Dynamics of Learning under User Choice: Overspecialization and Peer-Model Probing

Adhyyan Narang[1] **Sarah Dean**[2] **Lillian J. Ratliff**[1] **Maryam Fazel**[1]

[1]Electrical and Computer Engineering, University of Washington
[2]Computer Science, Cornell University

Abstract

In many economically relevant contexts where machine learning is deployed, multiple platforms obtain data from the same pool of users, each of whom selects the platform that best serves them. Prior work in this setting focuses exclusively on the “local” losses of learners on the distribution of data that they observe. We find that there exist instances where learners who use existing algorithms almost surely converge to models with arbitrarily poor global performance, even when models with low full-population loss exist. This happens through a feedback-induced mechanism, which we call the overspecialization trap: as learners optimize for users who already prefer them, they become less attractive to users outside this base, which further restricts the data they observe. Inspired by the recent use of knowledge distillation in modern ML, we propose an algorithm that allows learners to "probe" the predictions of peer models, enabling them to learn about users who do not select them. Our analysis characterizes when probing succeeds: this procedure converges almost surely to a stationary point with bounded full-population risk when probing sources are sufficiently informative, e.g., a known market leader or a majority of peers with good global performance. We verify our findings with semi-synthetic experiments on the MovieLens, Census, and Amazon Sentiment datasets.[a]

[a]Code for this paper is available at: `https://github.com/AdhyyanNarang/overspecialization-probing`.

1 INTRODUCTION

Traditional supervised learning theory typically assumes a single learner observing data drawn from a fixed distribution. However, this assumption is increasingly violated in modern machine learning markets, such as recommendation platforms and large language model (LLM) services. In these ecosystems, multiple learners operate on the same pool of users, and data is not assigned randomly. Instead, users choose which platform to engage with based on how well that platform serves their specific needs or preferences. Consequently, the data distribution observed by a learner is a function of the learner's own performance and the choices available in the market. This setting is increasingly garnering interest in the machine learning community [Dean et al., 2024, Su and Dean, 2024, Ginart et al., 2021, Bose et al., 2024, Shekhtman and Dean, 2024].

This coupling between model performance and user selection creates a feedback loop. As a learner optimizes for its current user base, it becomes increasingly specialized to that subpopulation. While this minimizes "local" loss on observed users, it often degrades performance on the unobserved population, a phenomenon we term *overspecialization*. Once a learner is overspecialized, it gets caught in an informational trap: it cannot learn to serve new users because it never observes them, and it never observes them because it cannot serve them. At a societal level, this dynamic fuels the formation of algorithmic echo chambers [Cinus et al., 2022, Beardow, 2021, Jiang et al., 2021, Interian et al., 2022], where platforms fragment the population rather than learning a robust, globally capable model.

Independently, another trend has become relevant in modern machine learning systems that has implications for the overspecialization problem: techniques such as knowledge distillation and training on synthetic data are becoming ubiquitous, particularly in the training of Large Language Models [Hinton et al., 2015, Werner, 2025]. While these methods are typically employed to improve reasoning capabilities or computational efficiency (through compression of data),

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:4733–4772.

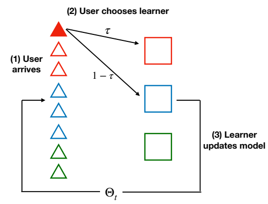

Figure 1: Illustration of our online multi-learner problem setting. The borders of users represent their highest ranked learner $\pi(z)$. For further details, see Section 2.

they introduce a structural change to the learning dynamic. Models are no longer limited to learning from organic user data, but can also "probe" other models to acquire synthetic labels. This enables learners to observe signals outside their siloed data distributions. In this work, we study whether the use of probing mechanisms in these machine learning markets can mitigate overspecialization.

Our Contributions. We model a market where users select learners based on a combination of inherent preferences and predictive loss. We analyze the resulting dynamics through a game-theoretic lens to understand the impact of peer probing. Our main findings are as follows:

1. **The Failure of Standard Learning:** We first analyze standard Multi-learner Streaming Gradient Descent (MSGD) in the absence of probing [Su and Dean, 2024]. We prove that due to the user selection mechanism, MSGD can converge to "bad" stationary points. In these equilibria, learners become overspecialized, achieving low loss on their niche but arbitrarily poor performance on the global population.
2. **Convergence of Peer Probing:** We propose a new algorithm, **MSGD with Probing (MSGD-P)**, where learners mix gradient updates from organic users with updates from pseudo-labeled queries sent to peer models. We prove that this multi-agent dynamic converges to a stationary point of a modified potential function (Theorem 3). To our knowledge, we are the first to observe and study the multi-agent dynamics that arise from synthetic data training.
3. **Restoring Global Competence:** We show that probing effectively mitigates overspecialization. By characterizing the stationary points of MSGD-P, we derive bounds on the full-population loss under appropriate informational conditions (e.g., probing a known market leader vs. aggregating diverse peers).
4. **Empirical Validation:** We validate our findings on the MovieLens, US Census and Amazon Sentiment datasets. We observe that while standard learning leaves some models trapped with poor accuracy, introducing peer probing closes this performance gap.

1.1 RELATED WORK

We study multi-learner markets where users choose among platforms based on inherent preferences and predictive quality, and analyze peer-model probing as a mitigation for overspecialization.

Performative Prediction. Our setting is an instance of performative prediction [Perdomo et al., 2020, Hardt et al., 2016, Miller et al., 2021], where deploying a model influences the data a learner subsequently observes. Multi-learner extensions [Piliouras and Yu, 2023, Narang et al., 2023, Li et al., 2022, Wang et al., 2023, Zhu et al., 2023, Góis et al., 2025, Wang et al., 2025] and gradient-based learning in games [Fiez et al., 2021] provide tools for endogenous distribution shift and strategic dynamics. We specialize to user-choice, where shift arises from selection rather than manipulation, enabling characterization of overspecialization and peer-model probing.

Learning under User Agency. Prior work on user agency focuses on single-learner settings where users opt out or selectively provide data [Hashimoto et al., 2018, Zhang et al., 2019, James et al., 2023, Ben-Porat and Tennenholtz, 2017, Cherapanamjeri et al., 2023, Harris et al., 2023, Raab et al., 2024] or sample-efficient interaction [Narang et al., 2024a,b]; these analyses do not capture our inter-learner feedback dynamics. In the multi-learner setting, Ginart et al. [2021] characterize the *existence* of performance gaps due to competition via batch retraining; we prove that streaming dynamics *converge to* overspecialized equilibria (Theorem 2) and propose probing as a mitigation. Most closely related, Dean et al. [2024] and Su and Dean [2024] study gradient-based dynamics in choice-driven settings; we extend their framework to analyze full-population risk and introduce peer probing. See Appendix A for detailed comparisons with Ginart et al. [2021], Kwon et al. [2022], Bose et al. [2024], and Su and Dean [2024].

Knowledge Distillation. Our probing mechanism draws on knowledge distillation [Hinton et al., 2015] and self-training [Scudder, 1965, Yarowsky, 1995]; unlike Deep Mutual Learning [Zhang et al., 2018] and codistillation [Anil et al., 2018], which assume shared data, or peer-predictive self-training [Feng et al., 2026], which lacks endogenous user-choice allocation, our setting couples distillation with user-driven selection, yielding novel multi-agent dynamics. This also distinguishes our setting from domain generalization methods, which train a centralized model from labeled source domains, and selective-label/sample-selection work, which typically studies a single learner under an exogenous censoring rule; here each learner's observed distribution is endogenously determined by user choices among competing models.

2 PROBLEM SETTING

Users and Learners. Consider a setting with m service providers (learners) serving a population of users distributed according to $\mathcal{P}$ over $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, where $\mathcal{X} \subseteq \mathbb{R}^d$ denotes covariates and $\mathcal{Y}$ denotes labels ($\mathcal{Y} \subseteq \mathbb{R}$ for regression, $\mathcal{Y} \subseteq \{1, \ldots, C\}$ for classification). We write $\mathcal{P}_X$ for the marginal distribution over covariates and $\mathcal{P}_{Y|X}(\cdot \mid x)$ for the conditional label distribution. When densities exist, we denote them by p_X and $p_{Y|X}$.

Each learner $i \in [m] := \{1, \ldots, m\}$ maintains a model with parameters θ_i, and we write $\Theta = (\theta_1, \ldots, \theta_m)$ for the joint parameter vector. The loss $\ell(x, y; \theta)$ measures the cost incurred by a model with parameters θ on a user with features x and label y. For regression, we consider linear predictors $h_\theta(x) = x^\top \theta$ with squared loss $\ell_{\mathrm{SQ}}(x, y; \theta) = (y - x^\top \theta)^2$. For classification with C classes, we use cross-entropy loss $\ell_{\mathrm{CE}}(x, y; \theta) = -\sum_{c=1}^{C} y_c \log h_\theta(x)_c$, where $h_\theta(x)_c = e^{x^\top \theta_c} / \sum_{j=1}^{C} e^{x^\top \theta_j}$ is the softmax output.

Assumption 1. *The distribution $\mathcal{P}$ has continuous density $p_Z(z) = p_X(x) p_{Y|X}(y|x)$ with p_X supported on $\{x : \|x\| \leq R\}$ for some $R > 0$. For regression, $p_{Y|X}$ is supported on $[-Y_{\max}, Y_{\max}]$.*

Under this assumption, both loss functions satisfy standard regularity conditions; see Lemma 7 in Appendix B.

User Preferences and Platform Choice. A central feature of our framework is that users have *inherent preferences* over platforms that exist independently of current model quality. These preferences capture factors such as brand loyalty, familiarity, network effects, or historical habits. We encode them via a function $\pi : \mathcal{Z} \to [m]$, where $\pi(z)$ denotes the platform that user z intrinsically prefers. This function is exogenous and fixed throughout learning.

The preference function π induces a partition of the user space. Let $S_i = \{z \in \mathcal{Z} : \pi(z) = i\}$ denote users who prefer platform i, with $\mathcal{P}_i = \mathcal{P}|_{S_i}$ the corresponding conditional distribution and $\alpha_i = \Pr_{z \sim \mathcal{P}}[\pi(z) = i]$ the fraction of users preferring platform i. Users do not only follow their inherent preferences; they also consider predictive quality. We model this tradeoff as follows:

Definition 1 (User Selection Rule). *Given models Θ, user z selects platform $M(z; \Theta) = \pi(z)$ with probability τ, and otherwise (with probability $1 - \tau$) selects $M(z; \Theta) = \arg\min_{i \in [m]} \ell(z; \theta_i)$.*

The parameter $\tau \in [0, 1]$ governs how strongly inherent preferences influence user behavior: $\tau = 1$ means users follow intrinsic preferences entirely, while $\tau = 0$ means they select the loss-minimizing platform. This generalizes Su and Dean [2024], who assume users either minimize loss or choose uniformly at random.

Model quality also induces a partition of users. Let $Z_i(\Theta) = \{z : i = \arg\min_{j \in [m]} \ell(z; \theta_j)\}$ denote users for whom platform i achieves minimal loss, with $\mathcal{D}_i(\Theta) = \mathcal{P}|_{Z_i(\Theta)}$ the conditional distribution and $a_i(\Theta) = \Pr_{z \sim \mathcal{P}}[z \in Z_i(\Theta)]$ the population fraction. Under Definition 1, learner i observes users from the mixture

$$\mathcal{O}_i(\Theta) = \tau \alpha_i \mathcal{P}_i + (1 - \tau) a_i(\Theta) \mathcal{D}_i(\Theta),$$

where $\mathcal{O}_i(\Theta)$ is a sub-probability measure with total mass $w_i(\Theta) = \tau \alpha_i + (1 - \tau) a_i(\Theta)$.

Dynamics and Learning Objective. We consider an online setting where learners interact with users over T timesteps, illustrated in Figure 1. At each step t:

1. A user $z^t \sim \mathcal{P}$ arrives and selects platform $M(z^t; \Theta^t)$.
2. The selected learner observes z^t, incurs loss $\ell(z^t; \theta^t_{M(z^t; \Theta^t)})$, and updates its parameters.

Learning Objective. Each learner's goal is to minimize the *full-population risk*

$$\mathcal{R}(\theta) = \mathbb{E}_{z \sim \mathcal{P}}[\ell(z; \theta)].$$

We write $\theta^\star = \arg\min_\theta \mathcal{R}(\theta)$ for the population-optimal model and $\epsilon = \mathcal{R}(\theta^\star)$ for the Bayes risk. This objective differs from prior work [Su and Dean, 2024, Dean et al., 2024], which focuses on the "local" loss over each learner's observed distribution $\mathcal{O}_i(\Theta)$. We focus on full-population risk because the signature of overspecialization is precisely the gap between local and global performance: a learner may achieve low loss on $\mathcal{O}_i(\Theta)$ while performing poorly on users outside its observed base. Understanding when this gap emerges and how to prevent it is the central question of this paper.

3 THE FAILURE OF STANDARD LEARNING DYNAMICS

We now analyze Multi-learner Streaming Gradient Descent (MSGD), the standard algorithm for this setting introduced by Su and Dean [2024], Algorithm 1.

3.1 ALGORITHM AND ASSUMPTIONS

In order to study the convergence behavior of the algorithm, we make the following standard assumptions on learning rates and loss geometry, which are the same as in Su and Dean [2024].

Assumption 2. *The learning rates satisfy $\sum_{t=1}^{\infty} \eta^t = \infty$ and $\sum_{t=1}^{\infty} (\eta^t)^2 < \infty$.*

Assumption 3. *For any $\theta \neq \theta'$, there exists $d_0 > 0$ such that for all $d < d_0$, the set $\{z : |\ell(z; \theta) - \ell(z; \theta')| < d\}$ has Lebesgue measure at most d.*

Algorithm 1 Multi-learner Streaming Gradient Descent (MSGD) [Su and Dean, 2024]

Require: Loss function $\ell(\cdot,\cdot) \geq 0$; initial models $\Theta^0 = (\theta_1^0, \ldots, \theta_m^0)$; learning rate $\{\eta^t\}_{t\geq 1}$
1: **for** $t = 0, 1, 2, \ldots, T$ **do**
2: Sample user $z^t \sim \mathcal{P}$
3: User selects learner $i = M(z^t; \Theta^t)$
4: $\theta_i^{t+1} \leftarrow \theta_i^t - \eta^t \nabla_\theta \ell(z^t; \theta_i^t)$
5: **end for**
6: **return** Θ^T

Assumption 3, from Su and Dean [2024], rules out pathological cases where a large mass of users is nearly indifferent between two learners. Intuitively, it ensures that small parameter perturbations move only a small mass of users across loss-induced decision boundaries, so the partition-dependent potential f can be differentiated without boundary terms. Assumption 4 below is the standard boundedness condition used in stochastic-approximation analyses.

3.2 CONVERGENCE TO STATIONARY POINTS

In MSGD, each learner i optimizes its expected loss over observed users. Define the potential function $f(\Theta)$ as the sum of the expected losses of all learners on the distributions that they observe.

$$f(\Theta) = \sum_{i=1}^{m} \mathbb{E}_{\mathcal{O}_i(\Theta)}[\ell(z;\theta_i)], \tag{1}$$

Note that this is the sum of the "local" losses of each of the learners on their observed distribution, unlike the global risk $\mathcal{R}(\cdot)$ that was introduced above.

In order to show convergence, we make the same boundedness assumptions as in Su and Dean [2024].

Assumption 4. *The parameter sequence $\{\Theta^t\}_{t\geq 0}$ is almost surely bounded: $\sup_{t\geq 0} \|\Theta^t\| < \infty$. Moreover, the set $\{\Theta : \nabla f(\Theta) = 0\}$ is compact.*

Our approach uses stochastic approximation [Borkar, 2008] and differs from that of Su and Dean [2024]. The key insight is that f serves as a Lyapunov function: despite each learner optimizing over a different, endogenously-determined user distribution, the aggregate of local losses forms a coherent potential. This is surprising because multi-agent gradient dynamics often cycle or diverge [Mazumdar et al., 2020]; The following lemma, proved using standard stochastic approximation arguments, establishes this connection.

Lemma 1. *Let Assumptions 1–4 hold. Then the iterates $\{\Theta^t\}$ of Algorithm 1 converge to a compact connected internally chain transitive invariant set of the ODE $\dot{\Theta} = -\nabla f(\Theta)$.*

See Appendix B.1 for definitions of the dynamical-systems terms used in Lemma 1. The lemma shows that the discrete stochastic dynamics behave, in the limit, like the continuous gradient flow on f. Since f decreases along trajectories of this ODE, i.e., $\frac{d}{dt} f(\Theta(t)) = -\|\nabla f\|^2 \leq 0$, the only invariant sets are stationary points. This yields our main convergence result.

Theorem 1. *Let Assumptions 1–4 hold. Then the iterates $\{\Theta^t\}$ of Algorithm 1 converge to the set of stationary points $\{\Theta : \nabla f(\Theta) = 0\}$ almost surely.*

The same potential argument extends to mini-batch updates, matching the implementation used in the experiments. If a batch $\mathcal{B}^t$ is sampled i.i.d. and learner i updates using the average gradient over the selected subset $S_i^t = \{z \in \mathcal{B}^t : M(z;\Theta^t) = i\}$, then conditional on $S_i^t \neq \emptyset$ this average is an unbiased gradient over $\mathcal{O}_i(\Theta^t)/w_i(\Theta^t)$. Since $w_i(\Theta) \geq \tau\alpha_i > 0$ for $\tau > 0$, the batch dynamics track the same potential with a bounded state-dependent rescaling of the step size.

3.3 THE OVERSPECIALIZATION TRAP

While Theorem 1 guarantees convergence, the stationary points may be highly undesirable. A learner cannot improve on users it never observes, and it never observes users it cannot serve well. This feedback loop is the overspecialization trap.

The following theorem shows that this trap can be severe: MSGD can converge to equilibria where some learners have arbitrarily poor global performance, even when models with low full-population loss exist.

Theorem 2. *Let Assumptions 1–4 hold. For any $\tau \geq \frac{1}{2}$ and any choice of ϵ, Γ with $0 < \epsilon < \Gamma$, there exists an instance $\mathcal{G} = (\mathcal{P}, \ell, \pi, \tau)$ such that:*

1. *There exists $\theta^\star$ with $\mathcal{R}(\theta^\star) \leq \epsilon$.*
2. *The MSGD iterates converge to a unique stationary point $\bar{\Theta}$ where $\mathcal{R}(\bar{\theta}_i) \geq \Gamma$ for some learner $i \in [m]$.*

The key mechanism is as follows. When $\tau \geq \frac{1}{2}$, inherent preferences dominate at equilibrium: the loss-induced partition collapses to the ranking partition, so that $Z_i(\Theta) = S_i$ for all i. Each learner optimizes exclusively for users who intrinsically prefer it, arriving at

$$\bar{\theta}_i = \arg\min_{\theta} \mathbb{E}_{z\sim\mathcal{P}_i}[\ell(z;\theta)].$$

This specialization occurs regardless of whether a better global model exists. In the constructed instance (detailed in Appendix C), learner 1 achieves *zero* loss on its observed population $\mathcal{P}_1$ while its full-population loss exceeds Γ. The learner has perfectly fit its niche while becoming arbitrarily poor globally. This dynamic formalizes the echo chamber phenomenon: platforms become increasingly specialized to their existing audience, unable to learn models that serve the broader population.

However, when $\tau < \frac{1}{2}$ and quality-based selection dominates, there may be multiple equilibria. Now, the limit point becomes initialization-dependent, which presents a technical obstacle to characterizing the limiting risk in closed form; however, experiments in Section 5 demonstrate similar phenomena across different values of τ.

4 MITIGATING OVERSPECIALIZATION THROUGH PEER PROBING

In many practical settings, learners can *probe* peer models to obtain pseudo-labels on unseen user segments. This practice has become increasingly common through knowledge distillation. Despite its growing practical importance, the theoretical implications of these multi-agent interactions remain largely unexplored. Under what circumstances can probing help overcome the overspecialization trap?

