level-k thinkers.

In Figure 1 we report the speed of convergence with $\epsilon = 0.5$ by presenting the iterate gap $\|\theta_{t+1} - \theta_t\|_2$ against the number of iterations. First, we can see that under all three mixtures the gap tends to zero and the dynamics converge. As the fraction of higher levels of thinking increases, the speed of convergence increases, which is in line with our theoretical finding in Theorem 3. We can verify empirically that the dynamics converge to a unique equilibrium independent of α.

[2]The implementation of the simulation can be found in `https://github.com/haiqingzhu543/Look-Ahead-Reasoning-on-Learning-Platforms`. The dataset is available at `https://www.kaggle.com/c/GiveMeSomeCredit/data`

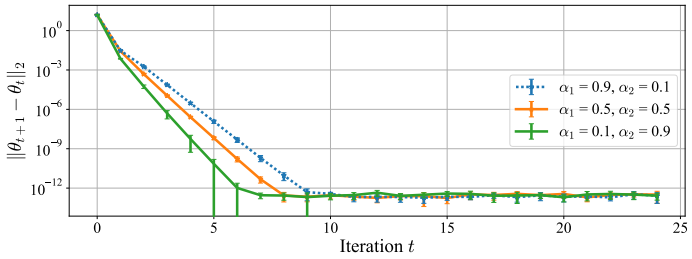


Figure 1: Convergence of repeated risk minimization. The x-axis is the number of iterations $\{1, \cdots, 25\}$ and the y-axis is the gap between iterations $\|\theta_{t+1} - \theta_t\|_2$. The error bars indicate one standard deviation over 10 runs.

6.2 Utility of collective participation

With the same credit-scoring data we investigate collective strategies of misreporting individual features. we provide intuition for alignment and how it impacts the utility gain for the collective and the effect of scaling the strategy.

To interpret the setting, we first look at the effect individual features have on the learner's predictive objectives to intuitively understand the alignment of the competing goals. To this end, let the feature of interest for the collective be feature i. Then, to simulate modifications to this feature, we replace z_i with a target value $\hat{z}_i$, which we sample independently from a standard normal distribution. Subsequently, we retrain a logistic regression classifier, using data in which the respective feature has been misreported. We repeat this experiment for each feature individually. It is intuitive that concealing or altering a particular feature will influence the classifier's performance. Figure 2 shows the resulting accuracy drops relative to a baseline classifier trained without misreporting. A larger accuracy reduction indicates that the concealed feature carries significant predictive information for determining the true labels. Intuitively, substantial accuracy drops imply a misalignment between the agents' incentive to misreport and the learner's objective of minimizing prediction loss. Thus, the magnitude of the performance drop serves as a proxi for the degree of alignment Φ.

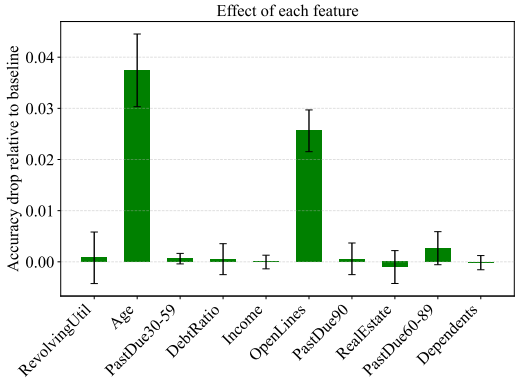


Figure 2: Simulation of the accuracy drop against concealing each feature. The values of the bars indicate the *drop* of test accuracy compared to the baseline classifier. The error bars indicate one standard diviations over 10 different train-test splits.

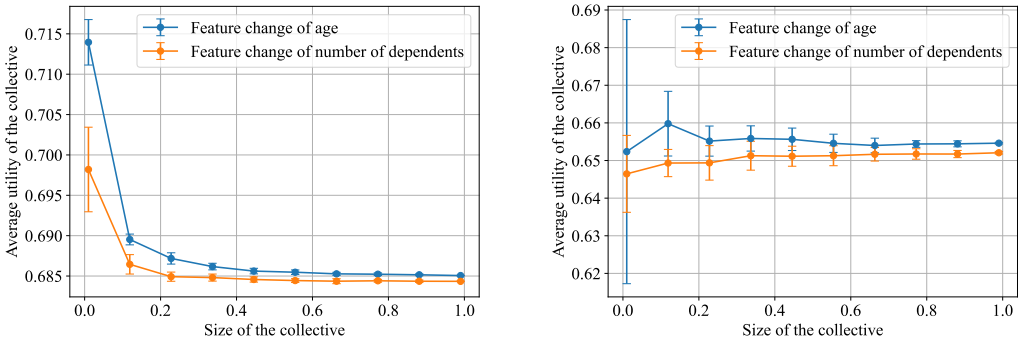
Next, we study a mixed-population setting as in (7) where an α-fraction of agents act collectively while the remaining $1 - \alpha$ report $\mathcal{D}_0$ truthfully. In particular, the agents consider the following objective:

$$u((x, y); \theta) = \mathrm{CE}(\theta^T x, y) - \lambda \cdot \|x - x_0\|^2,$$

where CE denotes the cross-entropy loss and x_0 represents the agent's initial feature vector. We assume the collective implements the optimal size aware-strategy $h^\sharp_\alpha$ under the constraint that the collective can only manipulate a single feature S (defining a constraint strategy space), i.e.,

$$h^\sharp_\alpha = \arg\max_{h \in H_S} \mathbb{E}_{z \sim \mathcal{D}_h}\Big[u\big(z; \mathcal{A}\big(\alpha \mathcal{D}_h + (1-\alpha)\mathcal{D}_0\big)\big)\Big],$$

where H_S denotes the space of strategies for changing feature S. We approximate the optimal strategy using gradient descent with learning rate 0.01 and 250 epochs.