4.1 ALGORITHM

We propose MSGD with Probing (MSGD-P), shown in Algorithm 2. The algorithm has two phases. In the *offline phase*, each probing learner $j \in U$ collects a dataset $\mathfrak{D}_j$ of pseudo-labeled examples by querying peer models on sampled covariates. In the *online phase*, learners interleave standard MSGD updates (on organic users) with gradient steps on their probing datasets.

More concretely, for each probing learner $i \in U$, we sample covariates $(\tilde{x}_i^1, \dots, \tilde{x}_i^n) \sim \mathcal{P}_X^n$. Given a query covariate x, the learner selects a subset of peers $T_i(x) \subseteq [m]$ to consult, and forms a pseudo-label via median aggregation

$$y_{\text{agg},i}(x, \Theta) = \text{median}\{h_{\theta_j}(x) : j \in T_i(x)\}.$$

We then define $\tilde{y}_i^q := y_{\text{agg},i}(\tilde{x}_i^q, \Theta^0)$ and the pseudo-labeled examples $\tilde{z}_i^q := (\tilde{x}_i^q, \tilde{y}_i^q)$, and collect the probing dataset $\mathfrak{D}_i := \{\tilde{z}_i^q\}_{q=1}^n$. The choice of $T_i(x)$ determines which peers are consulted; we discuss this in Section 4.3.

For a probing learner $i \in U$, the update can be interpreted as a stochastic gradient step on the following instantaneous loss:

$$\begin{aligned} L_i^t(\theta_i) = \tau \alpha_i \mathbb{E}_{z \sim \mathcal{P}_i}[\ell(z; \theta_i)] \\ + (1-\tau) a_i(\Theta^t) \mathbb{E}_{z \sim \mathcal{D}_i(\Theta^t)}[\ell(z; \theta_i)] \\ + \frac{p}{n} \sum_{q=1}^{n} \ell(\tilde{z}_i^q; \theta_i) + \frac{\lambda p}{2} \|\theta_i\|^2. \end{aligned} \quad (2)$$

The first two terms capture organic learning from users who select learner i (via inherent preference $\mathcal{P}_i$ or quality-based choice $\mathcal{D}_i$), while the third term captures learning from probing data. The parameter $p > 0$ controls the relative weight of probing gradients: larger p emphasizes pseudo-labels, smaller p prioritizes organic data.

Algorithm 2 Multi-learner Streaming Gradient Descent with Probing (MSGD-P)

Require: loss function $\ell(\cdot, \cdot) \geq 0$; Initial models $\Theta^0 = (\theta_1^0, \dots, \theta_m^0)$; Learning rate $\{\eta^t\}_{t=1}^{T+1}$, probing weight $p > 0$, set of probing learners $U \subseteq [m]$, regularization weight $\lambda \geq 0$

// Offline probing data collection
for $j \in U$: **do**
  Sample covariates $(\tilde{x}_j^1, \dots, \tilde{x}_j^n) \sim \mathcal{P}_X^n$.
  Collect pseudo-labels and store dataset $\mathfrak{D}_j = \{(\tilde{x}_j^1, y_{\text{agg},i}(\tilde{x}_j^1, \Theta^0)) \dots (\tilde{x}_j^n, y_{\text{agg},i}(\tilde{x}_j^n, \Theta^0))\}$
end for
// Online updates
for $t = 0, 1, 2, \dots, T$ **do**
  Sample data point $z^t \sim \mathcal{P}$
  User selects model $i = M(z^t; \Theta^t)$
  $\theta_i^{t+1} \leftarrow \theta_i^t - \eta^t \nabla \ell(z^t, \theta_i^t)$
  for $j \in U$: **do**
    Sample $\tilde{z}_j^t$ uniformly from $\mathfrak{D}_j$
    $\theta_j^{t+1} \leftarrow \theta_j^t - \eta^t p \left(\nabla \ell(\tilde{z}_j^t, \theta_j^t) + \lambda \theta_j^t\right)$
  end for
end for
return Θ^T

We assume that probing learners can sample covariates from the full distribution $\mathcal{P}_X$, but do not have access to true labels. This asymmetry is natural: covariates are often publicly available or easy to generate (e.g., movie metadata, user demographics, or text prompts), while labels require costly human annotation or reveal private user behavior (e.g., individual ratings or response quality judgments).

We focus on *offline probing*, where pseudo-labels are collected once at initialization from a fixed snapshot of peer models. This mirrors practical distillation workflows where teachers are queried to create a fixed dataset and student training proceeds independently [Werner, 2025]. Offline probing also ensures reproducibility by capturing a specific model version's behavior, avoiding inconsistencies from querying adapting peers. We discuss online probing in Section 6.

4.2 CONVERGENCE

With probing, each learner $i \in U$ now optimizes a blend of two objectives: the loss on observed users (as in standard MSGD) and the loss on probing data. This leads to a modified potential function:

$$\widetilde{f}(\Theta) = f(\Theta) + p \sum_{i \in U} \left(\frac{1}{n} \sum_{q=1}^{n} \ell(\tilde{z}_i^q, \theta_i) + \frac{\lambda}{2} \|\theta_i\|^2 \right), \quad (3)$$

where the second term captures the probing loss (with regularization) weighted by p.

The same stochastic approximation analysis from Section 3 extends to this setting: $\widetilde{f}$ serves as a Lyapunov function for the modified dynamics, yielding the following convergence guarantee.

Theorem 3. *Let Assumptions 1–3, Assumption 4 (as applied to $\widetilde{f}$), and Assumption 5 hold. Then, the iterates $\{\Theta^t\}$ of Algorithm 2 converge to the set of stationary points $\{\Theta : \nabla\widetilde{f}(\Theta) = 0\}$ almost surely.*

Crucially, the stationary points of $\widetilde{f}$ differ from those of f: probing changes *where* learners converge, not *whether* they converge.

4.3 WHEN DOES PROBING HELP?

Note that the convergence guarantee above holds for any choice of $T_i(x)$. However, in order for the probing data to be *helpful*, the pseudo-labels must be a good proxy for the ground-truth labels.

Assumption 5 (Accurate Probing)**.** *We say that the "accurate probing" condition holds for probing learner $i \in [m]$ if there exists $B \geq 0$ such that*

$$\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{P}}\left[\left(y_{agg,i}(x, \Theta_{-i}) - y\right)^2\right] \leq B.$$

This assumption is stated for a single probing learner i. Different learners may satisfy it via different scenarios (or not at all); performance guarantees in Section 4.4 apply to any learner for whom the assumption holds.

Below, we identify scenarios under which Assumption 5 holds. These scenarios differ along two axes: *what learner i must know* about the market, and *what must be true* about peer models at initialization. Table 1 summarizes these scenarios; the second and third columns display the knowledge-vs-peer-requirement tradeoff.

Definition 2 (Globally good peers)**.** *We say learner i can achieve accurate probing via* globally good peers *if any of the following scenarios hold:*

(i) ***Majority-good.*** *More than half of the learners satisfy $\theta_j^0 \in B_r(\theta^*)$ for a given proximity parameter $r > 0$.*

(ii) ***Market-leader.*** *There exists a single learner $j^* \in [m]$ such that*

$$\mathcal{R}(\theta_{j^*}^0) \leq \xi,$$

and the identity of j^* is known to learner i.

(iii) ***Partial knowledge.*** *There exists a subset $G \subseteq [m] \setminus \{i\}$ such that more than half of the learners in G satisfy $\theta_j^0 \in B_r(\theta^*)$ for a given $r > 0$, and learner i has knowledge of the subset G.*

Above, the parameter $r > 0$ in the majority-good and partial-knowledge scenarios controls how close peers must be to θ^*; smaller r yields tighter bounds. When no globally good learners exist or can be identified, probing may still be effective if the learner has access to ranking information.

Definition 3 (Preference-aware probing)**.** *Suppose all learners initialize at the parameters $\bar{\Theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_m)$. If learner i has knowledge of the inherent preference function $\pi(z)$, which identifies each user's preferred platform, we call this the* preference-aware *scenario.*

These definitions capture different approaches to probing: *Definition 2* covers settings where learners can identify or rely on peers with strong global performance, while *Definition 3* addresses settings where learners must instead leverage knowledge of user preferences. The scenarios exhibit a fundamental tradeoff: when peer models are favorable (e.g., many are globally good), learner i needs little knowledge to achieve accurate probing; conversely, with stronger knowledge (e.g., knowing $\pi(x)$), the learner can probe in a more targeted way, and probing succeeds even when no peer is globally competent.

In realistic markets, learners often naturally gain access to information enabling one of these approaches. For Scenarios (i) or (iii), platforms may observe broad industry benchmarks [Liang et al., 2023, Chiang et al., 2024, Russakovsky et al., 2015, Bennett and Lanning, 2007, Olson et al., 2017]. For Scenario (ii), there are examples in the LLM literature of this explicitly happening: for instance, the Alpaca model [Taori et al., 2023] was explicitly trained on data generated by text-davinci-003 (GPT-3.5), the Vicuna model [Chiang et al., 2023] was trained on data generated by GPT-4, and Gudibande et al. [2024] document the prevalence of this practice. For Definition 3, it is natural to maintain knowledge of user preference patterns [Guadagni and Little, 1983, Abdullah et al., 2021, Huang et al., 2018, Irani-Kermani et al., 2023].

The probing rule $T_i(x)$ (fourth column of Table 1) in each scenario is chosen so that median aggregation is robust: probing all peers when the majority are good, targeting the known leader when one exists, or routing to the locally-expert peer in the preference-aware case. The preference-aware scenario is particularly notable: it requires *no assumption* on peer quality, only knowledge of user preferences $\pi(x)$, enabling learner i to aggregate specialized knowledge into global competence even when every peer suffers from overspecialization.

Lemma 2. *For each scenario in Definitions 2–3, the probing rule $T_i(x)$ in Table 1 satisfies Assumption 5 with the stated accuracy parameter B.*

Scenario	What i knows	Peer requirement	$T_i(x)$	B
Majority-good	Nothing	$> 50\%$ in $B_r(\theta^*)$	$[m]$	$R^2r^2 + 2\epsilon$
Market-leader	Identity of j^*	$\mathcal{R}(\theta^0_{j^*}) \leq \xi$	$\{j^*\}$	ξ
Partial knowledge	Subset G	$> 50\%$ of G in $B_r(\theta^*)$	G	$R^2r^2 + 2\epsilon$
Preference-aware	$\pi(x)$	**Nothing**	$\{\pi(x)\}$	ϵ

Table 1: Probing scenarios with corresponding rules and accuracy bounds. The preference-aware scenario is notable: it requires no assumption on peer quality, only knowledge of user preferences.

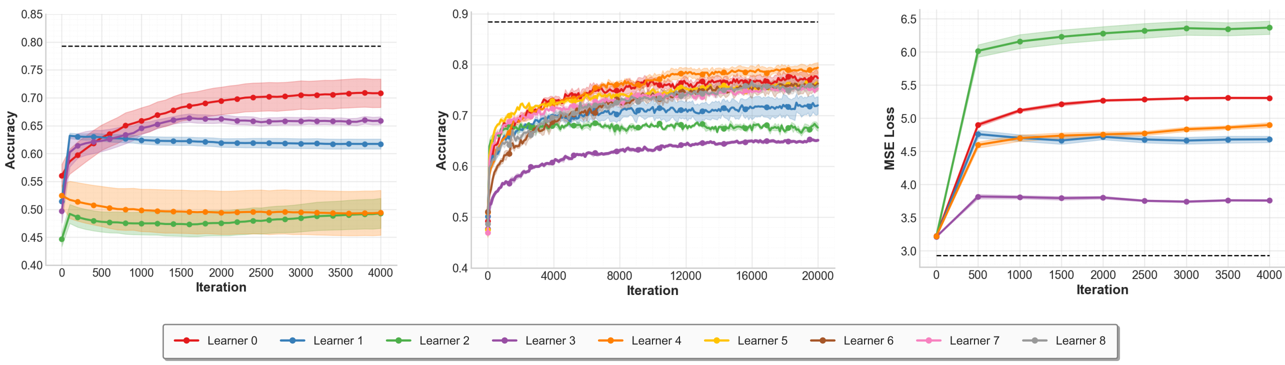

Figure 2: **MSGD full-population performance with random initialization (Preference-aware scenario).** Left: Census test accuracy. Mid: Amazon sentiment test accuracy Right: MovieLens test loss. The dashed black line represents the performance of a baseline θ^* trained on the full dataset, and error bars depict standard error. In all cases, $\tau = 0.3$; other dataset-specific hyperparameters are given in Table 3.

4.4 PERFORMANCE GUARANTEES

We now characterize the full-population risk of MSGD-P stationary points for the squared loss.

Theorem 4. *Let Assumptions 1–5 hold. For any $\kappa \in (0, 1)$, with probability at least $1 - \kappa$, every stationary point $\tilde{\Theta}$ of MSGD-P satisfies, for each probing learner $i \in U$, for the squared loss:*

$$\mathcal{R}(\tilde{\theta}_i) \ \leq \ O\Big(\Big(\frac{p+1}{p}\Big)\epsilon + B + \lambda\|\theta^\star\|^2 + \frac{(p+1)C_{gen}}{p\lambda}\sqrt{\frac{\log(1/\kappa)}{n}}\Big).$$

where B is quantified in Table 1 for each scenario, and C_{gen} depends on R, $Y_{\max}$, $\|\theta^\star\|$, $\max_{j\neq i}\|\theta^0_j\|$, with explicit form in the Appendix.

The four terms admit natural interpretations: (i) $\frac{p+1}{p}\epsilon$ is the irreducible Bayes error, scaled by the ratio of total to probing gradient weight; (ii) B is the probing bias from pseudo-label inaccuracy (see Table 1); (iii) $\lambda\|\theta^\star\|^2$ is the regularization bias; and (iv) the final term captures finite-sample generalization error from n probing queries. The parameter λ trades off regularization bias ($\lambda\|\theta^\star\|^2$) against generalization ($O(1/\sqrt{\lambda n})$).

Corollary 1. *Let Assumptions 1–5 hold. Fix $\lambda = \epsilon/\|\theta^\star\|^2$ and any $\kappa \in (0, 1)$. Define $M_0 := \max_{j\neq i}\|\theta^0_j\|$. If there are sufficiently many probing samples $n \geq \underline{n}(p, \kappa, \epsilon, R, Y_{\max}, M_0, \|\theta^\star\|)$, then with probability at least $1 - \kappa$, every stationary point $\tilde{\Theta}$ of MSGD-P satisfies, for each probing learner i,*

$$\mathcal{R}(\tilde{\theta}_i) \ \leq \ O\Big(\Big(\frac{p+1}{p}\Big)\epsilon + B\Big).$$

Hence, probing breaks the information barrier created by user-choice dynamics. While Theorem 2 shows that the risk of a learner under MSGD may be arbitrarily worse than ϵ, the bound above presents a ceiling on the risk of any probing learner for sufficiently large n.

See Appendix E for the explicit sample complexity $\underline{n}$ and proof. We note that this sample complexity bound is not tight in n: we show in Figure 6 that strong empirical recovery can occur with very small probing datasets.

Remark 1 (Cross-entropy loss)**.** *An analogous performance guarantee holds for cross-entropy loss; see Assumption 6 and Appendix F for the bound, and Table 2 for the corresponding accuracy parameters.*

5 NUMERICAL EXPERIMENTS

We evaluate on MovieLens-10M, ACS Employment from the US Census (Alabama, 2018), and Amazon Reviews 2023.

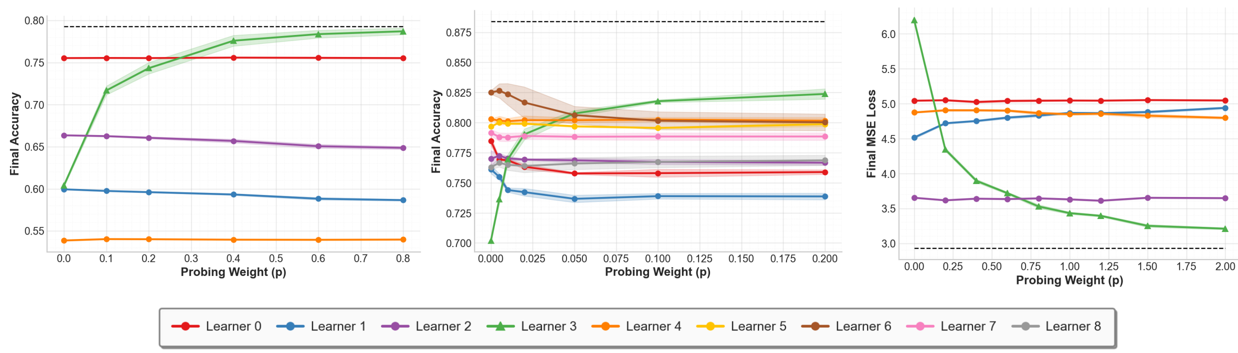

Figure 3: **Effect of probing on full-population performance when initialized at $\bar{\Theta}$ (Preference-aware scenario).** Left: Census final accuracy vs probing weight p. Mid: Amazon sentiment final accuracy vs probing weight p Right: MovieLens final loss vs p. In all cases, the green learner is the probing learner, and error bars depict standard error. Here $\tau = 0.7$; other dataset-specific hyperparameters are given in Table 3.

Datasets. **MovieLens-10M** [Harper and Konstan, 2015] has 69,474 users with $d = 16$ matrix-factorization embeddings; learners predict ratings for 200 movies with squared loss. **ACS Employment** [Ding et al., 2021] has 38,221 Alabama Census users with $d = 16$ demographic features; learners predict employment status. **Amazon Reviews 2023** has up to 30,000 reviews from nine categories with $d = 384$ sentence embeddings; learners predict binary sentiment. Full preprocessing details are in Appendix G.

User Preferences. For Census and MovieLens ($m = 5$), inherent preferences $\pi(z)$ are induced by K-means clustering on user features. For Amazon ($m = 9$), preferences are determined by product category, with all reviews from a given category preferring the same learner.

Evaluation. We use one fixed train/test split for all learners and do no validation tuning; metrics are test accuracy for Census/Amazon and masked test MSE for MovieLens.

5.1 EXPERIMENTAL RESULTS

We present results for the preference-aware scenario from Definition 3. The results are qualitatively similar in the other scenarios from Definition 2 as well, and are presented in Appendix G.

Expt 1: MSGD converges to equilibria with poor global performance. Our first set of experiments validates Theorem 2. Figure 2 shows the full-population performance trajectories of individual learners on all three datasets without probing ($p = 0$) over $T = 4000$ rounds, when initialized randomly. In all cases, the results reveal large overspecialization gaps relative to the dashed black baseline.

Expt 2: Peer Model Probing Mitigates Overspecialization. Figure 3 shows learner final performance at timestep T as a function of probing weight p using the offline probing budgets in Table 3; here, learner 3 (indicated by triangle markers) uses preference-aware probing while other learners use standard MSGD. On Census (left), Learner 3's accuracy improves from approximately 60% at $p = 0$ to 78% at $p = 0.8$, shrinking its baseline gap from roughly 18 percentage points to about 1 percentage point. On MovieLens (right), Learner 2's MSE loss decreases from approximately 6.2 to 3.2.

Expt 3: Probing is robust to noisy source selection. Figure 4 tests a noisy version of preference-aware probing: for each probe query, the learner routes to $\pi(x)$ with probability $1 - \kappa$ and to a random peer with probability κ. Across datasets, probing remains beneficial under moderate routing noise, showing that the method does not require perfect knowledge of user preferences.