(a) zero-sum case ($\lambda = 0$)

(b) non-zero-sum case ($\lambda = 0.5$)

Figure 3: Collective utility in a mixed population. Each line indicates the manipulation of a feature, either age or number of dependents. Figures show expected utility of the collective for varying collective size α. All the error bars indicate one standard deviation over 10 runs.

Alignment. Since the learner performs logistic regression our setup corresponds to a zero-sum game for $\lambda = 0$. The larger λ the more misaligned the objective are. To reflect these different scenarios we evaluate the expected utility of the collective at equilibrium in the cases of $\lambda = 0$ and $\lambda = 0.5$.

Figure 3 reports the results for the two cases where the feature S is either 'age' or 'number of dependents', where the former was found to be the most important feature and the latter is the least important feature. The figure report the gain of the agents as the collective size α increases. Both curves are decreasing with the increase of collective size. As shown in Figure 2, age is a more important feature than number of dependents. This means, changing the feature related to age leads to larger return for the collective, but at the same time, they face a stronger counterforce by the learner as the collective size increases, leading to a faster drop in the utility gain. Under the non-zero-sum setting (right panel of Figure 3), the collective encounters a much weaker counter-force. After accounting for the regularization term penalizing deviation from the base distribution, the overall utilities are lower than in the zero-sum case. The utility trend becomes flatter, and in some regimes even slightly increasing with collective size. This behavior suggests that, once penalized for large deviations, the collective gains more effective steering power, enabling larger groups to coordinate beneficially despite the penalty for deviating from the base distribution.

7 Conclusion

We introduce look-ahead reasoning as a new perspective on strategic reasoning on learning platforms. While traditional analyses of strategic classification treat users as reacting independently to a fixed model, look-ahead reasoning highlights that users' incentives and actions are inherently interdependent—each agent's actions influence future model deployments, and thus the utility of other agents in the population. Within this broad theme, we find that higher-order reasoning accelerates convergence toward equilibrium but does not improve individuals' long-run outcomes (Theorem 3), suggesting that attempts to "outsmart" others may offer only transient advantages. In contrast, collective reasoning—where users coordinate their behavior through their shared impact on the model—allows the agents to steer the model towards a desirable state. The benefits of coordination, however, may be limited (Theorem 5), and the excessive steering power that comes with larger collectives may harm the collective utility, leading to smaller utility for larger collectives in certain cases (Proposition 7).

Acknowledgements

Celestine Mendler-Dünner acknowledges the financial support of the Hector Foundation.

References

David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, pages 354–363. PMLR, 2018.

Joachim Baumann and Celestine Mendler-Dünner. Algorithmic collective action in recommender systems: Promoting songs by reordering playlists. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Yahav Bechavod, Chara Podimata, Zhiwei Steven Wu, and Juba Ziani. Information discrepancy in strategic learning, 2021.

Omri Ben-Dov, Jake Fawkes, Samira Samadi, and Amartya Sanyal. The role of learning algorithms in collective action. In *International Conference on Machine Learning*, pages 3443–3461. PMLR, 2024.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *International Conference on Artificial Intelligence and Statistics*, volume 151, pages 6045–6061. PMLR, 2022.

Michael Brückner and Tobias Scheffer. Stackelberg games for adversarial prediction problems. In *ACM SIGKDD*, pages 547–555, 2011.

Julie Yujie Chen. Thrown under the bus and outrunning it! the logic of didi and taxi drivers' labour and activism in the on-demand economy. *New Media & Society*, 20(8):2691–2711, 2018.

Yiling Chen, Yang Liu, and Chara Podimata. Learning strategy-aware linear classifiers. *Advances in Neural Information Processing Systems*, 33:15265–15276, 2020.

Jinshuo Dong, Aaron Roth, Zachary Schutzman, Bo Waggoner, and Zhiwei Steven Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 55–70, 2018.

Dmitriy Drusvyatskiy and Lin Xiao. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research*, 48(2):954–998, 2023.

Etienne Gauthier, Francis Bach, and Michael I. Jordan. Statistical collusion by collectives on learning platforms. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pages 18897–18919, 2025.

Ganesh Ghalme, Vineet Nair, Itay Eilat, Inbal Talgam-Cohen, and Nir Rosenfeld. Strategic classification in the dark. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 3672–3681. PMLR, 2021.

Moritz Hardt and Celestine Mendler-Dünner. Performative Prediction: Past and Future. *Statistical Science*, 40(3):417 – 436, 2025.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *ACM Conference on Innovations in Theoretical Computer Science*, page 111–122, 2016.

Moritz Hardt, Meena Jagadeesan, and Celestine Mendler-Dünner. Performative power. In *Advances in Neural Information Processing Systems*, volume 35, pages 22969–22981, 2022.

Moritz Hardt, Eric Mazumdar, Celestine Mendler-Dünner, and Tijana Zrnic. Algorithmic collective action in machine learning. In *International Conference on Machine Learning*, 2023.

Meena Jagadeesan, Celestine Mendler-Dünner, and Moritz Hardt. Alternative microfoundations for strategic classification. In *International Conference on Machine Learning*, volume 139, pages 4687–4697. PMLR, 2021.

Terri Kneeland. Identifying higher-order rationality. *Econometrica*, 83(5):2065–2079, 2015.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. In *Advances in Neural Information Processing Systems*, volume 33, pages 4929–4939, 2020.

Rosemarie Nagel. Unraveling in Guessing Games: An Experimental Study. *American Economic Review*, 85(5):1313–1326, December 1995.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian J. Ratliff. Multiplayer performative prediction: Learning in decision-dependent games. *Journal of Machine Learning Research*, 24(202):1–56, 2023.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, 2020.

Chara Podimata. Incentive-aware machine learning; robustness, fairness, improvement & causality. *arXiv preprint arXiv:2505.05211*, 2025.

Dorothee Sigg, Moritz Hardt, and Celestine Mendler-Dünner. Decline now: A combinatorial model for algorithmic collective action. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, 2025.

Arianna Tassinari and Vincenzo Maccarrone. Riders on the storm: Workplace solidarity among gig economy couriers in italy and the uk. *Work, Employment and Society*, 34(1):35–54, 2020.

Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. Data leverage: A framework for empowering the public in its relationship with technology companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, page 215–227, 2021.

Tijana Zrnic, Eric Mazumdar, Shankar Sastry, and Michael Jordan. Who leads and who follows in strategic classification? *Advances in Neural Information Processing Systems*, 34:15257–15269, 2021.

A Proofs

A.1 Auxiliary results

Lemma 9. *Let $\mathcal{W}$ denote the Wasserstein-1 distance, and $\sum_{i=1}^n \alpha_i \geq 0$ with $\alpha_i \geq 0$, then*

$$\mathcal{W}\left(\sum_{i=1}^n \alpha_i \mathcal{D}_i, \sum_{i=1}^n \alpha_i \mathcal{D}'_i\right) \leq \sum_{i=1}^n \alpha_i \mathcal{W}(\mathcal{D}_i, \mathcal{D}'_i).$$

Proof. By definition, the Wasserstein-1 distance could be rewrited as

$$\min_{\mu(X,Y)} \mathbb{E}_{\mu(X,Y)} \left[|X - Y|\right],$$

where μ is the joint distribution of X, Y and $X \sim \sum_{i=1}^n \alpha_i \mathcal{D}_i$, $Y \sim \sum_{i=1}^n \alpha_i \mathcal{D}'_i$. Then, consider the measures μ_i defined as

$$\mu_i = \min_{\mu(X_i,Y_i)} \mathbb{E}_{\mu(X_i,Y_i)} \left[|X_i - Y_i|\right],$$

where $X_i \sim \mathcal{D}_i$, $Y_i \sim \mathcal{D}'_i$. Then, with $\hat{\mu} = \sum_i \alpha_i \mu_i$, we can notice that

$$\mathcal{W}\left(\sum_{i=1}^n \alpha_i \mathcal{D}_i, \sum_{i=1}^n \alpha_i \mathcal{D}'_i\right) \leq \mathbb{E}_{\hat{\mu}} \left[|X - Y|\right] = \sum_{i=1}^n \alpha_i \mathbb{E}_{\mu_i} \left[|X - Y|\right] = \sum_{i=1}^n \alpha_i \mathcal{W}(\mathcal{D}_i, \mathcal{D}'_i),$$

where the first inequality follows from the minimisation property of the Wasserstein-1 distance. □

A.2 Proof of Theorem 3

The key step is to prove Lemma 10 below. The claim in the theorem follows by combining Lemma 10 with Theorem 3.5 in Perdomo et al. [2020],

Lemma 10. *Let α_k be the portion of the population with cognitive level k. Then, suppose ℓ is γ-strongly convex and β-smooth in z, θ, and that the distribution map $\overline{\mathcal{D}}(\theta) := \sum_{k=0}^\infty \alpha_k \mathcal{D}_k(\theta)$, then the sensitivity of $\overline{\mathcal{D}}$ is $\sum_{k=1}^\infty \alpha_k \left(\frac{\epsilon\beta}{\gamma}\right)^{k-1} \epsilon$.*

Proof. Denote $\theta^k := \mathcal{A}(\mathcal{D}_1(\theta^{k-1}))$ and $\theta^0 = \theta$. Similarly, $\phi^k := \mathcal{A}(\mathcal{D}_1(\phi^{k-1}))$ and $\phi^0 = \phi$. From Equation (2), we can see that $\mathcal{D}_k(\theta^0) = \mathcal{D}_1(\theta^{k-1})$. Consider the map $\mathcal{D}_k$, we have the recurrence:

$$\mathcal{W}(\mathcal{D}_k(\theta), \mathcal{D}_k(\phi)) = \mathcal{W}(\mathcal{D}_1(\theta^{k-1}), \mathcal{D}_1(\phi^{k-1})) \leq \epsilon \left\|\theta^{k-1} - \phi^{k-1}\right\|_2.$$

By Perdomo et al. [2020], Theorem 3.5, we also have

$$\left\|\theta^{k-1} - \phi^{k-1}\right\|_2 \leq \left(\frac{\epsilon\beta}{\gamma}\right)^{k-1} \|\theta - \phi\|_2.$$

Then, we notice that

$$\mathcal{W}\left(\sum_{k=0}^\infty \alpha_k \mathcal{D}_k(\theta), \sum_{k=0}^\infty \alpha_k \mathcal{D}_k(\phi)\right) \leq \sum_{k=0}^\infty \alpha_k \mathcal{W}(\mathcal{D}_k(\theta), \mathcal{D}_k(\phi)) \leq \sum_{k=1}^\infty \alpha_k \left(\frac{\epsilon\beta}{\gamma}\right)^{k-1} \epsilon \|\theta - \phi\|_2,$$

where the first inequality follows from Lemma 9 and the $k = 0$ terms could be dropped since both $\mathcal{D}_0(\theta) = \mathcal{D}_0(\phi) = \mathcal{D}_0$. □