Expt 4: How much probing data is needed? Figure 6 studies sample efficiency on Census: for the probing learner with $p \in \{0.5, 1.0\}$, accuracy rises from about 0.68 at $n = 5$ to about 0.78 by $n = 50$, a tiny fraction of the $38,221$ examples, consistent with Theorem 4.

Expt 5: Two-layer neural learners. Figure 5 shows the same overspecialization/probing-recovery pattern for Census two-layer ReLU MLP learners.

6 DISCUSSION

We showed that standard learning dynamics in competitive ML markets converge to overspecialized equilibria, while peer model probing can mitigate this failure. Probing breaks information barriers created by user-choice dynamics, with guarantees that degrade gracefully with pseudo-label quality.

Online versus offline probing. We simulate preferences via K-means clustering; richer choice models merit explo-

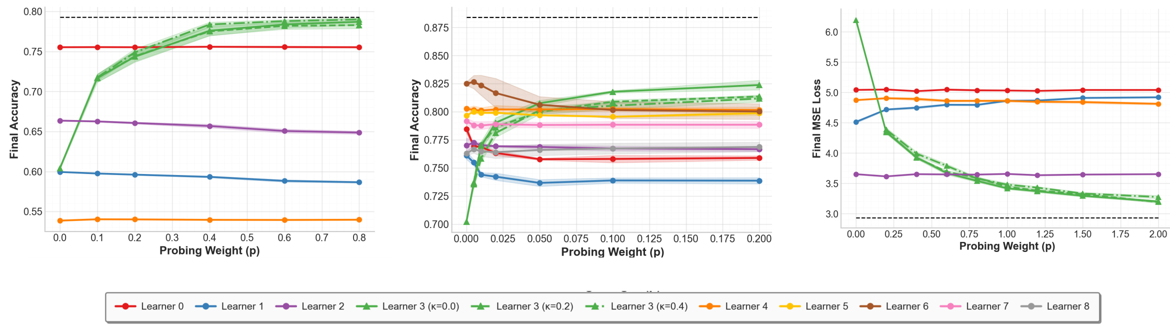

Figure 4: **Effect of noisy probing-source selection.** Probe queries route to $\pi(x)$ with probability $1-\kappa$ and to a random peer with probability κ. Left: Census accuracy vs. p. Middle: Amazon accuracy vs. p. Right: MovieLens MSE vs. p. The green learner probes; its multiple curves show different κ values.

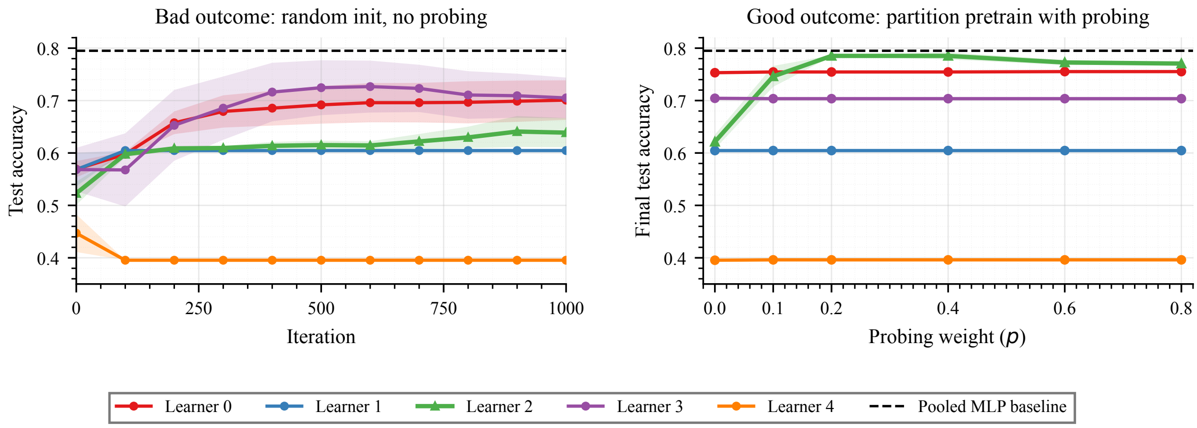

Figure 5: **Census experiments with two-layer neural learners.** Each logistic learner is replaced by a two-layer ReLU MLP ($d \to 64 \to 1$), keeping the Census split, clustering, rankings, and offline pseudo-label construction fixed. Left: random initialization, no probing ($\tau = 0.3, p = 0$). Right: partition-pretrained initialization with Learner 2 probing ($\tau = 0.7, \kappa = 0$). Bands show standard error over three seeds.

ration. Our offline probing collects pseudo-labels once at initialization, mirroring fixed-teacher distillation and avoiding queries to adapting peers. Online probing would introduce co-adaptation dynamics and possible instabilities reminiscent of model collapse [Shumailov et al., 2024].

Extensions to non-convex learners. Our theory is restricted to convex losses with linear predictors; Figure 5 suggests the phenomena persist for simple non-convex learners, but extending guarantees to deep and overparameterized models remains open [Narang et al., 2021].

References

Nor Aniza Abdullah, Rasheed Abubakar Rasheed, Mohd Hairul Nizam Nasir, and Md Mujibur Rahman. Eliciting auxiliary information for cold start user recommendation: A survey. *Applied Sciences*, 11(20):9608, 2021.

Rohan Anil, Gabriel Pereyra, Alexandre Passos, Róbert Ormandi, George E Dahl, and Geoffrey E Hinton. Large scale distributed neural network training through online distillation. In *International Conference on Learning Representations (ICLR)*, 2018.

Jye E. Beardow. Scroll, click, like, share, repeat: The algorithmic polarisation phenomenon. *ANU Journal of Law & Technology*, 2(1):153–164, 2021. Autumn 2021 issue.

Omer Ben-Porat and Moshe Tennenholtz. Best response regression. *Advances in Neural Information Processing Systems*, 30, 2017.

Omer Ben-Porat and Moshe Tennenholtz. Regression equilibrium. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 173–191, 2019.

James Bennett and Stan Lanning. The Netflix prize. In *Proceedings of KDD Cup and Workshop*. ACM, 2007.

Vivek S Borkar. *Stochastic approximation: a dynamical systems viewpoint*, volume 9. Springer, 2008.

Avinandan Bose, Mihaela Curmei, Daniel L Jiang, Jamie Morgenstern, Sarah Dean, Lillian J Ratliff, and Maryam Fazel. Initializing services in interactive ml systems for diverse users. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Yeshwanth Cherapanamjeri, Constantinos Daskalakis, Andrew Ilyas, and Manolis Zampetakis. What makes a good fisherman? linear regression under self-selection bias. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023. Association for Computing Machinery, 2023. doi: 10.1145/3564246.3585177. URL https://doi.org/10.1145/3564246.3585177.

Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. Vicuna: An open-source chatbot impressing GPT-4 with 90%* ChatGPT quality. https://lmsys.org/blog/2023-03-30-vicuna/, March 2023. Accessed: 2025.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E Gonzalez, and Ion Stoica. Chatbot arena: An open platform for evaluating LLMs by human preference. In *Proceedings of the 41st International Conference on Machine Learning*, 2024. URL https://arxiv.org/abs/2403.04132.

Federico Cinus, Marco Minici, Corrado Monti, and Francesco Bonchi. The effect of people recommenders on echo chambers and polarization. In *Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM '22)*, pages 90–101. Association for the Advancement of Artificial Intelligence (AAAI), 2022. ICWSM 2022.

Sarah Dean, Mihaela Curmei, Lillian Ratliff, Jamie Morgenstern, and Maryam Fazel. Emergent specialization from participation dynamics and multi-learner retraining. In *International Conference on Artificial Intelligence and Statistics*, pages 343–351. PMLR, 2024.

Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. Retiring Adult: New datasets for fair machine learning. *Advances in Neural Information Processing Systems*, 34:6478–6490, 2021.

Shi Feng, Hanlin Zhang, Fan Nie, Sham Kakade, and Yiling Chen. Peer-predictive self-training for language model reasoning. *arXiv preprint arXiv:2604.13356*, 2026.

Tanner Fiez, Lillian Ratliff, Eric Mazumdar, Evan Faulkner, and Adhyyan Narang. Global convergence to local minmax equilibrium in classes of nonconvex zero-sum games. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/f3507289cfdc8c9ae93f4098111a13f9-Abstract.html.

Tony Ginart, Eva Zhang, Yongchan Kwon, and James Zou. Competing ai: How does competition feedback affect machine learning? In *International Conference on Artificial Intelligence and Statistics*, pages 1693–1701. PMLR, 2021.

António Góis, Mehrnaz Mofakhami, Fernando P. Santos, Gauthier Gidel, and Simon Lacoste-Julien. Performative prediction on games and mechanism design. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 1855–1863. PMLR, 2025.

Peter M Guadagni and John DC Little. A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2(3):203–238, 1983.

Arnav Gudibande, Eric Wallace, Charlie Snell, Xinyang Geng, Hao Liu, Pieter Abbeel, Sergey Levine, and Dawn Song. The false promise of imitating proprietary LLMs. In *International Conference on Learning Representations (ICLR)*, 2024.

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

F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19, 2015.

Keegan Harris, Chara Podimata, and Zhiwei Steven Wu. Strategic apple tasting. *Adv. Neural Inf. Process. Syst.*, abs/2306.06250, June 2023.

Tatsunori Hashimoto, Megha Srivastava, Hongseok Namkoong, and Percy Liang. Fairness without demographics in repeated loss minimization. In *International Conference on Machine Learning*, pages 1929–1938. PMLR, 2018.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. URL https://arxiv.org/abs/1503.02531.

Hong Huang, Bo Zhao, Hao Zhao, Zhou Zhuang, Zhenxuan Wang, Xiaoming Yao, Xinggang Wang, Hai Jin, and Xiaoming Fu. A cross-platform consumer behavior analysis of large-scale mobile shopping data. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 1785–1794, Lyon, France, April 2018. International World Wide Web Conferences Steering Committee. doi: 10.1145/3178876.3186169. URL https://doi.org/10.1145/3178876.3186169.

Ruben Interian, Ruslán G. Marzo, Isela Mendoza, and Celso C. Ribeiro. Network polarization, filter bubbles, and echo chambers: An annotated review of measures and reduction methods. *arXiv preprint*, 2022. arXiv:2207.13799.

Roozbeh Irani-Kermani, Edward C. Jaenicke, and Ardalan Mirshani. Accommodating heterogeneity in brand loyalty estimation: Application to the U.S. beer retail market. *Journal of Marketing Analytics*, 11(4):820–835, 2023. doi: 10.1057/s41270-022-00187-2. URL https://doi.org/10.1057/s41270-022-00187-2.

Meena Jagadeesan, Michael I Jordan, and Nika Haghtalab. Competition, alignment, and equilibria in digital marketplaces. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(5):5689–5696, 2023a. doi: 10.1609/aaai.v37i5.25706.

Meena Jagadeesan, Michael I Jordan, Jacob Steinhardt, and Nika Haghtalab. Improved bayes risk can yield reduced social welfare under competition. *arXiv preprint arXiv:2306.14670*, 2023b.

Hailey James, Chirag Nagpal, Katherine A Heller, and Berk Ustun. Participatory personalization in classification. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Julie Jiang, Xiang Ren, and Emilio Ferrara. Social media polarization and echo chambers in the context of covid-19: Case study. *JMIRx med*, 2(3):e29570, 2021. doi: 10.2196/29570.

Yongchan Kwon, Antonio Ginart, and James Zou. Competition over data: how does data purchase affect users? *Transactions on Machine Learning Research*, 2022. URL https://openreview.net/forum?id=63sJsCmq6Q.

Qiang Li, Chung-Yiu Yau, and Hoi-To Wai. Multi-agent performative prediction with greedy deployment and consensus seeking agents. In *Advances in Neural Information Processing Systems*, volume 35, pages 38449–38460, 2022.

Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Cosgrove, Christopher D Manning, Christopher Ré, Diana Acosta-Navas, Drew A Hudson, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue Wang, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yüksekgönül, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri Chatterji, Omar Khessin, Peter Henderson, Qian Huang, Ryan Chi, Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chandu, William Wang, Xuechen Xie, Xuechen Zhang, Yushi Wang, Yuntao Zhou, and Yuta Koreeda. Holistic evaluation of language models. *Transactions on Machine Learning Research*, 2023. URL https://arxiv.org/abs/2211.09110.

Eric Mazumdar, Lillian J. Ratliff, and S. Shankar Sastry. On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, 2020. doi: 10.1137/18M1231298. URL https://doi.org/10.1137/18M1231298.

John Miller, Juan C Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7710–7720. PMLR, 2021.

Adhyyan Narang, Vidya Muthukumar, and Anant Sahai. Classification and adversarial examples in an overparameterized linear model: A signal processing perspective. *arXiv preprint arXiv:2109.13215*, 2021.

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.

Adhyyan Narang, Omid Sadeghi, Lillian Ratliff, Maryam Fazel, and Jeff Bilmes. Efficient interactive maximization of BP and weakly submodular objectives. In Negar Kiyavash and Joris M. Mooij, editors, *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pages 2670–2699. PMLR, 2024a. URL https://proceedings.mlr.press/v244/narang24a.html.

Adhyyan Narang, Andrew Wagenmaker, Lillian J. Ratliff, and Kevin Jamieson. Sample complexity reduction via policy difference estimation in tabular reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 37, 2024b. doi: 10.52202/079017-0717. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/28795419a

644f41ede3fa058b13fc622-Abstract-Conference.html.

Randal S Olson, William La Cava, Patryk Orzechowski, Ryan J Urbanowicz, and Jason H Moore. PMLB: A large benchmark suite for machine learning evaluation and comparison. *BioData Mining*, 10(1):36, 2017.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609. PMLR, 2020.

Georgios Piliouras and Fang-Yi Yu. Multi-agent performative prediction: From global stability and optimality to chaos. In *Proceedings of the 24th ACM Conference on Economics and Computation*, pages 1047–1048. ACM, 2023.

Reilly Raab, Ross Boczar, Maryam Fazel, and Yang Liu. Fair participation via sequential policies. In *AAAI Conference on Artificial Intelligence*, 2024.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115 (3):211–252, 2015.

H.J. Scudder. Probability of error of some adaptive pattern-recognition machines. *IEEE Transactions on Information Theory*, 11(3):363–371, July 1965.

Eliot Shekhtman and Sarah Dean. Strategic usage in a multi-learner setting. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2024.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. Ai models collapse when trained on recursively generated data. *Nature*, 631(8022):755–759, 2024.

Jinyan Su and Sarah Dean. Learning from streaming data when users choose. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 46772–46803. PMLR, 2024.

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. Stanford Alpaca: An instruction-following LLaMA model. https://github.com/tatsu-lab/stanford_alpaca, 2023. Accessed: 2025.

Guanghui Wang, Ioannis Panageas, Georgios Piliouras, and Fang-Yi Yu. Last-iterate convergence for symmetric, general-sum, 2×2 games under the exponential weights dynamic. *arXiv preprint arXiv:2502.08063*, 2025.

Xiaolu Wang, Chung-Yiu Yau, and Hoi To Wai. Network effects in performative prediction games. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 36514–36540. PMLR, 2023.

John Werner. Did deepseek copy off of openai? and what is distillation? *Forbes*, 2025. URL https://www.forbes.com/sites/johnwerner/2025/01/30/did-deepseek-copy-off-of-openai-and-what-is-distillation/.

David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *33rd Annual Meeting of the Association for Computational Linguistics*, pages 189–196, Cambridge, Massachusetts, USA, June 1995. Association for Computational Linguistics. doi: 10.3115/981658.981684.

Xueru Zhang, Mohammadmahdi Khaliligarekani, Cem Tekin, et al. Group retention when using machine learning in sequential decision making: the interplay between user dynamics and fairness. *Advances in Neural Information Processing Systems*, 32, 2019.

Ying Zhang, Tao Xiang, Timothy M Hospedales, and Huchuan Lu. Deep mutual learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4320–4328, 2018.

Zaiwei Zhu, Runyu Wan, Yingbin Cho, Haoming Luo, Zhuoran Yang, and Zhaoran Wang. Online performative gradient descent for learning nash equilibria in decision-dependent games. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Dynamics of Learning under User Choice: Overspecialization and Peer-Model Probing
(Supplementary Material)

Adhyyan Narang[1] **Sarah Dean**[2] **Lillian J. Ratliff**[1] **Maryam Fazel**[1]

[1]Electrical and Computer Engineering, University of Washington
[2]Computer Science, Cornell University

A EXTENDED RELATED WORK

This section provides detailed comparisons with related work summarized in Section 1.1 of the main paper.

Single-learner user agency. A body of work studies learning dynamics when users are treated as independent agents rather than passive data points. Hashimoto et al. [2018] show that empirical risk minimization can cause minority groups to opt out, creating a feedback loop that further degrades minority performance. Zhang et al. [2019] study a related dynamic where underrepresented groups receive worse predictions, leading to further data scarcity. James et al. [2023] analyze participatory data collection where users choose whether to contribute data. Ben-Porat and Tennenholtz [2017] study best-response dynamics in strategic classification. Cherapanamjeri et al. [2023] and Harris et al. [2023] study settings where data quality or availability depends on the learner's past performance. We build on this perspective in the multi-learner setting, where inter-learner interactions create additional feedback dynamics.

Multi-learner competition and user choice. Several works study multi-learner user-choice settings with explicitly strategic users who optimize their own utility functions [Shekhtman and Dean, 2024, Ben-Porat and Tennenholtz, 2017, 2019, Jagadeesan et al., 2023a,b]. Ginart et al. [2021] provide both empirical and theoretical analysis of how competition drives specialization: their Theorems 4.1–4.3 establish risk ratio bounds showing that competing predictors perform worse on the general population than a single predictor would. Their analysis considers batch retraining dynamics and characterizes the *existence* of performance gaps due to competition. We complement this by analyzing streaming gradient-based dynamics, proving that such dynamics *converge to* overspecialized equilibria (Theorem 2), and proposing probing as a mitigation. Kwon et al. [2022] study a related setting where learners may purchase user data, providing primarily empirical analysis of the resulting market dynamics.

Gradient-based dynamics in choice-driven settings. Most closely related to our work, Dean et al. [2024] and Su and Dean [2024] analyze gradient-based dynamics in choice-driven settings. Su and Dean [2024] introduce the MSGD algorithm and prove convergence to stationary points of an aggregate loss across learners. We build directly on their framework: our Algorithm 1 adapts MSGD to our user selection rule (Definition 1), and our potential function (Equation 1) extends theirs. Our Theorem 1 provides an alternate proof of their convergence result using stochastic approximation techniques, which we believe better illuminates the underlying dynamics. Beyond convergence, we analyze generalization to users outside each learner's observed population—a consideration absent from prior work—and introduce peer probing as a mechanism to restore global competence. Bose et al. [2024] propose intelligent initialization schemes to improve outcomes in similar settings, but also focus on the sum of all local losses on the observed distribution.

B TERMINOLOGY AND PRELIMINARY RESULTS

B.1 DYNAMICAL SYSTEMS TERMINOLOGY

This subsection records standard definitions underlying terms such as "invariant set" and "internally chain transitive" in Lemma 10. Let $\dot{x}(t) = h(x(t))$ be an ODE with locally Lipschitz h, so that solutions are unique. Let $\phi_t(x)$ denote the state at time $t \geq 0$ of the solution initialized at x at time 0.

Definition 4 (Invariant set)**.** *A set $A \subseteq \mathbb{R}^d$ is (forward)* invariant *for the ODE if $\phi_t(x) \in A$ for all $x \in A$ and all $t \geq 0$.*

Definition 5 ((ε, T)-chain)**.** *Fix $\varepsilon > 0$ and $T > 0$. An* (ε, T)-chain in A from x to y *is a finite sequence of points $x = x_0, x_1, \ldots, x_k = y$ in A and times $t_1, \ldots, t_k \geq T$ such that*

$$\|\phi_{t_i}(x_{i-1}) - x_i\| < \varepsilon \qquad \text{for all } i \in \{1, \ldots, k\}.$$

Definition 6 (Internally chain transitive)**.** *A compact invariant set A is* internally chain transitive *if for every $x, y \in A$ and every $\varepsilon > 0$, $T > 0$, there exists an (ε, T)-chain in A from x to y.*

B.2 REGULARITY OF SQUARED AND CROSS ENTROPY LOSSES

Lemma 3 (Regularity of the squared-error loss)**.** *Let $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Y} \subset \mathbb{R}$, $W \subset \mathbb{R}^d$ be compact. Write*

$$\ell(x, y; \theta) \ = \ \big(y - x^\top \theta\big)^2, \qquad B = \max_{\theta \in W} \|\theta\|.$$

Then for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and all $\theta_1, \theta_2 \in W$:

(i) ℓ is nonnegative, C^∞ in θ, convex, and β–smooth with $\beta = 2R^2$.

(ii) ℓ is locally Lipschitz on W:

$$\big|\ell(x, y; \theta_1) - \ell(x, y; \theta_2)\big| \leq L_W \, \|\theta_1 - \theta_2\|, \quad L_W = R\big(2|y| + 2RB\big).$$

(iii) The Hessian is constant in θ, so $\|\nabla^2 \ell(\theta_1) - \nabla^2 \ell(\theta_2)\| = 0$ (i.e. $\gamma_\ell = 0$).

Proof. (i) One computes $\nabla_\theta \ell = -2\,x\big(y - x^\top \theta\big)$ and $\nabla^2_\theta \ell = 2\,x\,x^\top \succeq 0$. Hence $\ell \geq 0$, is C^∞, convex, and $\|\nabla \ell(\theta_1) - \nabla \ell(\theta_2)\| \leq 2R^2 \|\theta_1 - \theta_2\|$.

(ii) Observe

$$\ell(\theta_1) - \ell(\theta_2) = \big(y - x^\top \theta_1\big)^2 - \big(y - x^\top \theta_2\big)^2 = x^\top (\theta_2 - \theta_1)\ \big(2y - x^\top (\theta_1 + \theta_2)\big).$$

Since $\|\theta_i\| \leq B$, $|x^\top (\theta_2 - \theta_1)| \leq R\|\theta_1 - \theta_2\|$ and $\big|2y - x^\top (\theta_1 + \theta_2)\big| \leq 2|y| + 2RB$, giving the stated bound.

(iii) Because $\nabla^2_\theta \ell \equiv 2\,x\,x^\top$ is independent of θ, its difference vanishes.

□

Lemma 4. *Under Assumption 1, for all $z \in \mathcal{Z}$ the loss $\ell(z, \cdot)$ is non-negative, convex, differentiable, locally Lipschitz and β_ℓ-smooth with*

$$\beta_\ell = 2R^2 \quad \text{for both the squared loss and cross-entropy loss.}$$

Moreover, let

$$\mathcal{R}(\theta) = \mathop{\mathbb{E}}_{z \sim \mathcal{P}}[\ell(z, \theta)], \quad \theta^* = \arg\min_\theta \mathcal{R}(\theta), \quad \epsilon = \mathcal{R}(\theta^*).$$

Then for any θ with $\|\theta - \theta^*\| \leq \gamma$, the following quadratic upper bound holds:

$$\mathcal{R}(\theta) \ \leq \ \epsilon + \tfrac{\beta_\ell}{2}\, \|\theta - \theta^*\|^2 \ \leq \ \epsilon + R^2\, \gamma^2.$$

Proof. The first part (non-negativity, convexity, differentiability, local Lipschitzness, and smoothness) follows directly by applying Lemmas 3 and 5 under Assumption 1, which together show each $\ell(z, \theta)$ is β_ℓ-smooth with $\beta_\ell = 2R^2$.

Since $f(\theta) = \mathbb{E}_z[\ell(z, \theta)]$, differentiability and smoothness carry over to f, and we have

$$\|\nabla f(\theta) - \nabla f(\theta^*)\| \le \beta_\ell \|\theta - \theta^*\|.$$

At the minimizer θ^*, $\nabla f(\theta^*) = 0$. Hence for any θ with $\|\theta - \theta^*\| \le \gamma$, the gradient satisfies

$$\|\nabla f(\theta)\| \le \beta_\ell \|\theta - \theta^*\| \le \beta_\ell \gamma.$$

Applying the standard smoothness inequality (second-order Taylor upper bound) around θ^*,

$$f(\theta) \le f(\theta^*) + \langle \nabla f(\theta^*), \theta - \theta^* \rangle + \tfrac{\beta_\ell}{2} \|\theta - \theta^*\|^2 = \epsilon + \tfrac{\beta_\ell}{2} \|\theta - \theta^*\|^2,$$

where we used $\nabla f(\theta^*) = 0$ and $f(\theta^*) = \epsilon$. Finally, substituting $\beta_\ell = 2R^2$ gives

$$f(\theta) \le \epsilon + R^2 \|\theta - \theta^*\|^2 \le \epsilon + R^2 \gamma^2,$$

completing the proof. □

Lemma 5 (Regularity of the multiclass cross-entropy loss)**.** *Let $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Y} = \{1, \dots, K\}$, $W \subset \mathbb{R}^{K \times d}$ compact, and let $\|x\| \le R$ for all $x \in \mathcal{X}$. Define for $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $\Theta = (\theta_1, \dots, \theta_K) \in W$*

$$\ell(x, y; \Theta) \;=\; -\log \frac{\exp(\theta_y^\top x)}{\sum_{k=1}^K \exp(\theta_k^\top x)}.$$

Then for all x, y, Θ_1, Θ_2 as above:

(i) *ℓ is nonnegative, C^∞ in Θ, convex, and β–smooth with $\beta \;=\; R^2$.*

(ii) *ℓ is (locally) Lipschitz on W:*

$$\big|\ell(x, y; \Theta_1) - \ell(x, y; \Theta_2)\big| \;\le\; L\,\|\Theta_1 - \Theta_2\|, \qquad L = \sqrt{2}\,R.$$

Proof. The proofs of each subpart are as follows:

(i) Let $p = \mathrm{softmax}(\Theta\, x) \in \Delta^{K-1}$. One checks

$$\nabla_\Theta \ell = (p - e_y)\, x^\top, \qquad \nabla^2_\Theta \ell = \big[\mathrm{diag}(p) - p\, p^\top\big] \;\otimes\; \big(x\, x^\top\big).$$

Since $\mathrm{diag}(p) - p\, p^\top \succeq 0$ $\ell \ge 0$, is C^∞ and convex, and

$$\|\nabla^2_\Theta \ell\| \;\le\; \|x\|^2 \,\lambda_{\max}\big(\mathrm{diag}(p) - p\, p^\top\big) \;\le\; R^2,$$

it follows that ℓ is β–smooth with $\beta = R^2$.

(ii) By the mean-value theorem,

$$\big|\ell(\Theta_1) - \ell(\Theta_2)\big| \;\le\; \sup_{\Theta \in W} \|\nabla_\Theta \ell\|\; \|\Theta_1 - \Theta_2\|.$$

But

$$\|\nabla_\Theta \ell\| = \|p - e_y\|\; \|x\| \;\le\; \sqrt{2}\,R,$$

giving the claimed Lipschitz constant $L = \sqrt{2}\,R$.

□

Lemma 6. *Let $\{h_\theta\}_{\theta \in \Theta}$ be the family of softmax classifiers mapping $x \in \mathbb{R}^d$ to a distribution over C classes,*

$$h_\theta(x)_c = \frac{e^{x^\top \theta_c}}{\sum_{j=1}^C e^{x^\top \theta_j}}.$$

Assume:

(i) $\|x\|_2 \le R$ *for all* x *in the support of* $\mathcal{P}$.

(ii) There exists $\alpha > 0$ *such that for all* x, θ, c, *we have* $h_\theta(x)_c \ge \alpha$.

(iii) θ, θ^* *lie in a convex set* $\Theta \subset \mathbb{R}^{C\times d}$.

Then the mapping $\theta \mapsto h_\theta(x)$ *is Lipschitz in the* ℓ_1*–norm: for all* $\theta, \theta^* \in \Theta$,

$$\|h_\theta(x) - h_{\theta^*}(x)\|_1 \;\le\; L_h\,\|\theta-\theta^*\|_2, \quad L_h = R\sqrt{2\,(1-\alpha)}.$$

Proof. By the multivariate mean-value theorem, there exists $\bar\theta$ on the line segment between θ^* and θ such that

$$h_\theta(x) - h_{\theta^*}(x) = D_\theta h(\bar\theta; x)\,\big(\theta-\theta^*\big),$$

where $D_\theta h(\bar\theta;x) \in \mathbb{R}^{C\times(Cd)}$ is the Jacobian w.r.t. all parameters. Write its cth row as the gradient vector $\nabla_\theta h_c(\bar\theta;x)$. A direct calculation shows for each class c and each block k:

$$\nabla_{\theta_k} h_c(\bar\theta;x) = h_c\big(\mathbf{1}\{c=k\} - h_k\big)\,x \quad\Longrightarrow\quad \|\nabla_\theta h_c(\bar\theta;x)\|_2 \le \|x\|_2\, h_c\,\sqrt{(1-h_c)^2 + \sum_{k\neq c} h_k^2}.$$

Using $\sum_k h_k = 1$, $h_k \le 1$ for each k, and the lower-bound $h_c \ge \alpha$, one checks

$$(1-h_c)^2 + \sum_{k\neq c} h_k^2 \le (1-h_c)^2 + (1-h_c) = (1-h_c)(2-h_c) \le 2\,(1-h_c) \le 2\,(1-\alpha),$$

where the first inequality uses $\sum_{k\neq c} h_k^2 \le \sum_{k\neq c} h_k = 1-h_c$. Hence

$$\|\nabla_\theta h_c(\bar\theta;x)\|_2 \le R\,h_c\,\sqrt{2\,(1-\alpha)}.$$

Therefore the operator norm from ℓ_2 to ℓ_1 satisfies

$$\|D_\theta h(\bar\theta;x)\|_{2\to 1} \;=\; \sum_{c=1}^{C} \|\nabla_\theta h_c(\bar\theta;x)\|_2 \;\le\; R\sqrt{2(1-\alpha)}\sum_{c=1}^{C} h_c = R\sqrt{2(1-\alpha)}.$$

Putting these together gives

$$\|h_\theta(x) - h_{\theta^*}(x)\|_1 \le \|D_\theta h(\bar\theta;x)\|_{2\to1}\,\|\theta-\theta^*\|_2 \le L_h\,\|\theta-\theta^*\|_2,$$

as claimed. □

Lemma 7. *Under Assumption 1, for all* $z \in \mathcal{Z}$, *the loss* $\ell(z,\cdot)$ *is non-negative, convex, differentiable, locally Lipschitz, and* β_ℓ*-smooth.*

Proof. The result follows by applying Lemmas 3 and 5 with $|x| \le R$ and choosing $\beta_\ell = \max 2R^2, R^2 = 2R^2$. Each loss is nonnegative, convex, differentiable, locally Lipschitz, and smooth under these bounds. □

C BAD OUTCOME

Notation Recall the potential function from the main text:

$$f(\Theta) = \sum_{i=1}^{m}\left(\tau\alpha_i \mathop{\mathbb{E}}_{z\sim\mathcal{P}_i}[\ell(z,\theta_i)] + (1-\tau)a_i(\Theta)\mathop{\mathbb{E}}_{z\sim\mathcal{D}_i(\Theta)}[\ell(z,\theta_i)]\right) \tag{4}$$

From Lemma 1, we have that the MSGD iterates converge almost surely to the set $\{\nabla f(\Theta) = 0\}$. Additionally, for learner $i \in [m]$, define $\bar\theta_i$ as the solution that each learner would choose if they trained independently on the subpopulation of users who rank them highest:

$$\bar\theta_i = \arg\min_{\theta_i} \mathop{\mathbb{E}}_{x\sim\mathcal{P}_i} \ell(x,\theta_i) \tag{5}$$

Example 1. *Specify the family of instances $\mathcal{G}_{bad}(\tau, C)$ as follows.*

1. *Distribution $\mathcal{P}$: is defined as a mixture of subpopulations. $\mathcal{P} = \alpha\mathcal{P}_1 + (1-\alpha)\mathcal{P}_2$. Here, $\{\mathcal{P}_1, \mathcal{P}_2\}$ are 2 subpopulations. For either subpopulation, the covariates are generated from the zero mean and unit-variance uniform distribution:*

$$x \sim \textit{Unif}([-\sqrt{3}, \sqrt{3}]) \quad \textit{for} \quad (x, y) \sim \mathcal{P}_i, \tag{6}$$

   For each subpopulation, the response variable is generated as:

$$y = Cx \quad \textit{for} \quad (x, y) \sim \mathcal{P}_1 \tag{7}$$

$$y = -x \quad \textit{for} \quad (x, y) \sim \mathcal{P}_2 \tag{8}$$

2. *Loss function: $\ell(x, y, \theta) = (y - \theta^T x)^2$ is the squared loss*
3. *Ranking $\pi(z) = i$ for $(x, y) \sim \mathcal{P}_i$.*

Lemma 8. *Consider the 1-D bad-outcome family in Example 1 with mixture weight $\alpha \in (0, 1)$ and slope parameter $C > 1$, and let $\mathcal{R}(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{P}}[(y - \theta x)^2]$.*

(i) The least-squares predictor on the full mixture, $\theta^\star = \alpha C - (1-\alpha)$, satisfies

$$\mathcal{R}(\theta^\star) = \alpha(1-\alpha)(C+1)^2.$$

(ii) The specialist trained on $\mathcal{P}_1$ is $\bar{\theta}_1 = C$, and its mixture risk is

$$\mathcal{R}(\bar{\theta}_1) \;=\; (1-\alpha)(C+1)^2.$$

(iii) The specialist trained on $\mathcal{P}_2$ is $\bar{\theta}_2 = -1$, and its mixture risk is

$$\mathcal{R}(\bar{\theta}_2) \;=\; \alpha\,(C+1)^2.$$

Proof. We have that $\mathbb{E}[x] = 0$ and $\mathbb{E}[x^2] = 1$, so that the squared risk reduces to $(\beta - \theta)^2$ when the true slope is β.

(i) Global compromise. On the mixture, the conditional label is linear: $y = \beta x$ with $\beta = \alpha C + (1-\alpha)(-1) = \alpha C - (1-\alpha)$. For squared loss with $\mathbb{E}[x] = 0$ and $\mathbb{E}[x^2] = 1$, the population least-squares minimizer is the regression slope, so $\theta^\star = \beta$. The corresponding risk is

$$\mathcal{R}(\theta^\star) = \alpha\,(C-\beta)^2 + (1-\alpha)\,(-1-\beta)^2 = \alpha(1-\alpha)(C+1)^2.$$

(ii) Specialist for $\mathcal{P}_1$. Training least squares on $\mathcal{P}_1$ alone yields $\bar{\theta}_1 = C$ since $\mathbb{E}[x\,y] = C$ and $\mathbb{E}[x^2] = 1$. The mixture risk is

$$\mathcal{R}(\bar{\theta}_1) = \alpha\,(C-C)^2 + (1-\alpha)\,(-1-C)^2 = (1-\alpha)(C+1)^2.$$

(iii) Specialist for $\mathcal{P}_2$. Training least squares on $\mathcal{P}_2$ gives $\bar{\theta}_2 = -1$ (its true slope). The mixture risk is

$$\mathcal{R}(\bar{\theta}_2) = \alpha\,(C+1)^2 + (1-\alpha)\,(-1-(-1))^2 = \alpha\,(C+1)^2.$$

□

Lemma 9. *Let $(\tilde{\theta}_1, \tilde{\theta}_2) \in \{\nabla f(\Theta) = 0\}$ be a stationary point of $f(\Theta)$ (Equation 1). Then, under the assumption that $\tau \geq \frac{1}{2}$, it must be true that $\mathcal{D}_1(\tilde{\theta}_1, \tilde{\theta}_2) = \mathcal{P}_1$ and $\mathcal{D}_2(\tilde{\theta}_1, \tilde{\theta}_2) = \mathcal{P}_2$.*

Proof. The proof proceeds by considering the decisions of the loss-minimizing users, and shows that the stationary point condition implies the conditions on $\mathcal{D}_i(\Theta)$ in the lemma statement. Consider any (x, y) from $\mathcal{P}_1$. For any $\theta \in \mathbb{R}$, we have that

$$\ell(x, y, \theta) = (Cx - \theta x)^2 = (C-\theta)^2 x^2. \tag{9}$$

Hence,

$$\arg\min_{i\in\{1,2\}} \ell(x, y, \theta_i) = \arg\min_{i\in\{1,2\}} (C-\theta_i)^2 x^2 = \arg\min_{i\in\{1,2\}} |C-\theta_i|.$$

Thus every loss-minimizing user in $\mathcal{P}_1$ picks the learner whose parameter is closer to C:

$$M(z;\Theta) \;=\; \arg\min_{i\in\{1,2\}} |C-\theta_i|, \quad z\in\mathcal{P}_1.$$

Likewise, every loss-minimizing user in $\mathcal{P}_2$ picks the learner whose parameter is closer to -1:

$$M(z;\Theta) \;=\; \arg\min_{i\in\{1,2\}} |-1-\theta_i|, \quad z\in\mathcal{P}_2.$$

Hence, within each subpopulation $\mathcal{P}_i$, all loss-minimizing users make the same choice. Consequently, for any Θ, the observed-data distributions $\mathcal{D}_i(\Theta)$ can only take one of the following four forms:

1. $\mathcal{D}_1(\Theta) = \mathcal{P}$ and $\mathcal{D}_2(\Theta)$ has zero mass under $\mathcal{P}$
2. $\mathcal{D}_1(\Theta)$ has zero mass under $\mathcal{P}$ and $\mathcal{D}_2(\Theta) = \mathcal{P}$
3. $\mathcal{D}_1(\Theta) = \mathcal{P}_2$ and $\mathcal{D}_2(\Theta) = \mathcal{P}_1$
4. $\mathcal{D}_1(\Theta) = \mathcal{P}_1$ and $\mathcal{D}_2(\Theta) = \mathcal{P}_2$

For any stationary point $(\tilde\theta_1, \tilde\theta_2)$, we consider each case separately and show that all cases other than the last one lead to a contradiction.

Case 1: In this case, the second learner observes only negative labels, and the first observes a mix of both positive and negative labels. Hence, it should be intuitively true that $\tilde\theta_1 > \tilde\theta_2 > -1$. This would lead to a contradiction since this would imply that users from $\mathcal{P}_2$ would strictly prefer the second learner over the first one. Now, we can verify this formally. Define the total probability masses seen by each learner

$$M_1 = \tau\,\alpha \;+\; (1-\tau)\,\alpha \;+\; (1-\tau)(1-\alpha) \;=\; \alpha \;+\; (1-\tau)(1-\alpha), \tag{10}$$
$$M_2 = \tau\,(1-\alpha). \tag{11}$$

The stationary conditions are given by

$$\tilde\theta_1 = \arg\min_{\theta_1} \tau\,\alpha \mathop{\mathbb{E}}_{P_1}\big[(y-\theta_1\,x)^2\big] + (1-\tau)\,\alpha \mathop{\mathbb{E}}_{P_1}\big[(y-\theta_1\,x)^2\big] + (1-\tau)(1-\alpha) \mathop{\mathbb{E}}_{P_2}\big[(y-\theta_1\,x)^2\big], \tag{12}$$

$$\tilde\theta_2 = \arg\min_{\theta_2} \tau\,(1-\alpha) \mathop{\mathbb{E}}_{P_2}\big[(y-\theta_2\,x)^2\big]. \tag{13}$$

We can solve these optimization problems in closed form:

$$\tilde\theta_1 = \frac{\alpha\,C \;-\; (1-\tau)(1-\alpha)}{M_1}, \tag{14}$$
$$\tilde\theta_2 = -1. \tag{15}$$

Now,

$$\tilde\theta_1 - \tilde\theta_2 = \tilde\theta_1 + 1 = \frac{\alpha\,(C+1)}{M_1} > 0,$$

which holds for every $\tau \in [0,1]$ since $\alpha > 0$ and $C > 0$. Since $\tilde\theta_2 = -1$ and $\tilde\theta_1 > -1$, $\mathcal{P}_2$ users would strictly prefer learner 2 over learner 1, leading to a contradiction.