A.3 Proof of Corollary 1

We start from Theorem 3. Define the contraction factor

$$\rho_\alpha = \left(\sum_{k=1}^\infty \left(\frac{\epsilon\beta}{\gamma}\right)^k \alpha_k\right).$$

It can be seen that if ρ_α is positive, it holds for any α such that $\sum_{k=1}^\infty \alpha_k = 1$. Similarly, if it is zero, this holds for any α. Thus, a simple contraction argument shows that the trajectory converges to the same stable point independent of α. The same holds for the special case $\alpha_k = 1$. At this point, no agent is moving and thus the equilirbium strategies $h_{\theta^*}^{(k)}$ are identical. So is the utility:

$$h_{\theta^*}^{(k)} = h_{\theta^*}^{(k')} \quad \Rightarrow \quad U(h_{\theta^*}^{(k)}) = U(h_{\theta^*}^{(k')})$$

A.4 Proof of Proposition 4

Since (h^*, θ^*) is the performative stable point. By definition, θ^*, it is clear that

$$\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta \ell(z,\theta^*)\right] = 0.$$

Therefore, since $u = c\cdot\ell$, we can conclude that $\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta u(z,\theta^*)\right] = c\cdot\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta \ell(z,\theta^*)\right] = 0$. Then, by Theorem 11 below, we have $B = 0$.

A.5 Generalized version of Theorem 5

We first state a general version of Theorem 5 without the use linearity assumption 1.

Theorem 11 (Generalization of Theorem 5)**.** *Suppose $U(h)$ is γ-strongly concave. Then, we have*

$$0 \le B \le \frac{1}{2\gamma}\cdot\left\|\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_h\nabla_\theta\ell(z,\theta^*)\right](\mathbf{H}^\star)^{-1}\mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[\nabla_\theta u(z,\theta^*)\right]\right\|_2^2.$$

where θ^* denotes the stable point corresponding to the selfish strategy h^* and $\theta^\sharp$ denotes the stable point corresponding to the strategy $h^\sharp = \arg\max_h U(h)$.

Proof. For the simplicity of notations, we set $f(h,\theta) := \mathbb{E}_{z\sim\mathcal{D}_h}\left[u(z;\theta)\right]$ and $g(h,\theta) := \mathbb{E}_{z\sim\mathcal{D}_h}\left[\ell(z;\theta)\right]$. Recall that

$$U(h) = f(h,\mathcal{A}(\mathcal{D}_h)) = \mathbb{E}_{z\sim\mathcal{D}_h}\left[u(z,\mathcal{A}(\mathcal{D}_h))\right],$$

where the first argument in f only applies in distribution that the distribution is taken against and the second argument is corresponding to the second argument of u. Then, by the implicit function theorem we have

$$\nabla_h U(h) = \nabla_h f(h,\mathcal{A}(\mathcal{D}_h)) = \nabla_1 f(h,\mathcal{A}(\mathcal{D}_h)) - \nabla^2_{1,2}g(h,\mathcal{A}(\mathcal{D}_h))\left[\nabla^2_{2,2}g(h,\mathcal{A}(\mathcal{D}_h))\right]^{-1}\nabla_2 f(h,\mathcal{A}(\mathcal{D}_h)),$$

where ∇_1 and ∇_2 denote the gradient operator on the first/second argument of the function. For the NE (h^*,θ^*), we must have $\nabla_1 f(h,\mathcal{A}(\mathcal{D}_h)) = 0$. Hence, by the PL-inequality, we could obtain that

$$f\left(h^\sharp\right) - f\left(h^*\right) \le \frac{1}{2\gamma}\cdot\left\|\nabla^2_{1,2}g(h^*,\mathcal{A}(\mathcal{D}_{h^*}))\left[\nabla^2_{2,2}g(h^*,\mathcal{A}(\mathcal{D}_{h^*}))\right]^{-1}\nabla_2 f(h^*,\mathcal{A}(\mathcal{D}_{h^*}))\right\|_2^2.$$

□

A.6 Proof of Theorem 5

For notional simplicity, we use h in replace of the parameterisation $\eta(h)$ whenever the usage is clear. We consider a hypothetical mixture population which is represented as $\alpha\mathcal{D}_{h^\sharp} + (1-\alpha)\mathcal{D}_{h^*}$. Note that (h^*,θ^*) is the stable point such that

$$h^* = \arg\max_h \mathbb{E}_{z\sim\mathcal{D}_h}\left[u(z,\theta^*)\right],$$
$$\theta^* = \arg\min_\theta \mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\ell(z,\theta)\right].$$

Therefore, for the mixture population, $\alpha = 0$ indicates the equilibrium of selfish action. It means that, if we fix the learner's output is θ^* of this case, the action $h_\alpha = \alpha h^\sharp + (1-\alpha)h^*$ will maximize the population's expected utility only when $\alpha = 0$. Formally, consider the function $\iota(\alpha) = \mathbb{E}_{z\sim\mathcal{D}_{\alpha h^\sharp+(1-\alpha)h^*}}\left[u(z,\theta^*)\right] = \mathbb{E}_{z\sim\alpha\mathcal{D}_{h^\sharp}+(1-\alpha)\mathcal{D}_{h^*}}\left[u(z,\theta^*)\right]$, we must have

$$\left.\frac{\partial\iota}{\partial\alpha}\right|_{\alpha=0} = \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[u(z,\theta^*)\right] - \mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[u(z,\theta^*)\right] = 0.$$

Also at the stable point (h^*,θ^*), we have the fact that $\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta\ell(z,\theta^*)\right] = 0$. Then, consider the function $e(\alpha) = U\left(\alpha h^\sharp + (1-\alpha)h^*\right)$, we have

$$\begin{aligned}
\left.\frac{\partial e}{\partial\alpha}\right|_{\alpha=0} &= \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[u(z,\theta^*)\right] - \mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[u(z,\theta^*)\right] \\
&\quad - \left\langle\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta u(z,\theta^*)\right], \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[\nabla_\theta\ell(z,\theta^*)\right] - \mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta\ell(z,\theta^*)\right]\right\rangle_{(\mathbf{H}^\star)^{-1}} \\
&= -\left\langle\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta u(z,\theta^*)\right], \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[\nabla_\theta\ell(z,\theta^*)\right] - \mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta\ell(z,\theta^*)\right]\right\rangle_{(\mathbf{H}^\star)^{-1}} \\
&= -\left\langle\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_\theta u(z,\theta^*)\right], \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[\nabla_\theta\ell(z,\theta^*)\right]\right\rangle_{(\mathbf{H}^\star)^{-1}}.
\end{aligned}$$

Then, the result follows from the PL-inequality. In the following, we state a more general version of Theorem 5 which does not rely on Assumption 1. One could see that PS is still governed by the alignment between the gradient of utility and the Jacobian term $\mathbb{E}_{z\sim\mathcal{D}_{h^*}}\left[\nabla_h\nabla_\theta\ell(z,\theta^*)\right]$.

A.7 Proof of Proposition 6

By Lemma 10, the sensitivity of the mixture distribution could be computed as

$$\sum_{k=1}^{\infty} \alpha_k \left(\frac{\epsilon\beta}{\gamma}\right)^{k-1} \epsilon = (1-\alpha)\epsilon,$$

where only $\alpha_1 = 1-\alpha$ and $\alpha_0 = \alpha$. Combining this with Theorem 3.5 in Perdomo et al. [2020] will yield the result.

A.8 Proof of Proposition 7

Consider the derivative of U_α with respect to variable α,

$$\frac{\partial U_\alpha}{\partial \alpha} = \frac{\partial \mathbb{E}_{z\sim\mathcal{D}_h}\left[u(z;\theta^*_\alpha)\right]}{\partial \alpha} = \frac{\partial \mathbb{E}_{z\sim\mathcal{D}_h}\left[u(z;\theta^*_\alpha)\right]}{\partial \theta} \cdot \frac{\partial \theta^*_\alpha}{\partial \alpha},$$

Next, we notice that

$$\nabla_\theta \mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right] = 0,$$

where we can further write the LHS as

$$\nabla_\theta \mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right] = \alpha \cdot \mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right] + (1-\alpha)\cdot \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right] = 0. \quad (10)$$

Therefore, $\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right] = -\frac{\alpha}{1-\alpha}\cdot \mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right]$. Then, we consider the term $\frac{\partial \theta^*_\alpha}{\partial \alpha}$. By implicit function theorem, with $\alpha > 0$, we have

$$\begin{aligned}
\frac{\partial \theta^*_\alpha}{\partial \alpha} &= -\left(\nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right]\right)^{-1}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right]\right) \\
&= -\frac{1}{1-\alpha}\left(\nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right]\right)^{-1}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right]\right).
\end{aligned}$$

Hence, we could finally write

$$\frac{\partial U_\alpha}{\partial \alpha} = -\frac{1}{1-\alpha}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta u(z;\theta^*_\alpha)\right]\right)^T\left(\nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right]\right)^{-1}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right]\right).$$

A.9 Proof of Proposition 8

Essentially, the proof is the same as the proof of Proposition 7 up to a use of an envelop theorem. For completeness, we restate the proof here and make it clear the usage of the envelop theorem. For notational simplicity, we abbreviate $h^\sharp_\alpha$ as h. Consider the derivative of U^*_α with respect to variable α,

$$\frac{\partial U^*_\alpha}{\partial \alpha} = \frac{\partial \mathbb{E}_{z\sim\mathcal{D}_h}\left[u(z;\theta^*_\alpha)\right]}{\partial \alpha} = \frac{\partial \mathbb{E}_{z\sim\mathcal{D}_h}\left[u(z;\theta^*_\alpha)\right]}{\partial \theta} \cdot \frac{\partial \theta^*_\alpha}{\partial \alpha},$$

where the second equality follows from the implicit function theorem and *the envelop theorem*. Next, we notice that

$$\nabla_\theta \mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right] = 0,$$

where we can further write the LHS as

$$\nabla_\theta \mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right] = \alpha \cdot \mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right] + (1-\alpha)\cdot \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right] = 0. \quad (11)$$

Therefore, $\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right] = -\frac{\alpha}{1-\alpha}\cdot \mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right]$. Then, we consider the term $\frac{\partial \theta^*_\alpha}{\partial \alpha}$. By implicit function theorem, with $\alpha > 0$, we have

$$\begin{aligned}
\frac{\partial \theta^*_\alpha}{\partial \alpha} &= -\left(\nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right]\right)^{-1}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right]\right) \\
&= -\frac{1}{1-\alpha}\left(\nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right]\right)^{-1}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right]\right).
\end{aligned}$$

Hence, we could finally write

$$\frac{\partial U^*_\alpha}{\partial \alpha} = -\frac{1}{1-\alpha}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta u(z;\theta^*_\alpha)\right]\right)^T\left(\nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right]\right)^{-1}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right]\right).$$

B Additional results

B.1 Benefit of participation

We contrast the utility of participating agents with those that opt out, assuming the mixture model (7). Here the base distribution $\mathcal{D}_0$ will play in and determine the utility of non participating agents in comparison to those that participate.