Case 2: In this case, the first learner observes only positive labels, and the second observes a mix of both positive and negative labels. Hence, it is easy to verify using a similar procedure as the previous case that $C > \tilde\theta_1 > \tilde\theta_2$, which would lead to a contradiction since $\mathcal{P}_1$ users would strictly prefer learner 1.

Case 3: We follow a similar argument as in Case 1. Define the total probability masses seen by each learner

$$M_1 = \tau\,\alpha \;+\; (1-\tau)(1-\alpha), \tag{16}$$
$$M_2 = \tau\,(1-\alpha) \;+\; (1-\tau)\,\alpha. \tag{17}$$

The stationary points in this case are given by

$$\tilde{\theta}_1 \;=\; \frac{\tau\,\alpha\,C \;-\; (1-\tau)(1-\alpha)}{\tau\,\alpha \;+\; (1-\tau)(1-\alpha)}, \qquad \tilde{\theta}_2 \;=\; \frac{(1-\tau)\,\alpha\,C \;-\; \tau(1-\alpha)}{(1-\tau)\,\alpha \;+\; \tau(1-\alpha)}.$$

It is easy to verify that both $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are greater than -1. Hence, the distance from -1 is given by

$$d_1 = \left|\tilde{\theta}_1 - (-1)\right| = \frac{\tau\,\alpha\,(C+1)}{\tau\,\alpha \;+\; (1-\tau)(1-\alpha)},$$
$$d_2 = \left|\tilde{\theta}_2 - (-1)\right| = \frac{(1-\tau)\,\alpha\,(C+1)}{(1-\tau)\,\alpha \;+\; \tau(1-\alpha)}.$$

A straightforward comparison gives

$$d_1 > d_2 \iff \tau\,M_2 > (1-\tau)\,M_1 \iff (1-\alpha)\,(2\tau-1) > 0 \iff \tau > \tfrac{1}{2}.$$

Therefore, whenever $\tau > \frac{1}{2}$, the free users from $\mathcal{P}_2$ strictly prefer learner 2 over learner 1. This contradicts the Case 3 hypothesis that $\mathcal{D}_1(\Theta) = \mathcal{P}_2$. Hence Case 3 cannot be a stationary point when $\tau > 1/2$. Having ruled out Cases 1–3, only the fourth configuration remains, completing the proof. □

Theorem 2. *Let Assumptions 1–4 hold. For any $\tau \geq \frac{1}{2}$ and any choice of ϵ, Γ with $0 < \epsilon < \Gamma$, there exists an instance $\mathcal{G} = (\mathcal{P}, \ell, \pi, \tau)$ such that:*

1. *There exists $\theta^\star$ with $\mathcal{R}(\theta^\star) \leq \epsilon$.*
2. *The MSGD iterates converge to a unique stationary point $\bar{\Theta}$ where $\mathcal{R}(\bar{\theta}_i) \geq \Gamma$ for some learner $i \in [m]$.*

Proof. The proof follows by considering Example 1.

Part (i) The proof follows from Lemma 8. Choosing $\alpha = \frac{\epsilon}{(C+1)^2}$ satisfies the condition.

Characterizing the stationary points From Lemma 1, we have that the MSGD iterates converge almost surely to the set $\{\nabla f(\Theta) = 0\}$. By Lemma 9, we have that $(\tilde{\theta}_1, \tilde{\theta}_2) \in \{\nabla f(\Theta) = 0\}$ if and only if $\mathcal{D}_1(\tilde{\theta}_1, \tilde{\theta}_2) = \mathcal{P}_1$ and $\mathcal{D}_2(\tilde{\theta}_1, \tilde{\theta}_2) = \mathcal{P}_2$. Hence, any stationary point must satisfy

$$\tilde{\theta}_1 = \arg\min_{\theta\in\mathbb{R}} \alpha \mathop{\mathbb{E}}_{z\sim\mathcal{P}_1}[\ell(z,\theta)]$$

and

$$\tilde{\theta}_2 = \arg\min_{\theta\in\mathbb{R}} (1-\alpha) \mathop{\mathbb{E}}_{z\sim\mathcal{P}_2}[\ell(z,\theta)]$$

Clearly, $(\bar{\theta}_1, \bar{\theta}_2)$ is the unique solution to these equations, and MSGD must converge to this point almost surely.

Characterizing the loss From Lemma 8, at the stationary point we have $\bar{\theta}_1 = C$ and $\bar{\theta}_2 = -1$, so the mixture risks are

$$\mathcal{R}(\bar{\theta}_1) = (1-\alpha)(C+1)^2, \qquad \mathcal{R}(\bar{\theta}_2) = \alpha(C+1)^2.$$

Moreover, the optimal global (compromise) risk satisfies $\mathcal{R}(\theta^*) \leq \epsilon$ by our choice in Part (i). To ensure learner 1's risk exceeds Γ, choose

$$C \;=\; \sqrt{\Gamma+\epsilon}\; - 1, \qquad \alpha \;=\; \frac{\epsilon}{(C+1)^2}.$$

Then $\alpha \in (0,1)$ and

$$\mathcal{R}(\theta^*) = \alpha(1-\alpha)(C+1)^2 \leq \alpha(C+1)^2 = \epsilon,$$

while

$$\mathcal{R}(\bar{\theta}_1) = (1-\alpha)(C+1)^2 = (C+1)^2 - \epsilon \geq \Gamma.$$

Thus the two claims in the theorem are simultaneously satisfied. □

Algorithm 3 Multi-learner Streaming Gradient Descent

Require: loss function $\ell(\cdot,\cdot) \geq 0$; Initial models $\Theta^0 = (\theta_1^0, \ldots, \theta_m^0)$; Learning rate $\{\eta^t\}_{t=1}^{T+1}$

for $t = 0, 1, 2, \ldots, T$ **do**
  Sample data point $z^t \sim \mathcal{P}$
  User selects model $i = M(z^t; \Theta^t)$
  $\theta_i^{t+1} \leftarrow \theta_i^t - \eta^t \nabla \ell(z^t, \theta_i^t)$
end for
return Θ^T

D CONVERGENCE OF Algorithm 1 AND Algorithm 2

D.1 CONVERGENCE OF ALGORITHM 1

The convergence results for MSGD (Algorithm 1) follow as special cases of the MSGD-P analysis when the probing set $U = \emptyset$. In this case, all probing terms vanish and the augmented potential $\widetilde{f}$ reduces to the original potential f.

Lemma 1. *Let Assumptions 1–4 hold. Then the iterates $\{\Theta^t\}$ of Algorithm 1 converge to a compact connected internally chain transitive invariant set of the ODE $\dot{\Theta} = -\nabla f(\Theta)$.*

Proof. This follows from the proof of Theorem 3 with $U = \emptyset$. When $U = \emptyset$:

- The indicator $\mathbf{1}_{i \in U} = 0$ for all $i \in [m]$, so all probing terms vanish.
- The augmented potential reduces to $\widetilde{f}(\Theta) = f(\Theta)$.
- The ODE (20) becomes $\dot{\Theta} = -\nabla f(\Theta)$.

The stochastic approximation argument proceeds identically: by Lemma 10, verifying local Lipschitzness of $-\nabla f$ (Lemma 7, Lemma 13), the step-size conditions (Assumption 2), the martingale variance bound (Lemma 11 with $U = \emptyset$, so $K_{\text{probe}} = 0$), and bounded iterates (Assumption 4), the iterates converge almost surely to a compact connected internally chain transitive invariant set of $\dot{\Theta} = -\nabla f(\Theta)$. □

Theorem 1. *Let Assumptions 1–4 hold. Then the iterates $\{\Theta^t\}$ of Algorithm 1 converge to the set of stationary points $\{\Theta : \nabla f(\Theta) = 0\}$ almost surely.*

Proof. Follows directly from Theorem 3 with $U = \emptyset$. □

D.2 CONVERGENCE OF MSGD-P (ALGORITHM 2)

Define the empirical probing loss and augmented potential:

$$\hat{L}_i(\theta_i) = \tfrac{1}{n} \sum_{q=1}^{n} \ell(\tilde{z}_i^q, \theta_i), \quad i \in U, \tag{18}$$

$$\widetilde{f}(\Theta) = f(\Theta) + p \sum_{i \in U} \left(\hat{L}_i(\theta_i) + \frac{\lambda}{2} \|\theta_i\|^2 \right). \tag{19}$$

Define the ordinary differential equation (ODE) $\dot{\Theta} = \widetilde{F}(\Theta)$ with:

$$\widetilde{F}_i(\Theta) = -\Big(\tau \alpha_i \mathop{\mathbb{E}}_{z \sim \mathcal{P}_i} [\nabla \ell(z, \theta_i)] + (1-\tau) a_i(\Theta) \mathop{\mathbb{E}}_{z \sim \mathcal{D}_i(\Theta)} [\nabla \ell(z, \theta_i)] + p\, \mathbf{1}_{i \in U} \left(\nabla \hat{L}_i(\theta_i) + \lambda\, \theta_i \right) \Big) \tag{20}$$

Theorem 3. *Let Assumptions 1–3, Assumption 4 (as applied to $\widetilde{f}$), and Assumption 5 hold. Then, the iterates $\{\Theta^t\}$ of Algorithm 2 converge to the set of stationary points $\{\Theta : \nabla \widetilde{f}(\Theta) = 0\}$ almost surely.*

Proof. The proof follows the stochastic approximation template by showing the iterates track the ODE $\dot{\Theta} = \widetilde{F}(\Theta)$ and then using a Lyapunov argument for $\widetilde{f}$.

Let $z_1 \sim \mathcal{P}$, $z_2 \sim \mathcal{P}_i$, and $z_3 \sim \mathcal{D}_i(\Theta^t)$. For the on-platform update, define for each coordinate i the random direction

$$g^t_{i,\text{plat}}(\Theta^t) = \begin{cases} \nabla\ell(z_2, \theta^t_i), & \text{w.p. } \tau\alpha_i, \\ \nabla\ell(z_3, \theta^t_i), & \text{w.p. } (1-\tau)\, a_i(\Theta^t), \\ 0, & \text{otherwise.} \end{cases}$$

For the probing update, for each $j \in U$ sample $\tilde{z}^t_j$ uniformly from the fixed dataset $\mathfrak{D}_j$ and set

$$g^t_{i,\text{probe}}(\Theta^t) = p\,\mathbf{1}_{i\in U}\left(\nabla\ell(\tilde{z}^t_i, \theta^t_i) + \lambda\,\theta^t_i\right).$$

Let $\tilde{g}^t_i(\Theta^t) = g^t_{i,\text{plat}}(\Theta^t) + g^t_{i,\text{probe}}(\Theta^t)$ and $\tilde{g}^t = (\tilde{g}^t_1, \ldots, \tilde{g}^t_m)$. The algorithmic iterate satisfies

$$\Theta^{t+1} = \Theta^t - \eta_t\,\tilde{g}^t(\Theta^t) = \Theta^t - \eta_t\,(\widetilde{F}(\Theta^t) + v^t),$$

where $v^t = \tilde{g}^t(\Theta^t) - \mathbb{E}[\tilde{g}^t(\Theta^t) \mid \mathcal{F}_t]$.

Drift identity: By construction and by uniform sampling from $\mathfrak{D}_i$,

$$\begin{aligned} \mathbb{E}[\tilde{g}^t_i(\Theta^t) \mid \mathcal{F}_t] &= \tau\alpha_i \mathop{\mathbb{E}}_{z\sim\mathcal{P}_i}[\nabla\ell(z, \theta^t_i)] + (1-\tau)\,a_i(\Theta^t) \mathop{\mathbb{E}}_{z\sim\mathcal{D}_i(\Theta^t)}[\nabla\ell(z, \theta^t_i)] \\ &\quad + p\,\mathbf{1}_{i\in U}\left(\nabla\hat{L}_i(\theta^t_i) + \lambda\,\theta^t_i\right) \\ &= -\widetilde{F}_i(\Theta^t). \end{aligned}$$

Below, we verify the assumptions of Lemma 10:

- From Lemma 7 and Lemma 13, the drift $\widetilde{F}(\Theta)$ is locally Lipschitz; the probe part $\theta_i \mapsto p\,(\nabla\hat{L}_i(\theta_i) + \lambda\,\theta_i)$ is a finite average of locally Lipschitz gradients plus a globally Lipschitz linear term.
- From Assumption 2, the step sizes satisfy $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$.
- From Lemma 11, augmented to include the probe term, there exists $K > 0$ such that $\mathbb{E}[\|v^t\|^2 \mid \mathcal{F}_t] \le K(1 + \|\Theta^t\|^2)$.
- From Assumption 4, we have $\sup_t \|\Theta^t\| < \infty$ almost surely.

By Lemma 10, the iterates converge almost surely to a compact, connected, internally chain transitive invariant set of $\dot{\Theta} = \widetilde{F}(\Theta)$. Since $\widetilde{F}(\Theta) = -\nabla\widetilde{f}(\Theta)$, along ODE trajectories

$$\tfrac{d}{dt}\,\widetilde{f}(\Theta(t)) = \langle\nabla\widetilde{f}, \dot{\Theta}\rangle = -\|\nabla\widetilde{f}\|^2 \le 0,$$

with equality iff $\nabla\widetilde{f} = 0$. Thus $\widetilde{f}$ is a strict Lyapunov function and the only invariant sets are stationary points, yielding the claim. □

Lemma 10 (Borkar [2008], Chapter 2, Theorem 2)**.** *Let $\{x_n\}$ be a sequence generated by the stochastic approximation algorithm*

$$x_{n+1} = x_n + a(n)[h(x_n) + M_{n+1}], \quad n \ge 0,$$

where:

1. $h : \mathbb{R}^d \to \mathbb{R}^d$ *is Lipschitz continuous*
2. *The step sizes $\{a(n)\}$ satisfy $\sum_n a(n) = \infty$ and $\sum_n a(n)^2 < \infty$*
3. *$\{M_n\}$ is a martingale difference sequence satisfying $E[||M_{n+1}||^2|\mathcal{F}_n] \le K(1 + ||x_n||^2)$ for some $K > 0$*
4. $\sup_n ||x_n|| < \infty$ *almost surely*

Then almost surely, the sequence $\{x_n\}$ converges to a (possibly sample path dependent) compact connected internally chain transitive invariant set of the ODE

$$\dot{x}(t) = h(x(t)).$$

Lemma 11 (Martingale variance bound)**.** *Suppose that Assumption 1 holds and that the probing datasets $\{\mathfrak{D}_i\}_{i\in U}$ are fixed finite sets. Let $g^t(\Theta^t) = (g_1^t(\Theta^t), \ldots, g_m^t(\Theta^t))$ be the stochastic update vector defined by*

$$g_i^t(\Theta^t) \;=\; g_{i,\text{plat}}^t(\Theta^t) \;+\; g_{i,\text{probe}}^t(\Theta^t),$$

where

$$g_{i,\text{plat}}^t(\Theta^t) = \begin{cases} \nabla\ell(z^t, \theta_i^t), & \text{if } i = i_t, \\ 0, & \text{if } i \neq i_t, \end{cases} \qquad g_{i,\text{probe}}^t(\Theta^t) = p\,\mathbf{1}_{i\in U}\left(\nabla\ell(\tilde{z}_i^t, \theta_i^t) + \lambda\,\theta_i^t\right),$$

with $\tilde{z}_i^t$ sampled uniformly from $\mathfrak{D}_i$ independently of i_t. Define the filtration $\mathcal{F}_t = \sigma(\Theta^0, z^1, \ldots, z^{t-1}, \{\mathfrak{D}_i\}_{i\in U})$, which contains all information available prior to time t. Define the martingale difference sequence:

$$v^t = g^t(\Theta^t) - \mathbb{E}[g^t(\Theta^t) \mid \mathcal{F}_t].$$

Then there exists a constant $K > 0$ (made explicit below) such that:

$$\mathbb{E}[\|v^t\|^2 \mid \mathcal{F}_t] \leq K\,(1 + \|\Theta^t\|^2).$$

Proof. Gradient bounds. From Lemmas 3 and 5, for all (x, y) with $\|x\| \leq R$ and all θ we have bounds of the form

$$\|\nabla_\theta \ell(x, y; \theta)\| \leq A_0 + A_1\,\|\theta\|,$$

with

$$(A_0, A_1) = \begin{cases} (2R\,Y_{\max},\ 2R^2), & \text{(squared loss)}, \\ (\sqrt{2}\,R,\ 0), & \text{(cross-entropy)}. \end{cases}$$

For the probe term under squared loss, labels are pseudo-labels $\hat{y} = \text{median}\{\langle x, \theta_0^j\rangle : j \in [m]\}$ so that $|\hat{y}| \leq R\,\max_{j\in[m]}\|\theta_0^j\| =: Y_{\max,\text{probe}}$, giving $\|\nabla_\theta \ell(x, \hat{y}; \theta)\| \leq \widetilde{A}_0 + A_1\,\|\theta\|$ with $\widetilde{A}_0 := 2R\,Y_{\max,\text{probe}}$. Including the L2 regularization term (which appears only in the probe update) and using the triangle inequality,

$$\|p\,(\nabla_\theta \ell(x, \hat{y}; \theta) + \lambda\,\theta)\| \leq p\,\widetilde{A}_0 + p\,(A_1 + \lambda)\,\|\theta\|.$$

Define the block constants

$$K_{\text{plat}} := 2\max\bigl(A_0^2, A_1^2\bigr), \qquad K_{\text{probe}} := 2\max\bigl((p\,\widetilde{A}_0)^2, \bigl(p\,(A_1 + \lambda)\bigr)^2\bigr),$$

where for cross-entropy we take $A_1 = 0$ and use the same $A_0 = \sqrt{2}R$ for both platform and probe.