Proposition 12 (Benefit of participation)**.** *For any $\alpha \in (0,1)$ and fixed strategy h, and let θ^*_α be the resulting stable point. Then, the cost of participation*

$$C_h(\alpha) := U_\alpha - \mathbb{E}_{z\sim\mathcal{D}_0}\left[u(z;\theta^*_\alpha)\right].$$

changes with the collective size as follows

$$\frac{\partial C_h(\alpha)}{\partial \alpha} = -\big\langle \mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta u(z;\theta^*_\alpha)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta u(z;\theta^*_\alpha)\right], \\ \mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta^*_\alpha)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta^*_\alpha)\right]\big\rangle_{\mathbf{H}^{-1}}, \tag{12}$$

where

$$\mathbf{H} := \nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right].$$

The above result indicates that the change in the benefit of participation also depends on a notion of alignment. Once the utilities are misaligned, as the benefit of participation decreases, it is harder to form a collective implementing some certain strategy h. Conversely, aligned utilities makes the implementation strategy h more appealing when the collective is growing. In particular, the RHS of Equation (12) is negative whenever the expected gradient of utility and loss change in a similar direction when considering the distributional shift from $\mathcal{D}_0$ to $\mathcal{D}_h$.

Finally, we uncover an interesting fact that, in the totally aligned (potential game) or misaligned (zero-sum game) cases. The utility of the population will not change by implementing a fixed strategy. It means that the incentive structure remains the same independent of the size of the collective.

Corollary 2. *For any $\alpha \in (0,1)$ and fixed strategy h, and let θ^*_α be the resulting stable point. Assume $u = k \cdot \ell$ for some constant $k \neq 0$. Then,*

$$\frac{\partial C_h}{\partial \alpha} = 0, \quad \forall \alpha \in (0,1).$$

Once the alignment/misalignment of the utilities goes to the extreme, the benefit of participation degenerates to 0. Intuitively, this is caused by the fact that the alignment maximize of the sensitivity learner. Then, the change to a fixed strategy could be fully captured by the learner.

Proof of Proposition 12. Similar to the proof of Proposition 8, we have

$$\frac{\partial \mathbb{E}_{z\sim\mathcal{D}_0}\left[u(h_\alpha(z);\theta^*_\alpha)\right]}{\partial \alpha} = -\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta u(z;\theta^*_\alpha)\right]\right)^T \left(\nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right]\right)^{-1} \\ \left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right]\right).$$

Using the chain rule, and implicit theorem again, we also have

$$\frac{\partial \mathbb{E}_{z\sim\mathcal{D}_0}\left[u(z;\theta^*_\alpha)\right]}{\partial \alpha} = -\left(\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta u(z;\theta^*_\alpha)\right]\right)^T \left(\nabla^2_{\theta,\theta}\mathbb{E}_{z\sim\alpha\mathcal{D}_h+(1-\alpha)\mathcal{D}_0}\left[\ell(z;\theta)\right]\right)^{-1} \\ \left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right]\right).$$

In summary, taking the derivative through the α dependence of the strategy we have

$$\frac{\partial B_\alpha}{\partial \alpha} = -\left\langle \mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta u(z;\theta^*_\alpha)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta u(z;\theta^*_\alpha)\right], \mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla_\theta \ell(z;\theta)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla_\theta \ell(z;\theta)\right]\right\rangle_{\mathbf{H}^{-1}},$$

hence the result above holds for any strategy h_α. And it is clear from the definition of h_α that $B_0 \geq 0$.

Proof of Corollary 2. First, by Equation (12), if one have $u = k \cdot l$ with $k > 0$. It turns out that $\frac{\partial C_h}{\partial \alpha} \geq 0$ always holds. Conversely, if one considers the mixture $(1-\alpha)\mathcal{D}_h + \alpha\mathcal{D}_0$, and consider the same derivative again. It is still nonnegative. Therefore, both increasing or decreasing of α will yield increasing of C_h. Hence, the only possible case is the mixed utility remains unchanged.

B.2 Altruistic objectives

Next, we explore the amount of shift necessary to move a model. Again, consider the mixture model (7).

Assume the collective aims to steer the model parameters to a target state θ_{target}. This can for example be an altruistic objective of optimizing the utility under $\mathcal{D}_0$.

The next result provides a lower bound on the amount of shift necessary to move the model. It illustrates that smaller collectives either have smaller influence, or they need to invest more effort to steer the learner to a target state. Note that the required shift is entirely a property of the learner's loss function and the suboptimality of the target state you aim to reach under $\mathcal{D}_0$.

Proposition 13 (Collective shift)**.** *Suppose ℓ is β-smooth and the collective aims to reach a state θ_{target} with strategy h. Then, the amount of distribution shift necessary is at least*

$$\mathcal{W}(\mathcal{D}_h, \mathcal{D}_0) \geq \frac{1}{\alpha \cdot \beta} \cdot \left\| \mathbb{E}_{z \sim \mathcal{D}_0} \left[\nabla_\theta \ell(z; \theta_{\text{target}}) \right] \right\|_2 .$$

Proof of Proposition 13 By the fact that optimizing h must induce the learner strategy θ_{target}. Moreover, for any α, this observation also must hold. Analogous to Lemma 3, the optimality condition of θ_{target} implies

$$\mathbb{E}_{z \sim \mathcal{D}_h} \left[\nabla_\theta \ell(z; \theta_{\text{target}}) \right] = -\frac{1-\alpha}{\alpha} \cdot \mathbb{E}_{z \sim \mathcal{D}_0} \left[\nabla_\theta \ell(z; \theta_{\text{target}}) \right].$$