Variance decomposition. By definition of v^t and $\text{Var}(Y) = \mathbb{E}[\|Y\|^2] - \|\,\mathbb{E}[Y]\|^2$,

$$\mathbb{E}[\|v^t\|^2 \mid \mathcal{F}_t] \;\leq\; \mathbb{E}[\|g^t(\Theta^t)\|^2 \mid \mathcal{F}_t].$$

We bound the RHS. Since g^t has at most one nonzero platform block and up to $|U|$ nonzero probe blocks,

$$\|g^t(\Theta^t)\|^2 = \sum_{i=1}^m \|g_{i,\text{plat}}^t + g_{i,\text{probe}}^t\|^2 \;\leq\; 2\sum_{i=1}^m \|g_{i,\text{plat}}^t\|^2 + 2\sum_{i=1}^m \|g_{i,\text{probe}}^t\|^2.$$

Taking conditional expectations and using the selection probabilities for the platform block as in the MSGD lemma,

$$\begin{aligned} \mathbb{E}[\|g^t(\Theta^t)\|^2 \mid \mathcal{F}_t] \leq 2\sum_{i=1}^m \mathbb{P}(i_t = i \mid \mathcal{F}_t)\ \mathbb{E}[\|\nabla\ell(z^t, \theta_i^t)\|^2 \mid \mathcal{F}_t, i_t = i] \\ + 2\sum_{j\in U} \mathbb{E}[\|p\,\nabla\ell(\tilde{z}_j^t, \theta_j^t)\|^2 \mid \mathcal{F}_t]. \end{aligned}$$

Applying the block bounds gives

$$\begin{aligned} \mathbb{E}[\|g^t(\Theta^t)\|^2 \mid \mathcal{F}_t] &\leq 2K_{\text{plat}}\sum_{i=1}^m \mathbb{P}(i_t = i \mid \mathcal{F}_t)\,(1 + \|\theta_i^t\|^2) \\ &\quad + 2K_{\text{probe}}\sum_{j\in U}(1 + \|\theta_j^t\|^2) \\ &\leq 2K_{\text{plat}}\,(1 + \|\Theta^t\|^2) \;+\; 2|U|K_{\text{probe}} \;+\; 2K_{\text{probe}}\,\|\Theta^t\|^2 \\ &= 2\bigl(K_{\text{plat}} + |U|K_{\text{probe}}\bigr) \;+\; 2\bigl(K_{\text{plat}} + K_{\text{probe}}\bigr)\,\|\Theta^t\|^2. \end{aligned}$$

Therefore, setting

$$K := 2\max\{K_{\text{plat}} + |U| K_{\text{probe}},\ K_{\text{plat}} + K_{\text{probe}}\}$$

yields

$$\mathbb{E}[\|v^t\|^2 \mid \mathcal{F}_t] \le K\,(1 + \|\Theta^t\|^2).$$

This completes the proof. ☐

Lemma 12 (Gradient of the augmented objective)**.** *For every $i \in [m]$ the gradient of $\widetilde{f}$ with respect to θ_i is*

$$\nabla_{\theta_i} \widetilde{f}(\Theta) = \tau \alpha_i \underset{z\sim\mathcal{P}_i}{\mathbb{E}}[\nabla \ell(z, \theta_i)] + (1-\tau)\, a_i(\Theta) \underset{z\sim\mathcal{D}_i(\Theta)}{\mathbb{E}}[\nabla \ell(z, \theta_i)] \\ + p\,\mathbf{1}_{i\in U}\left(\nabla \hat{L}_i(\theta_i) + \lambda\,\theta_i\right).$$

Proof. We treat a single index i; all other coordinates of Θ are held fixed. The term $(1-\tau) a_i(\Theta)\,\mathbb{E}_{z\sim\mathcal{D}_i(\Theta)}[\ell(z,\theta_i)]$ depends on θ_i both explicitly (inside ℓ) and implicitly through $a_i(\Theta)$ and $\mathcal{D}_i(\Theta)$. Lemma 4.3 of Su and Dean [2024] proves that for any differentiable ℓ satisfying the stated regularity,

$$\nabla_{\theta_i}\Big(a_i(\Theta) \underset{z\sim\mathcal{D}_i(\Theta)}{\mathbb{E}}[\ell(z,\theta_i)]\Big) = a_i(\Theta) \underset{z\sim\mathcal{D}_i(\Theta)}{\mathbb{E}}[\nabla \ell(z,\theta_i)].$$

Multiplying by $(1-\tau)$ yields the corresponding contribution, while the $\tau\alpha_i$-term is immediate. Local L-Lipschitzness of ℓ ensures the required directional limits.

The probing part depends only on θ_i through the finite average $\hat{L}_i(\theta_i) = \frac{1}{n}\sum_{q=1}^{n} \ell(\tilde{z}_i^q, \theta_i)$ and the regularization term $\frac{\lambda}{2}\|\theta_i\|^2$, so $\nabla_{\theta_i}\big(p\,(\hat{L}_i(\theta_i) + \frac{\lambda}{2}\|\theta_i\|^2)\big) = p\,(\nabla \hat{L}_i(\theta_i) + \lambda\,\theta_i)$ if $i \in U$ and zero otherwise. Summing the contributions gives the stated expression. ☐

Lemma 13 (Local Lipschitz of $a_i(\Theta)$)**.** *Let the standard regularity and boundedness assumptions for the idealized MSGD dynamics hold. Then $a_i(\Theta)$ is locally Lipschitz in Θ.*

Proof. From Lemma 7, we know that the loss function ℓ is locally Lipschitz. In other words: for every compact set $\mathcal{K} \subset \mathbb{R}^{k\times d}$ there is a constant $L_{\mathcal{K}} < \infty$ such that

$$\big|\ell(x,\theta) - \ell(x,\theta')\big| \ \le\ L_{\mathcal{K}}\,\|\theta - \theta'\| \quad \forall\, x \in B(0,R),\ \theta,\theta' \in \mathcal{K}.$$

Fix any compact neighborhood $\mathcal{K}$ containing both Θ and Θ'. Let

$$p_{\max}\ =\ \sup_{x\in B(0,R)} p(x), \qquad L_{\mathcal{K}} \text{ as above.}$$

We will show $\big|a_i(\Theta) - a_i(\Theta')\big| \le C_{\mathcal{K}}\|\Theta - \Theta'\|$ for some $C_{\mathcal{K}}$.

Case $m = 2$**.** With services $i = 1, 2$,

$$a_1(\Theta) - a_1(\Theta') = \int_{X_1(\Theta)\setminus X_1(\Theta')} p(x)\,dx\ -\ \int_{X_1(\Theta')\setminus X_1(\Theta)} p(x)\,dx,$$

so

$$\big|a_1(\Theta) - a_1(\Theta')\big|\ \le\ p_{\max}\Big[\lambda\big(X_1(\Theta)\setminus X_1(\Theta')\big) + \lambda\big(X_1(\Theta')\setminus X_1(\Theta)\big)\Big].$$

For any $x \in X_1(\Theta')\setminus X_1(\Theta)$ we have $\ell(x,\theta_1') < \ell(x,\theta_2')$ and $\ell(x,\theta_2) < \ell(x,\theta_1)$. Hence

$$0\ <\ \ell(x,\theta_1) - \ell(x,\theta_2) = \big[\ell(x,\theta_1) - \ell(x,\theta_1')\big] + \big[\ell(x,\theta_1') - \ell(x,\theta_2')\big] + \big[\ell(x,\theta_2') - \ell(x,\theta_2)\big].$$

Since $x \in X_1(\Theta')$, the middle term satisfies $\ell(x,\theta_1') - \ell(x,\theta_2') \le 0$. The first and third terms are each bounded by $L_{\mathcal{K}}\|\Theta - \Theta'\|$ by Lipschitzness. Thus $\ell(x,\theta_1) - \ell(x,\theta_2) \le 2L_{\mathcal{K}}\|\Theta - \Theta'\|$. Define

$$S\ =\ \big\{\,x : \big|\ell(x,\theta_1) - \ell(x,\theta_2)\big| \le 2L_{\mathcal{K}}\|\Theta - \Theta'\|\big\}.$$

Since $\lambda(S) \leq (2L_{\mathcal{K}}/C)\|\Theta - \Theta'\|$ for some constant C from Assumption 3 and $X_1(\Theta') \setminus X_1(\Theta) \subset S$, we get $\lambda(X_1(\Theta') \setminus X_1(\Theta)) \leq C'\|\Theta - \Theta'\|$. The same argument applies to the other set difference, so altogether

$$\left|a_1(\Theta) - a_1(\Theta')\right| \;\leq\; 4\, p_{\max}\, L_{\mathcal{K}}\, \|\Theta - \Theta'\|.$$

General m. The same pairwise argument shows for each i and $j \neq i$, $\lambda\big(X_i(\Theta) \cap X_j(\Theta')\big) \leq C_{\mathcal{K}}\|\Theta - \Theta'\|$. Summing over all $j \neq i$ gives $\lambda\big(X_i(\Theta) \triangle X_i(\Theta')\big) \leq C''_{\mathcal{K}}\|\Theta - \Theta'\|$, and hence $\left|a_i(\Theta) - a_i(\Theta')\right| \leq p_{\max}\, C''_{\mathcal{K}}\, \|\Theta - \Theta'\|$.

Since all constants depend only on the compact set $\mathcal{K}$, this proves that $a_i(\Theta)$ is Lipschitz on $\mathcal{K}$, i.e. locally Lipschitz in Θ. □

E SQUARED LOSS: PERFORMANCE GUARANTEE

Notation. Let $\{(x_i^q, y_i^q)\}_{q=1}^n$ be the probing sample with true labels y_i^q and pseudo-labels $\tilde{y}_i^q$. Here, the true labels y_i^q are hidden from the learner, and the pseudo-labels $\tilde{y}_i^q$ are observed. For any $\theta \in \mathbb{R}^d$, define

$$\widehat{L}_i(\theta) := \frac{1}{n}\sum_{q=1}^{n}\big(\langle x_i^q, \theta\rangle - \tilde{y}_i^q\big)^2,$$

and

$$\widehat{L}_{i,\mathrm{true}}(\theta) := \frac{1}{n}\sum_{q=1}^{n}\big(\langle x_i^q, \theta\rangle - y_i^q\big)^2.$$

Additionally, define the empirical pseudo–true discrepancy

$$\Delta_i^2 := \frac{1}{n}\sum_{q=1}^{n}\big(\tilde{y}_i^q - y_i^q\big)^2.$$

E.1 PROOF OF Lemma 2

We restate the lemma for ease of reference, and then provide the proof.

Lemma 2. *For each scenario in Definitions 2–3, the probing rule $T_i(x)$ in Table 1 satisfies Assumption 5 with the stated accuracy parameter B.*

Proof. We treat each scenario in turn.

(i) Majority-good. Fix any x with $\|x\| \leq R$. Write the peer deviations $u_j := \langle x, \theta_j^0\rangle - \langle x, \theta^\star\rangle$. If strictly more than half of the peers satisfy $\|\theta_j^0 - \theta^\star\| \leq r$, then at least half of the $\{u_j\}$ lie in $[-Rr, Rr]$, so the median obeys $|\tilde{y}(x) - \langle x, \theta^\star\rangle| \leq Rr$ and hence

$$(\tilde{y} - y)^2 \;=\; \big(\tilde{y} - \langle x, \theta^\star\rangle + \langle x, \theta^\star\rangle - y\big)^2 \;\leq\; 2\,(\tilde{y} - \langle x, \theta^\star\rangle)^2 \;+\; 2\,(\langle x, \theta^\star\rangle - y)^2 \;\leq\; 2R^2r^2 \;+\; 2\,(\langle x, \theta^\star\rangle - y)^2.$$

Taking expectation over $(x, y) \sim \mathcal{P}$ yields $\mathbb{E}[(\tilde{y} - y)^2] \leq 2R^2r^2 + 2\,\mathbb{E}[(\langle x, \theta^\star\rangle - y)^2] = 2R^2r^2 + 2\epsilon$.

(ii) Market-leader. Here $\tilde{y}(x) = x^\top\theta_{j^*}$ and by assumption $\mathbb{E}\left[(\tilde{y} - y)^2\right] = \mathbb{E}\left[(x^\top\theta_{j^*} - y)^2\right] \leq \xi$, which directly verifies Accurate Probing with $B = \xi$.

(iii) Partial knowledge. Fix any x with $\|x\| \leq R$. The probing rule $T_i(x) = G$ uses median aggregation over the subset $G \subseteq [m]\setminus\{i\}$. Write the peer deviations for $j \in G$: $u_j := \langle x, \theta_j^0\rangle - \langle x, \theta^\star\rangle$. Since all learners in G satisfy $\|\theta_j^0 - \theta^\star\| \leq r$, all deviations $\{u_j\}_{j\in G}$ lie in $[-Rr, Rr]$. Because $|G| > (m-1)/2$, the set G contains more than half of the peers (excluding i), and thus the median over G obeys $|\tilde{y}(x) - \langle x, \theta^\star\rangle| \leq Rr$. The remainder of the proof follows identically to case (i):

$$(\tilde{y} - y)^2 \;=\; \big(\tilde{y} - \langle x, \theta^\star\rangle + \langle x, \theta^\star\rangle - y\big)^2 \;\leq\; 2\,(\tilde{y} - \langle x, \theta^\star\rangle)^2 \;+\; 2\,(\langle x, \theta^\star\rangle - y)^2 \;\leq\; 2R^2r^2 \;+\; 2\,(\langle x, \theta^\star\rangle - y)^2.$$

Taking expectation over $(x, y) \sim \mathcal{P}$ yields $\mathbb{E}[(\tilde{y} - y)^2] \le 2R^2r^2 + 2\ \mathbb{E}[(\langle x, \theta^\star\rangle - y)^2] = 2R^2r^2 + 2\epsilon$.

(iv) Preference-aware. By the probing rule $T_i(x) = \{\pi(x)\}$, we have $\tilde{y}(x) = x^\top\bar{\theta}_{\pi(x)}$, so

$$\mathbb{E}\left[(\tilde{y} - y)^2\right] \ = \ \sum_{i=1}^{m} \alpha_i \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{P}_i}\left[(y - x^\top\bar{\theta}_i)^2\right].$$

By optimality of $\bar{\theta}_i$ for the ERM objective on $\mathcal{P}_i$, for every i and any θ (in particular $\theta^\star$),

$$\mathop{\mathbb{E}}_{\mathcal{P}_i}\left[(y - x^\top\bar{\theta}_i)^2\right] \ \le \ \mathop{\mathbb{E}}_{\mathcal{P}_i}\left[(y - x^\top\theta^\star)^2\right].$$

Summing over i with weights α_i gives

$$\mathbb{E}\left[(\tilde{y} - y)^2\right] \ \le \ \sum_{i=1}^{m} \alpha_i \mathop{\mathbb{E}}_{\mathcal{P}_i}\left[(y - x^\top\theta^\star)^2\right] \ = \ \epsilon.$$

□

E.2 PROOF OF Theorem 4

Proof of Corollary 1. Set $\lambda = \epsilon/\|\theta^\star\|^2$ in Lemma 14. Define

$$S(\kappa, \lambda) := (4b_\star + 6C_0^2)\sqrt{2\log(2/\kappa)} + 4(Y_{\max} + B_\theta R)B_\theta R + b_{\mathrm{u}}\sqrt{2\log(2/\kappa)},$$

where B_θ and b_{u} are the λ-dependent quantities from Lemma 14. If

$$n \ \ge \ \underline{n} \ := \ \frac{S(\kappa, \lambda)^2}{\epsilon^2},$$

then the concentration terms in Lemma 14 satisfy $S(\kappa, \lambda)/\sqrt{n} \le \epsilon$. With this choice and $\lambda\|\theta^\star\|^2 = \epsilon$, the explicit bound yields

$$\mathcal{R}(\tilde{\theta}_i) \ \le \ 6B + \left(4 + \frac{2}{p}\right)\epsilon + \epsilon + \epsilon \ = \ 6B + \left(6 + \frac{2}{p}\right)\epsilon,$$

which is $O\Big(\big(\frac{p+1}{p}\big)\epsilon + B\Big)$.

To see the stated scaling of $\underline{n}$, note that $B_\theta = \Theta(1/\sqrt{\lambda})$ and $b_{\mathrm{u}} = (Y_{\max} + B_\theta R)^2 = O(1/\lambda)$ with constants depending on $(R, Y_{\max}, M_0, p)$. To surface the dominant dependence on $(R, Y_{\max}, M_0, p)$, write $A := Y_{\max}^2 + pR^2M_0^2$ so that

$$B_\theta R = R\sqrt{\frac{2A}{\lambda p}} \quad \text{and} \quad (B_\theta R)^2 = \frac{2R^2 A}{\lambda p}.$$

The leading terms (in terms of ϵ) in $S(\kappa, \lambda)$ scale as $(B_\theta R)^2$, so one can take

$$\underline{n} = O\left(\frac{R^4\|\theta^\star\|^4}{\epsilon^4}\left(\frac{Y_{\max}^2}{p} + R^2M_0^2\right)^2 \log\frac{1}{\kappa}\right),$$

where we suppress lower-order terms in ϵ coming from $b_\star$ and C_0. □

Theorem 4. *Let Assumptions 1–5 hold. For any $\kappa \in (0, 1)$, with probability at least $1 - \kappa$, every stationary point $\tilde{\Theta}$ of MSGD-P satisfies, for each probing learner $i \in U$, for the squared loss:*

$$\mathcal{R}(\tilde{\theta}_i) \ \le \ O\Big(\Big(\frac{p+1}{p}\Big)\epsilon + B + \lambda\|\theta^\star\|^2 + \frac{(p+1)C_{gen}}{p\lambda}\sqrt{\frac{\log(1/\kappa)}{n}}\Big).$$

where B is quantified in Table 1 for each scenario, and C_{gen} depends on R, $Y_{\max}$, $\|\theta^\star\|$, $\max_{j\neq i}\|\theta_j^0\|$, with explicit form in the Appendix.

Proof. By Lemma 14, for any $\kappa \in (0,1)$ and with probability at least $1-\kappa$,

$$\mathcal{R}(\tilde{\theta}_i) \;\le\; 6\,B + \Big(4 + \frac{2}{p}\Big)\epsilon + \lambda\,\|\theta^\star\|^2 + \frac{T(\kappa)}{\sqrt{n}},$$

where every $n^{-1/2}$ contribution has been grouped into

$$T(\kappa) := \big(4b_\star + 6C_0^2\big)\sqrt{2\log(2/\kappa)} + b_{\mathrm{u}}\sqrt{2\log(2/\kappa)} + 4(Y_{\max} + B_\theta R)B_\theta R. \tag{21}$$

The terms that depend on λ and p are through $B_\theta = \sqrt{2(Y_{\max}^2 + p\,Y_{\max,\text{probe}}^2)/(\lambda\,p)}$, which also appears inside $b_{\mathrm{u}} = (Y_{\max} + B_\theta R)^2$. Expanding the B_θ-dependent pieces, from (21),

$$T(\kappa) = \big(4b_\star + 6C_0^2\big)\sqrt{2\log(2/\kappa)} + \underbrace{\Big[(Y_{\max} + B_\theta R)^2\sqrt{2\log(2/\kappa)} + 4(Y_{\max} + B_\theta R)B_\theta R\Big]}_{B_\theta\text{-dependent}}.$$

Thus the λ-independent part is already $O\big(\sqrt{\log(1/\kappa)}\big)$, so it remains to control the bracketed term. Expanding gives

$$(Y_{\max} + B_\theta R)^2\sqrt{2\log(2/\kappa)} + 4(Y_{\max} + B_\theta R)B_\theta R = O\Big(\sqrt{\log(1/\kappa)}\,\big[Y_{\max}^2 + Y_{\max}\,B_\theta R + B_\theta^2 R^2\big]\Big).$$

Plugging in the value of B_θ from Lemma 19 and assuming $\lambda < 1$, we can combine with the (λ, p)-independent part of $T(\kappa)$ to get

$$C_{\text{gen}} = R(Y_{\max}^2 + Y_{\max,\text{probe}}^2) + 4b_\star + 6C_0^2$$

□

Lemma 14 (Squared-loss performance Full Statement)**.** *Let Assumption 5 hold with parameter B. Then, for any $\kappa \in (0,1)$, with probability at least $1-\kappa$ over the probing sample, every stationary point $\tilde{\Theta}$ of MSGD-P satisfies, for probing learner $i \in [m]$,*

$$\begin{aligned}\mathcal{R}(\tilde{\theta}_i) \;&\le\; 6\,B + \Big(4 + \frac{2}{p}\Big)\,\epsilon + \lambda\,\|\theta^\star\|^2 \\ &\quad + \frac{(4\,b_\star + 6\,C_0^2)\,\sqrt{2\log(2/\kappa)}}{\sqrt{n}} + \frac{4\,(Y_{\max} + B_\theta R)\,B_\theta R}{\sqrt{n}} + b_{\mathrm{u}}\sqrt{\frac{2\log(2/\kappa)}{n}},\end{aligned}$$

where the constants are defined as follows:

- $C_0 \;:=\; Y_{\max} + Y_{\max,\text{probe}}$.
- $B_\theta \;:=\; \sqrt{\dfrac{2(Y_{\max}^2 + p\,Y_{\max,\text{probe}}^2)}{\lambda\,p}}$ *is a radius (independent of $\theta^\star$) with* $Y_{\max,\text{probe}} \;:=\; R\,M_0,\;\; M_0 \;:=\; \max_{j\neq i}\|\theta_j^0\|$, *which bounds the pseudo-label magnitudes via $|\tilde{y}_i^q| \le Y_{\max,\text{probe}}$.*
- $b_\star \;:=\; \big(Y_{\max} + R\,\|\theta^\star\|\big)^2$.
- $b_{\mathrm{u}} \;:=\; \big(Y_{\max} + B_\theta R\big)^2$.

Proof. We work on the event $\mathcal{E}_{\mathrm{u}} \cap \mathcal{E}_\star$ where $\mathcal{E}_{\mathrm{u}}$ is the event of Lemma 17 (probability $\ge 1-\kappa/2$ with B_θ from Lemma 19) and $\mathcal{E}_\star$ is the event of Lemma 16 (probability $\ge 1-\kappa/2$). By a union bound,

$$\mathbb{P}(\mathcal{E}_{\mathrm{u}} \cap \mathcal{E}_\star) \ge 1-\kappa.$$

Bound on $\widehat{L}_i(\tilde{\theta}_i)$. By stationarity, $\tilde{\theta}_i$ minimizes

$$\hat{\Phi}_i^{\text{probe}}(\theta;\tilde{\Theta}) = \tau\alpha_i \mathop{\mathbb{E}}_{\mathcal{P}_i}[\ell(z,\theta)] + (1-\tau)a_i(\tilde{\Theta}) \mathop{\mathbb{E}}_{\mathcal{D}_i(\tilde{\Theta})}[\ell(z,\theta)] + p\,\widehat{L}_i(\theta) + \tfrac{\lambda p}{2}\|\theta\|^2.$$

Thus, comparing the objective at $\tilde{\theta}_i$ and $\theta^\star$,

$$p\,\widehat{L}_i(\tilde{\theta}_i) \;\le\; \tau\alpha_i \mathop{\mathbb{E}}_{\mathcal{P}_i}[\ell(z,\theta^\star)] + (1-\tau)a_i(\tilde{\Theta}) \mathop{\mathbb{E}}_{\mathcal{D}_i(\tilde{\Theta})}[\ell(z,\theta^\star)] + p\,\widehat{L}_i(\theta^\star) + \tfrac{\lambda\,p}{2}\|\theta^\star\|^2.$$

Since

$$\tau\alpha_i \mathop{\mathbb{E}}_{\mathcal{P}_i} \ell(\theta^\star) + (1-\tau)a_i(\tilde{\Theta}) \mathop{\mathbb{E}}_{\mathcal{D}_i(\tilde{\Theta})} \ell(\theta^\star) \le \epsilon,$$

dividing by p, we obtain

$$\widehat{L}_i(\tilde{\theta}_i) \;\le\; \frac{\epsilon}{p} \;+\; \widehat{L}_i(\theta^\star) \;+\; \frac{\lambda}{2}\,\|\theta^\star\|^2.$$

By Lemma 15 and Lemma 16,

$$\widehat{L}_i(\theta^\star) \le 2\,\widehat{L}_{i,\mathrm{true}}(\theta^\star) + 2\,\Delta_i^2 \le 2\,\epsilon + 2\,\Lambda_\star + 2\,\Delta_i^2.$$

Hence

$$\widehat{L}_i(\tilde{\theta}_i) \;\le\; \frac{\epsilon}{p} + 2\,\epsilon + 2\,\Lambda_\star + 2\,\Delta_i^2 \;+\; \frac{\lambda}{2}\,\|\theta^\star\|^2.$$

Bound on $\widehat{L}_{i,\mathrm{true}}(\tilde{\theta}_i)$. Applying Lemma 15 again gives

$$\widehat{L}_{i,\mathrm{true}}(\tilde{\theta}_i) \;\le\; 2\,\widehat{L}_i(\tilde{\theta}_i) + 2\,\Delta_i^2.$$

Substituting the bound from Step 1,

$$\widehat{L}_{i,\mathrm{true}}(\tilde{\theta}_i) \;\le\; \Big(\tfrac{2}{p} + 4\Big)\epsilon \;+\; 4\,\Lambda_\star \;+\; 6\,\Delta_i^2 \;+\; \lambda\,\|\theta^\star\|^2.$$

By Lemma 18, on $\mathcal{E}_\star$,

$$\Delta_i^2 \le B + C_0^2\sqrt{\frac{2\log(2/\kappa)}{n}}.$$

Therefore,

$$\widehat{L}_{i,\mathrm{true}}(\tilde{\theta}_i) \;\le\; 6\,B \;+\; \Big(\tfrac{2}{p} + 4\Big)\epsilon \;+\; 4\,\Lambda_\star \;+\; 6\,C_0^2\sqrt{\frac{2\log(2/\kappa)}{n}} \;+\; \lambda\,\|\theta^\star\|^2.$$

Bound on $\mathcal{R}(\tilde{\theta}_i)$. Finally, on $\mathcal{E}_\mathrm{u}$ (with $\|\tilde{\theta}_i\| \le B_\theta$ from Lemma 19),

$$\mathcal{R}(\tilde{\theta}_i) \le \widehat{L}_{i,\mathrm{true}}(\tilde{\theta}_i) + \Lambda_\mathrm{u}.$$

□

Alternative scaling with explicit λ. If we keep λ explicit and only choose n to control the concentration terms in Lemma 14, then for any target tolerance $\delta > 0$ it suffices to take

$$n \;\ge\; \bar{n}(\lambda,\delta) \;:=\; \frac{S(\kappa,\lambda)^2}{\delta^2},$$

which yields

$$\mathcal{R}(\tilde{\theta}_i) \;\le\; 6B + \Big(4 + \frac{2}{p}\Big)\epsilon + \lambda\|\theta^\star\|^2 + \delta.$$

Writing $Y_{\max,\mathrm{probe}} = RM_0$ and $A := Y_{\max}^2 + pR^2M_0^2$, we have

$$B_\theta = \sqrt{\frac{2A}{\lambda p}} \quad \text{and} \quad (Y_{\max} + B_\theta R)^2 \le 2Y_{\max}^2 + \frac{4R^2A}{\lambda p}.$$

Using $b_\star = (Y_{\max} + R\|\theta^\star\|)^2$ and $C_0 = Y_{\max} + RM_0$, this implies

$$S(\kappa,\lambda)^2 = O\left(\log\frac{1}{\kappa}\right)\left[(Y_{\max} + R\|\theta^\star\|)^4 + (Y_{\max} + RM_0)^4 + \frac{R^4A^2}{\lambda^2 p^2}\right],$$

and therefore a sufficient choice is

$$n = O\left(\frac{\log(1/\kappa)}{\delta^2}\left[(Y_{\max} + R\|\theta^\star\|)^4 + (Y_{\max} + RM_0)^4 + \frac{R^4(Y_{\max}^2 + pR^2M_0^2)^2}{\lambda^2 p^2}\right]\right).$$

Setting $\delta = \epsilon$ yields the same bias expression as above, but now with n explicit in $(R, Y_{\max}, M_0, p, \lambda, \epsilon, \kappa)$. $\square$

Lemma 15. *We have*

$$\widehat{L}_{i,\mathrm{true}}(\theta) \;\le\; \left(\sqrt{\widehat{L}_i(\theta)} + \Delta_i\right)^2 \quad \textit{and} \quad \widehat{L}_i(\theta) \;\le\; \left(\sqrt{\widehat{L}_{i,\mathrm{true}}(\theta)} + \Delta_i\right)^2.$$

In particular, $\widehat{L}_{i,\mathrm{true}}(\theta) \le 2\,\widehat{L}_i(\theta) + 2\,\Delta_i^2$ *and* $\widehat{L}_i(\theta) \le 2\,\widehat{L}_{i,\mathrm{true}}(\theta) + 2\,\Delta_i^2$.

Proof of Lemma 15. Let $X \in \mathbb{R}^{n\times d}$ be the design matrix, $\mathbf{a} := X\theta - \tilde{\mathbf{y}}$ and $\mathbf{b} := \tilde{\mathbf{y}} - \mathbf{y}$. Then

$$X\theta - \mathbf{y} = \mathbf{a} + \mathbf{b}.$$

By the triangle inequality,

$$\sqrt{\widehat{L}_{i,\mathrm{true}}(\theta)} = \frac{\|X\theta - \mathbf{y}\|}{\sqrt{n}} \le \frac{\|\mathbf{a}\|}{\sqrt{n}} + \frac{\|\mathbf{b}\|}{\sqrt{n}} = \sqrt{\widehat{L}_i(\theta)} + \Delta_i.$$

Squaring yields the first inequality; the second is analogous. $\square$

Lemma 16 (One–point deviation at $\theta^\star$)**.** *For any* $\kappa \in (0,1)$*, with probability at least* $1 - \kappa/2$*,*

$$\left|\widehat{L}_{i,\mathrm{true}}(\theta^\star) - \mathcal{R}(\theta^\star)\right| \;\le\; \Lambda_\star \quad \textit{where} \quad \Lambda_\star \;:=\; b_\star\sqrt{\frac{2\log(2/\kappa)}{n}},\;\; b_\star \;:=\; \left(Y_{\max} + R\,\|\theta^\star\|\right)^2.$$

Proof of Lemma 16. For each q, define the random variable

$$Z_q := (\langle \tilde{x}_i^q, \theta^\star\rangle - y_i^q)^2.$$

Since $|y| \le Y_{\max}$ and $\|x\| \le R$ a.s., we have

$$Z_q \in [0,\, (Y_{\max} + R\|\theta^\star\|)^2] = [0, b_\star].$$

By Hoeffding's inequality, with probability at least $1 - \kappa/2$,

$$\left|\frac{1}{n}\sum_{q=1}^{n} Z_q - \mathbb{E}[Z]\right| \le b_\star\sqrt{\frac{2\log(2/\kappa)}{n}}.$$

$\square$

Lemma 17 (Uniform deviation over $\{\|\theta\| \le B_\theta\}$)**.** *Fix any radius* $B_\theta > 0$ *and let* $\kappa \in (0,1)$*. Define the squared-loss class*

$$\mathcal{F}_{B_\theta} \;:=\; \left\{ f_\theta(x,y) \;=\; \left(y - \langle x, \theta\rangle\right)^2 \;\middle|\; \|\theta\| \le B_\theta \right\}.$$

Then, with probability at least $1 - \kappa/2$ *(over the draw of the probing sample of size* n*),*

$$\sup_{\|\theta\| \le B_\theta} \left|\widehat{L}_{i,\mathrm{true}}(\theta) - \mathcal{R}(\theta)\right| \;\le\; \Lambda_{\mathrm{u}},$$

where

$$\Lambda_{\mathrm{u}} \;:=\; \frac{4\,(Y_{\max} + B_\theta R)\,B_\theta R}{\sqrt{n}} \;+\; b_{\mathrm{u}}\sqrt{\frac{2\log(2/\kappa)}{n}}, \qquad b_{\mathrm{u}} \;:=\; (Y_{\max} + B_\theta R)^2.$$

Proof. We apply Wainwright's Theorem 4.10.

Uniform bound b for $\mathcal{F}_{B_\theta}$. For any $\|\theta\| \le B_\theta$ and any (x, y) with $\|x\| \le R$, $|y| \le Y_{\max}$, we have

$$\big|y - \langle x, \theta\rangle\big| \;\le\; |y| + |\langle x, \theta\rangle| \;\le\; Y_{\max} + \|x\|\,\|\theta\| \;\le\; Y_{\max} + B_\theta R.$$

Hence

$$0 \;\le\; f_\theta(x, y) \;=\; \big(y - \langle x, \theta\rangle\big)^2 \;\le\; (Y_{\max} + B_\theta R)^2 \;=:\; b_{\mathrm{u}}.$$

Thus $\mathcal{F}_{B_\theta}$ is b_{u}–uniformly bounded.

Bound $R_n(\mathcal{F}_{B_\theta})$ via contraction to the linear class. Fix a sample $S = \{(x_i, y_i)\}_{i=1}^n$ and let $\sigma_1, \dots, \sigma_n$ be i.i.d. Rademacher signs. The empirical Rademacher average of $\mathcal{F}_{B_\theta}$ on S is

$$\mathrm{Rad}_S(\mathcal{F}_{B_\theta}) \;:=\; \mathop{\mathbb{E}}_{\sigma}\Big[\sup_{\|\theta\| \le B_\theta} \frac{1}{n}\sum_{i=1}^n \sigma_i\big(y_i - \langle x_i, \theta\rangle\big)^2\Big].$$

Define $u_{i,\theta} := \langle x_i, \theta\rangle$ and, for each i, the recentered function

$$\psi_i(u) \;:=\; \big(y_i - u\big)^2 - \big(y_i - 0\big)^2 \;=\; u^2 - 2\,y_i\,u, \qquad \psi_i(0) = 0.$$

Since $\mathbb{E}[\sigma_i] = 0$, the constant offsets $\{(y_i - 0)^2\}$ vanish in the Rademacher average. Thus

$$\mathrm{Rad}_S(\mathcal{F}_{B_\theta}) \;=\; \mathop{\mathbb{E}}_{\sigma}\Big[\sup_{\|\theta\| \le B_\theta} \frac{1}{n}\sum_{i=1}^n \sigma_i\,\psi_i\big(u_{i,\theta}\big)\Big].$$

We now bound the Lipschitz constants of ψ_i over the relevant range. For any $u, v \in [-B_\theta R, B_\theta R]$,

$$\big|\psi_i(u) - \psi_i(v)\big| = \big|\,(u - v)\,\big[(u - y_i) + (v - y_i)\big]\,\big| \;\le\; |u - v|\,\big(\,|u - y_i| + |v - y_i|\,\big) \;\le\; 2(Y_{\max} + B_\theta R)\,|u - v|.$$

Hence each ψ_i is L–Lipschitz with

$$L \;:=\; 2(Y_{\max} + B_\theta R), \qquad \psi_i(0) = 0.$$

By the Ledoux–Talagrand contraction inequality (applied elementwise to $\{\psi_i\}$ with common Lipschitz constant L and $\psi_i(0) = 0$), we have

$$\mathrm{Rad}_S(\mathcal{F}_{B_\theta}) \;\le\; L \cdot \mathop{\mathbb{E}}_{\sigma}\Big[\sup_{\|\theta\| \le B_\theta} \frac{1}{n}\sum_{i=1}^n \sigma_i\,u_{i,\theta}\Big] \;=\; L \cdot \mathrm{Rad}_S(\mathcal{G}_{B_\theta}),$$

where $\mathcal{G}_{B_\theta} := \{(x, y) \mapsto \langle x, \theta\rangle : \; \|\theta\| \le B_\theta\}$ is the linear class (note: dependence on y disappears).

The empirical Rademacher average of $\mathcal{G}_{B_\theta}$ on S is standard:

$$\mathrm{Rad}_S(\mathcal{G}_{B_\theta}) = \mathop{\mathbb{E}}_{\sigma}\Big[\sup_{\|\theta\| \le B_\theta} \frac{1}{n}\sum_{i=1}^n \sigma_i\,\langle x_i, \theta\rangle\Big] = \frac{B_\theta}{n}\,\mathop{\mathbb{E}}_{\sigma}\Big\|\sum_{i=1}^n \sigma_i x_i\Big\| \;\le\; \frac{B_\theta}{n}\,\sqrt{\mathop{\mathbb{E}}_{\sigma}\Big\|\sum_{i=1}^n \sigma_i x_i\Big\|^2} = \frac{B_\theta}{n}\,\sqrt{\sum_{i=1}^n \|x_i\|^2}.$$

Using $\|x_i\| \le R$, we conclude $\mathrm{Rad}_S(\mathcal{G}_{B_\theta}) \le B_\theta R/\sqrt{n}$. Combining,

$$\mathrm{Rad}_S(\mathcal{F}_{B_\theta}) \;\le\; L \cdot \mathrm{Rad}_S(\mathcal{G}_{B_\theta}) \;\le\; 2(Y_{\max} + B_\theta R)\,\frac{B_\theta R}{\sqrt{n}}.$$

Taking expectation over the sample (or simply noting that this bound holds for any S satisfying $\|x_i\| \le R$) yields

$$R_n(\mathcal{F}_{B_\theta}) \;\le\; 2(Y_{\max} + B_\theta R)\,\frac{B_\theta R}{\sqrt{n}}.$$

Apply Theorem 4.10 and choose δ. By Theorem 4.10, for any $\delta > 0$,

$$\sup_{\|\theta\| \le B_\theta} \big|\widehat{L}_{i,\mathrm{true}}(\theta) - \mathcal{R}(\theta)\big| \;\le\; 2\,R_n(\mathcal{F}_{B_\theta}) + \delta \;\le\; \frac{4(Y_{\max} + B_\theta R)\,B_\theta R}{\sqrt{n}} + \delta,$$

with probability at least $1 - \exp\big(-\frac{n\,\delta^2}{2b_{\mathrm{u}}^2}\big)$. Set $\delta = b_{\mathrm{u}}\sqrt{\frac{2\log(2/\kappa)}{n}}$ so that $\exp\big(-\frac{n\,\delta^2}{2b_{\mathrm{u}}^2}\big) = \exp\big(-\log(2/\kappa)\big) = \kappa/2$. The stated deviation bound Λ_{u} follows. $\square$

Lemma 18 (Concentration of Δ_n^2 under Accurate Probing)**.** *Assume that* $|\tilde{y}| \le Y_{\max,\text{probe}}$ *almost surely (e.g.,* $Y_{\max,\text{probe}} = R \max_j \|\theta_j^0\|$ *for linear peers at initialization). Then for any* $\kappa \in (0,1)$*, with probability at least* $1-\kappa/2$*,*

$$\Delta_n^2 \;\le\; B \;+\; C_0^2 \sqrt{\frac{2\log(2/\kappa)}{n}}, \qquad \textit{where} \quad C_0 \;:=\; Y_{\max} + Y_{\max,\text{probe}}.$$

Proof. Let $W = (\tilde{y} - y)^2$. By the boundedness assumptions, $0 \le W \le (Y_{\max} + Y_{\max,\text{probe}})^2 =: U$ almost surely, and $\mathbb{E}[W] \le B$ by Accurate Probing. By Hoeffding's inequality,

$$\Pr\left(\frac{1}{n}\sum_{q=1}^{n} W_q - \mathbb{E}[W] \ge t\right) \;\le\; \exp\left(-\frac{2nt^2}{U^2}\right).$$

Choosing $t = U\sqrt{\frac{2\log(2/\kappa)}{n}}$ yields $\Pr\big(\Delta_n^2 \ge \mathbb{E}[W] + U\sqrt{\frac{2\log(2/\kappa)}{n}}\big) \le \kappa/2$. Using $\mathbb{E}[W] \le B$ gives the stated bound with $C_0^2 = U$. □

Lemma 19 (A priori norm bound and choice of B_θ)**.** *For any probing learner* $i \in U$*,*

$$\|\tilde{\theta}_i\| \;\le\; B_\theta \quad \textit{with} \quad B_\theta \;:=\; \sqrt{\frac{2(Y_{\max}^2 + p\,Y_{\max,\text{probe}}^2)}{\lambda\,p}}.$$

Proof of Lemma 19. At any stationary point $\tilde{\Theta}$, $\tilde{\theta}_i$ minimizes the convex function

$$\hat{\Phi}_i^{\text{probe}}(\theta;\tilde{\Theta}) \;:=\; \tau\alpha_i\,\mathbb{E}_{z\sim\mathcal{P}_i}\big[\ell(z,\theta)\big] \;+\; (1-\tau)\,a_i(\tilde{\Theta})\,\mathbb{E}_{z\sim\mathcal{D}_i(\tilde{\Theta})}\big[\ell(z,\theta)\big] \;+\; p\,\widehat{L}_i(\theta) \;+\; \frac{\lambda\,p}{2}\,\|\theta\|^2.$$

By optimality of $\tilde{\theta}_i$ for $\hat{\Phi}_i^{\text{probe}}(\cdot;\tilde{\Theta})$,

$$\hat{\Phi}_i^{\text{probe}}(\tilde{\theta}_i;\tilde{\Theta}) \le \hat{\Phi}_i^{\text{probe}}(0;\tilde{\Theta}).$$

Using squared loss and $|y| \le Y_{\max}$, we have

$$\widehat{L}_i(0) = \frac{1}{n}\sum_{q=1}^{n}(\tilde{y}_i^q)^2 \le Y_{\max,\text{probe}}^2.$$

Hence,

$$\frac{\lambda\,p}{2}\,\|\tilde{\theta}_i\|^2 \;\le\; \tau\alpha_i\,Y_{\max}^2 \;+\; (1-\tau)\,a_i(\tilde{\Theta})\,Y_{\max}^2 \;+\; p\,\widehat{L}_i(0) \;\le\; Y_{\max}^2 + p\,Y_{\max,\text{probe}}^2.$$

Thus

$$\|\tilde{\theta}_i\| \;\le\; \sqrt{\frac{2(Y_{\max}^2 + p\,Y_{\max,\text{probe}}^2)}{\lambda\,p}} \;=\; B_\theta.$$

□

F CROSS-ENTROPY LOSS: PERFORMANCE GUARANTEE

Notation. In this section, we provide a high-probability upper bound on the full-population cross-entropy risk for a probing learner in MSGD-P at a stationary point. The proof mirrors the squared-loss analysis, with key differences arising from the geometry of the softmax and cross-entropy.