We note that the RHS is a fixed constant for any α. Then, for any $\|\mathbf{v}\|_2 = 1$, we have

$$\left(\mathbb{E}_{z \sim \mathcal{D}_h} \left[\nabla_\theta \ell(z; \theta_{\text{target}}) \right] - \mathbb{E}_{z \sim \mathcal{D}_0} \left[\nabla_\theta \ell(z; \theta_{\text{target}}) \right] \right)^T \mathbf{v} \leq \beta \cdot \mathcal{W}(\mathcal{D}_h, \mathcal{D}_0).$$

Combining the above, we have

$$-\frac{1}{\alpha} \cdot \mathbb{E}_{z \sim \mathcal{D}_0} \left[\nabla_\theta \ell(z; \theta_{\text{target}}) \right]^T \mathbf{v} \leq \beta \cdot \mathcal{W}(\mathcal{D}_h, \mathcal{D}_0).$$

Take $\mathbf{v} = \frac{-\frac{1}{\alpha} \cdot \mathbb{E}_{z \sim \mathcal{D}_0} [\nabla_\theta \ell(z; \theta_{\text{target}})]}{\left\| -\frac{1}{\alpha} \cdot \mathbb{E}_{z \sim \mathcal{D}_0} [\nabla_\theta \ell(z; \theta_{\text{target}})] \right\|_2}$, we have

$$\left\| \mathbb{E}_{z \sim \mathcal{D}_0} \left[\nabla_\theta \ell(z; \theta_{\text{target}}) \right] \right\|_2 \leq \alpha \cdot \beta \cdot \mathcal{W}(\mathcal{D}_h, \mathcal{D}_0).$$

B.3 Selfish agents meet collectives

In the following result we characterize how this hurts the collective and agents engaging in collective reasoning only partially reach their goal if some agents act selfishly.

Theorem 14. *Consider the mixed population according to* (5) *with $D(\theta) = D_1(\theta)$ and let θ^*_α denote the stable point for a fixed α. Further, suppose $U(h)$ is γ-strongly concave in h and Assumption 1 holds. Then, the utility loss due to partial participation can be bounded as*

$$U_1 - U_\alpha \leq \frac{(1-\alpha)^2}{2\gamma} \cdot \left(\left\langle \mathbb{E}_{z \sim \mathcal{D}_\alpha(\theta^*_\alpha)} \left[\nabla_\theta u(z, \theta^*_\alpha) \right], \mathbb{E}_{z \sim \mathcal{D}_{h^\sharp}} \left[\nabla_\theta \ell(z, \theta^*_\alpha) \right] \right\rangle_{(\mathbf{H}^\star)^{-1}} \right)^2,$$

where

$$\mathbf{H}^* = \nabla^2_\theta \mathbb{E}_{z \sim \mathcal{D}_\alpha(\theta^*_\alpha)} \left[\nabla_\theta \ell(z, \theta^*_\alpha) \right]$$

The above result explains that the utility loss is proportional to $1 - \alpha$. Once there are more agents joining the collective, the overall utility will approach optimality. With more agents joining the collective, the expected utility across the agents in the mixture distribution will be improved. Note that this is different to Section 5.2 where we considered the utility of participants only.

Proof of Theorem 14 First, for the function $\iota(\beta) = \mathbb{E}_{z \sim \beta \mathcal{D}_{h^\sharp} + (1-\beta)\mathcal{D}_{h^*_\alpha}} [u(z; \theta^*_\alpha)]$, by looking at the derivative of β at $\beta = 0$. ι is maximised at $\beta = 0$ and hence $\mathbb{E}_{z \sim \mathcal{D}_{h^\sharp}} [u(z; \theta^*_\alpha)] = \mathbb{E}_{z \sim \mathcal{D}_{h^*_\alpha}} [u(z; \theta^*_\alpha)]$. Then, consider another auxiliary function:

$$e(\beta) = \mathbb{E}_{z \sim \beta \mathcal{D}_{h^\sharp} + (1-\beta)\mathcal{D}_{h^*_\alpha}} \left[u(z, \theta^*_\beta) \right],$$

where $\theta^*_\beta = \mathcal{A}(\beta\mathcal{D}_{h^\sharp} + (1-\beta)\mathcal{D}_{h^*_\alpha})$. Then, we consider

$$\begin{aligned}\frac{\partial e}{\partial \beta}\Big|_{\beta=\alpha} &= -\left\langle \mathbb{E}_{z\sim\mathcal{D}_\alpha}\left[\nabla_\theta u(z,\theta^*_\alpha)\right], \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[\nabla_\theta \ell(z,\theta^*_\alpha)\right]\right\rangle_{(\mathbf{H}^\star)^{-1}} \\ &= -(1-\alpha)\left\langle \mathbb{E}_{z\sim\mathcal{D}_\alpha}\left[\nabla_\theta u(z,\theta^*_\alpha)\right], \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp}}\left[\nabla_\theta \ell(z,\theta^*_{\alpha'=1})\right]\right\rangle_{(\mathbf{H}^\star)^{-1}},\end{aligned}$$

where the last line follows from applying Lemma 15 below. Again, the result follows from the PL inequality.

Lemma 15. *For any $\alpha \in [0,1]$, and some fixed h, we must have*

$$\frac{\partial \frac{1}{\alpha}\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right]}{\partial\alpha} = 0,$$

which implies that $\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right] = \alpha\cdot\mathbf{v}$ for some $\mathbf{v}$ depending on h.

Proof. By direct calculation, we have

$$\begin{aligned}&\frac{\partial \frac{1}{\alpha}\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right]}{\partial\alpha} \\ &= -\frac{1}{\alpha^2}\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right] + \frac{1}{\alpha}\frac{\partial \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right]}{\partial\alpha} \\ &= -\frac{1}{\alpha^2}\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right] \\ &\quad -\frac{1}{\alpha}\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla^2\ell(z;\theta^*_\alpha)\right]\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla^2\ell(z;\theta^*_\alpha)\right]^{-1}\left(\mathbb{E}_{z\sim\mathcal{D}_h}\left[\nabla\ell(z;\theta^*_\alpha)\right] - \mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right]\right) \\ &= -\frac{1}{\alpha^2}\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right] - \frac{1}{\alpha}\left(-\frac{1}{\alpha}\mathbb{E}_{z\sim\mathcal{D}_0}\left[\nabla\ell(z;\theta^*_\alpha)\right]\right) = 0,\end{aligned}$$

where the third line follows from the implicit function theorem and the last line follows from the identity eq. (11). □

C Experiments

C.1 Additional simulation on the utilities of individual k-level agents

We also compare the utilities of agents with different cognitive level with the same data setup as Section 6. The utilities are calculated following Equation (9). The simulation is run by the choice of parameter $(\alpha_0, \alpha_1, \alpha_2) = (0, 0.5, 0.5)$ and $\epsilon = 0.5$. (See Section 6.1 for more detailed setups of the experiment). We report the results in Figure 4 where the y-axis represents $u^1_\epsilon, u^2_\epsilon$ where we can see that the differences are decreasing with error bars breaking across 0. Such observations align with our theoretical finding outlined in Corollary 1, where we showed that agents with different levels will finally converge to a stable point with the same utility level. Besides, we note that there is no evidence of any monotonicity of the individual utilities against their cognitive levels.

C.2 Alignment metric in the simulation of Section 6.2

Under the same setup as Section 6.2, we also report the following alignment metric in the zero-sum case:

$$\left\langle \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp_\alpha}}\left[\nabla_\theta u(z;\theta^*_\alpha)\right], \mathbb{E}_{z\sim\mathcal{D}_{h^\sharp_\alpha}}\left[\nabla_\theta \ell(z;\theta^*_\alpha)\right]\right\rangle_{\mathbf{H}^{-1}}$$

as defined in Proposition 7. One can observe that the alignment metric, as the indicator of the counter-force experienced by the collective, have similar pattern as Figure 3 (a), which verifies our theoretical observations in Proposition 7.

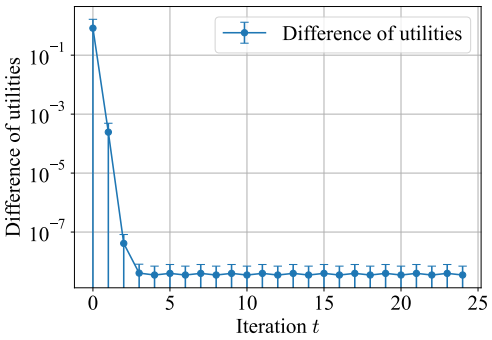


Figure 4: Differences of utilities between level-1 and level-2 agents. The error bars indicate one standard deviation over 10 different selections of individuals.

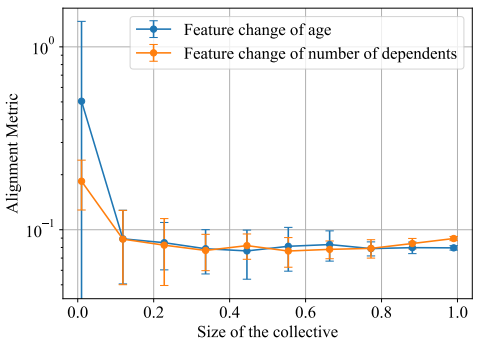


Figure 5: The alignment metrics (as defined in Proposition 7) versus different sizes of the collective. The error bars represent one standard deviation over 10 runs.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The abstract and introduction accurately reflect the scope and contributions.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: Limitations are discussed where necessary.

   Guidelines:

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

3. **Theory assumptions and proofs**

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

   Answer: [Yes]

Justification: The proofs are complete. They are either in the main text or the supplementary material.

Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: The source code of the experiments are included in the supplementary material.

   Guidelines:

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

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

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

Answer: [Yes]

Justification: Data is available publicly/randomly generated (in the source code). The code is in the supplementary material.

Guidelines:

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

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: The setup is discussed in the simulation section. The implementation is in the supplementary material.

   Guidelines:

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

7. **Experiment statistical significance**

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

   Answer: [Yes]

   Justification: Error bars and other setup info are discussed in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: All simulations in this paper are executable with a standard personal laptop.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: [NA]

   Guidelines:

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

10. **Broader impacts**

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

    Answer: [Yes]

    Justification: Our paper mainly focuses on the theoretical contributions. The societal implications/explanations of the results are discussed in the paper.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

    Answer: [NA]

    Justification: [NA]

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: The data used in the experimental section is open for public and research use.

    Guidelines:

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

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

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

    Answer: [NA]

    Justification: [NA]

    Guidelines:

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

14. **Crowdsourcing and research with human subjects**

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

    Answer: [NA]

    Justification: No experiments involving human participants are conducted.

    Guidelines:

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

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

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

    Answer: [NA]

    Justification: No experiments involving human participants are conducted.

    Guidelines:

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

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLM is only used for wording and grammar checking.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.