Throughout this section, let K denote the number of classes. Let $\theta \equiv W \in \mathbb{R}^{K\times d}$ be the matrix of class-wise linear weights and define logits $z(x) = Wx \in \mathbb{R}^K$ and predicted probabilities $q_W(x) = \text{softmax}(z(x))$. The multiclass cross-entropy loss is

$$\ell_{\text{CE}}((x,y),W) \;=\; -\sum_{c=1}^{K} y_c \,\log q_W(x)_c \;=\; -\log q_W(x)_Y,$$

where $y \in \{e_1,\ldots,e_K\}$ is one-hot and Y is the true class index. We assume $\|x\| \le R$ almost surely.

For a probing learner $i \in U$ (with probing weight $p > 0$), we define its augmented objective at Θ (as per MSGD-P) by

$$\begin{aligned}\Phi_i(W;\Theta) \;=\; &\tau\alpha_i \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{P}_i}\big[\ell_{\mathrm{CE}}((x,y),W)\big]\\ &+\;(1-\tau)\,a_i(\Theta)\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_i(\Theta)}\big[\ell_{\mathrm{CE}}((x,y),W)\big]\\ &+\;p\,\widehat{L}_{\mathrm{probe}}(W)\;+\;\frac{\lambda p}{2}\,\|W\|_F^2,\end{aligned}$$

where

$$\widehat{L}_{\mathrm{probe}}(W) \;=\; \frac{1}{n}\sum_{q=1}^{n}\ell_{\mathrm{CE}}\big((\tilde{x}_i^q,\tilde{y}_i^q),W\big) \;=\; -\frac{1}{n}\sum_{q=1}^{n}\sum_{c=1}^{K}\tilde{y}_{i,c}^q\log q_W(\tilde{x}_i^q)_c.$$

Here $\{(\tilde{x}_i^q)\}_{q=1}^n$ are probing covariates drawn i.i.d. from $\mathcal{P}_X$, and $\{\tilde{y}_i^q\}_{q=1}^n$ are soft pseudo-labels. For analysis, let y_i^q be the (hidden) true labels of $\tilde{x}_i^q$, and define the empirical "true" probing CE:

$$\widehat{L}_{\mathrm{true}}(W) \;=\; \frac{1}{n}\sum_{q=1}^{n}\ell_{\mathrm{CE}}\big((\tilde{x}_i^q,y_i^q),W\big) \;=\; -\frac{1}{n}\sum_{q=1}^{n}\log q_W(\tilde{x}_i^q)_{y_i^q}.$$

Let $\theta^\star \in \arg\min_W \mathcal{R}(W)$ be a (finite-norm) population minimizer of the unregularized CE risk $\mathcal{R}(W) = \mathbb{E}[\ell_{\mathrm{CE}}((x,y),W)]$, with $\epsilon := \mathcal{R}(\theta^\star)$ and $\|\theta^\star\|_F =: M_\star < \infty$.

Aggregation Rule. we define peer logits $z_j(x,\Theta) = \theta_j^0 x \in \mathbb{R}^K$ and the coordinatewise median of logits

$$\tilde{z}_c(x,\Theta) := \mathrm{median}(z_{j,c}(x,\Theta) : j \in T_i(x)), \quad c \in [K].$$

Then, we set the probing label as $y_{\mathrm{agg},i}(x,\Theta) = \mathrm{softmax}(\tilde{z}(x,\Theta))$.

F.1 ACCURATE PROBING ASSUMPTION AND PROOFS

Scenario	$T_i(x)$	B
Majority-good	$[m]$	$\epsilon + 2Rr$
Market-leader	$\{j^*\}$	ξ
Partial knowledge	G	$\epsilon + 2Rr$
Preference-aware	$\{\pi(x)\}$	ϵ

Table 2: Probing accuracy parameters for cross-entropy loss.

Assumption 6 (Accurate Probing for CE)**.** *There exists $B_{\mathrm{CE}} \geq 0$ such that*

$$\mathbb{E}\big[\mathrm{CE}(y,\tilde{y})\big] \;=\; \mathbb{E}\big[-\log\tilde{y}_Y\big] \;\leq\; B_{\mathrm{CE}},$$

where the expectation is over $(x,y)\sim\mathcal{P}$ and $\tilde{y} = \tilde{y}(x)$ produced by the robust aggregator.

We now present sufficient conditions ensuring Assumption 6.

Lemma 20 (Sufficient conditions for Accurate Probing (CE))**.** *Assumption 6 holds in each of the following scenarios:* ***(i) Majority-good.*** *Suppose strictly more than half of peers satisfy $\|\theta_j^0 - \theta^\star\|_F \leq r$. Then*

$$B_{\mathrm{CE}} \;=\; \mathbb{E}\big[-\log\tilde{y}_Y\big] \;\leq\; \epsilon \;+\; 2\,R\,r.$$

(ii) Market-leader. *Suppose learner i probes a single peer $j^\star$ with $\mathbb{E}[\ell_{\mathrm{CE}}((x,y),\theta_{j^\star}^0)] \leq \xi$. If $\tilde{y}(x) = q_{\theta_{j^\star}^0}(x)$, then $B_{\mathrm{CE}} \leq \xi$.*

(iii) Partial knowledge. *If a known subset G of peers satisfies $|G| > (m-1)/2$ and $\|\theta_j^0 - \theta^\star\|_F \leq r$ for all $j \in G$, and the aggregator uses the coordinatewise median over G, the same bound as in case (i) holds.*

(iv) Preference-aware. *Suppose each peer j solves the ERM on its preference partition $\mathcal{P}_j$:*

$$\bar{\theta}_j \in \arg\min_{W} \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{P}_j} \big[\ell_{\mathrm{CE}}((x,y),W)\big].$$

If learner i probes $\tilde{y}(x) = q_{\bar{\theta}_{\pi(x)}}(x)$, then

$$B_{\mathrm{CE}} \;\le\; \sum_{j=1}^{K} \alpha_j \mathop{\mathbb{E}}_{\mathcal{P}_j} \big[\ell_{\mathrm{CE}}((x,y),\bar{\theta}_j)\big] \;\le\; \epsilon.$$

Proof. We treat each scenario in turn.

(i) Majority-good. Fix any x with $\|x\| \le R$ and write $z^\star(x) = \theta^\star x$. For any good peer j (i.e., $\|\theta_j^0 - \theta^\star\|_F \le r$) and any class c,

$$|z_{j,c}(x) - z_c^\star(x)| \;=\; |(\theta_{j,c}^0 - \theta_c^\star)^\top x| \;\le\; \|\theta_{j,c}^0 - \theta_c^\star\|_2 \cdot \|x\| \;\le\; \|\theta_j^0 - \theta^\star\|_F \cdot \|x\| \;\le\; r\,R.$$

Since strictly more than half of the $\{z_{j,c}(x)\}_{j\neq i}$ lie in $[z_c^\star(x){-}rR,\; z_c^\star(x){+}rR]$, their coordinatewise median must also lie in this interval. Hence

$$\|\tilde{z}(x) - z^\star(x)\|_\infty \;\le\; r\,R.$$

Let $f(z) = -\log \mathrm{softmax}(z)_Y$. By Lemma 21,

$$-\log \tilde{y}_Y \;=\; f(\tilde{z}(x)) \;\le\; f\big(z^\star(x)\big) + 2\,\|\tilde{z}(x) - z^\star(x)\|_\infty \;\le\; -\log q_Y^\star(x) + 2\,r\,R,$$

where $q^\star = \mathrm{softmax}(z^\star)$. Taking expectations over $(x,y) \sim \mathcal{P}$, we get

$$B_{\mathrm{CE}} \;=\; \mathbb{E}\big[-\log \tilde{y}_Y\big] \;\le\; \mathbb{E}\big[-\log q_Y^\star(x)\big] + 2\,r\,R \;=\; \epsilon + 2\,r\,R.$$

(ii) Market-leader. Direct: $B_{\mathrm{CE}} = \mathbb{E}[-\log \tilde{y}_Y] = \mathbb{E}[\ell_{\mathrm{CE}}((x,y),\theta_{j^\star}^0)] \le \xi$.

(iii) Partial knowledge. The proof is identical to case (i), restricting the median to G. Since $|G| > (m-1)/2$ and all learners in G satisfy $\|\theta_j^0 - \theta^\star\|_F \le r$, the coordinatewise median over G satisfies the same bound.

(iv) Preference-aware. By optimality of $\bar{\theta}_j$ for the ERM objective,

$$\mathop{\mathbb{E}}_{\mathcal{P}_j} \big[\ell_{\mathrm{CE}}((x,y),\bar{\theta}_j)\big] \;\le\; \mathop{\mathbb{E}}_{\mathcal{P}_j} \big[\ell_{\mathrm{CE}}((x,y),\theta^\star)\big].$$

Summing over j with weights α_j gives

$$B_{\mathrm{CE}} \;\le\; \sum_j \alpha_j \mathop{\mathbb{E}}_{\mathcal{P}_j} \ell_{\mathrm{CE}}(\bar{\theta}_j) \;\le\; \sum_j \alpha_j \mathop{\mathbb{E}}_{\mathcal{P}_j} \ell_{\mathrm{CE}}(\theta^\star) \;=\; \epsilon.$$

□

Lemma 21 (CE is 2-Lipschitz in logits under $\|\cdot\|_\infty$). *Fix a class $Y \in [K]$ and define $f(z) := -\log \mathrm{softmax}(z)_Y = -z_Y + \mathrm{logsumexp}(z)$ for $z \in \mathbb{R}^K$. Then for any $z, z' \in \mathbb{R}^K$,*

$$\big|f(z') - f(z)\big| \;\le\; 2\,\|z' - z\|_\infty.$$

Proof. We have $\nabla f(z) = q(z) - e_Y$, where $q(z) = \mathrm{softmax}(z)$ and e_Y is the one-hot at Y. By the mean value inequality with dual norms $(\|\cdot\|_\infty, \|\cdot\|_1)$,

$$|f(z') - f(z)| \;\le\; \sup_{\xi \in [z,z']} \|\nabla f(\xi)\|_1 \cdot \|z' - z\|_\infty \;=\; \sup_{\xi} \sum_{k=1}^{K} |q_k(\xi) - (e_Y)_k| \cdot \|z' - z\|_\infty.$$

Since $y = e_Y$ is one-hot and $q \in \Delta^{K-1}$, $\sum_{k=1}^K |q_k - (e_Y)_k| = |q_Y - 1| + \sum_{k\neq Y} q_k = (1 - q_Y) + (1 - q_Y) \le 2$. Hence $|f(z') - f(z)| \le 2\,\|z' - z\|_\infty$. □

F.2 PERFORMANCE BOUND

We will use two probability floors:

$$\gamma_\star \;:=\; \frac{1}{K}\,\exp\bigl(-2R\,\|\theta^\star\|_F\bigr), \quad \Gamma_\star \;:=\; \log\frac{1}{\gamma_\star} \;=\; \log K + 2R\|\theta^\star\|_F,$$

and

$$B_\theta \;:=\; \sqrt{\frac{2(1+p)\log K}{\lambda\,p}}, \qquad \gamma_B \;:=\; \frac{1}{K}\,\exp\bigl(-2R\,B_\theta\bigr), \qquad \Gamma_B \;:=\; \log\frac{1}{\gamma_B} \;=\; \log K + 2RB_\theta.$$

Corollary 2 (Big-O summary)**.** *Under the assumptions of Theorem 5, with probability at least* $1-\kappa$*,*

$$\mathcal{R}(\tilde{\theta}_i) \le O\Bigl(\Bigl(\frac{p+1}{p}\Bigr)\epsilon \;+\; \lambda\,\|\theta^\star\|_F^2 \;+\; C_{bias}\sqrt{B_{\mathrm{CE}}} \;+\; \frac{(p+1)\,C_{gen}}{p}\sqrt{\tfrac{\log(1/\kappa)}{\lambda n}}\Bigr),$$

where the constant $C_{bias} = R\left(\|\theta^\star\|_F + \sqrt{2\log K/\lambda}\right)$ *and* $C_{gen} = C_{gen}(R, \|\theta^\star\|, \log(K))$ *is specified in the appendix.*

Theorem 5 (CE performance bound with probing)**.** *Let Assumptions 1, 2, 3, 4, and 6 hold. Let* $\tilde{\Theta}$ *be a stationary point of MSGD-P and let* $\tilde{\theta}_i$ *be the parameter for a probing learner* $i \in U$*. Then for any* $\kappa \in (0,1)$*, with probability at least* $1-\kappa$ *over the probing sample,*

$$\mathcal{R}(\tilde{\theta}_i) \le \Bigl(1+\frac{1}{p}\Bigr)\epsilon + \frac{\lambda}{2}\|\theta^\star\|_F^2 + R(\|\theta^\star\|_F + B_\theta)\sqrt{2B_{\mathrm{CE}}} + 4R\sqrt{\frac{(1+p)K\log K}{\lambda\,p\,n}}$$
$$+ \bigl(\Gamma_\star + \Gamma_B + R(\|\theta^\star\|_F + B_\theta)\bigr)\sqrt{\frac{\log(3/\kappa)}{2n}}.$$

Proof. We work on the event $\mathcal{E}_{\mathrm{pinsker}} \cap \mathcal{E}_\star \cap \mathcal{E}_{\mathrm{u}}$ where:

- $\mathcal{E}_{\mathrm{pinsker}}$ is the event of Lemma 24 (probability $\ge 1-\kappa/3$),
- $\mathcal{E}_\star$ is the event of Lemma 25 (probability $\ge 1-\kappa/3$),
- $\mathcal{E}_{\mathrm{u}}$ is the event of Lemma 26 (probability $\ge 1-\kappa/3$).

By a union bound,

$$\mathbb{P}(\mathcal{E}_{\mathrm{pinsker}} \cap \mathcal{E}_\star \cap \mathcal{E}_{\mathrm{u}}) \ge 1-\kappa.$$

Bound on $\widehat{L}_{\mathrm{probe}}(\tilde{\theta}_i)$. By Lemma 22, dividing by p,

$$\widehat{L}_{\mathrm{probe}}(\tilde{\theta}_i) \;\le\; \frac{\epsilon}{p} \;+\; \widehat{L}_{\mathrm{probe}}(\theta^\star) \;+\; \frac{\lambda}{2}\,\|\theta^\star\|_F^2.$$

By Corollary 28 applied at $W=\theta^\star$ and Lemma 25,

$$\widehat{L}_{\mathrm{probe}}(\theta^\star) \;\le\; \widehat{L}_{\mathrm{true}}(\theta^\star) \;+\; R\,\|\theta^\star\|_F\,\Delta_{1,n} \;\le\; \epsilon \;+\; \Gamma_\star\sqrt{\frac{\log(3/\kappa)}{2n}} \;+\; R\,\|\theta^\star\|_F\,\Delta_{1,n}.$$

Hence

$$\widehat{L}_{\mathrm{probe}}(\tilde{\theta}_i) \;\le\; \frac{\epsilon}{p} + \epsilon \;+\; \frac{\lambda}{2}\,\|\theta^\star\|_F^2 \;+\; \Gamma_\star\sqrt{\frac{\log(3/\kappa)}{2n}} \;+\; R\,\|\theta^\star\|_F\,\Delta_{1,n}.$$

Bound on $\widehat{L}_{\mathrm{true}}(\tilde{\theta}_i)$. By Corollary 28 at $W=\tilde{\theta}_i$,

$$\widehat{L}_{\mathrm{true}}(\tilde{\theta}_i) \;\le\; \widehat{L}_{\mathrm{probe}}(\tilde{\theta}_i) \;+\; R\,\|\tilde{\theta}_i\|_F\,\Delta_{1,n}.$$

Substituting the bound from the previous step,

$$\widehat{L}_{\mathrm{true}}(\tilde{\theta}_i) \;\le\; \frac{\epsilon}{p} + \epsilon \;+\; \frac{\lambda}{2}\,\|\theta^\star\|_F^2 \;+\; \Gamma_\star\sqrt{\frac{\log(3/\kappa)}{2n}} \;+\; R\,(\|\theta^\star\|_F + \|\tilde{\theta}_i\|_F)\,\Delta_{1,n}.$$

On $\mathcal{E}_{\text{pinsker}}$, by Lemma 24,

$$\Delta_{1,n} \;\le\; \sqrt{2B_{\text{CE}}} \;+\; \sqrt{\frac{\log(3/\kappa)}{2n}}.$$

Therefore,

$$\widehat{L}_{\text{true}}(\tilde{\theta}_i) \;\le\; \frac{\epsilon}{p} + \epsilon \;+\; \frac{\lambda}{2}\,\|\theta^\star\|_F^2 \;+\; R\,(\|\theta^\star\|_F + B_\theta)\,\sqrt{2B_{\text{CE}}} \;+\; \Big(\Gamma_\star + R\,(\|\theta^\star\|_F + B_\theta)\Big)\,\sqrt{\frac{\log(3/\kappa)}{2n}}.$$

Bound on $\mathcal{R}(\tilde{\theta}_i)$. By Lemma 26, on $\mathcal{E}_{\text{u}}$ and since $\|\tilde{\theta}_i\| \le B_\theta$ by Lemma 23,

$$\mathcal{R}(\tilde{\theta}_i) \;\le\; \widehat{L}_{\text{true}}(\tilde{\theta}_i) \;+\; 2\sqrt{2}\,RB_\theta\sqrt{\frac{K}{n}} \;+\; \Gamma_B\,\sqrt{\frac{\log(3/\kappa)}{2n}}.$$

Combining with the previous bound and substituting $B_\theta = \sqrt{2(1+p)\log K/(\lambda\,p)}$ (which yields $2\sqrt{2}RB_\theta\sqrt{K/n} = 4R\sqrt{(1+p)(K\log K)/(\lambda\,p\,n)}$) gives the stated bound. □

Lemma 22 (Stationarity bound)**.** *At a stationary point* $\tilde{\Theta}$*, for learner* $i \in U$*,*

$$p\,\widehat{L}_{\text{probe}}(\tilde{\theta}_i) \;\le\; \epsilon \;+\; p\,\widehat{L}_{\text{probe}}(\theta^\star) \;+\; \frac{\lambda\,p}{2}\,\|\theta^\star\|_F^2.$$

Proof of Lemma 22. By optimality of $\tilde{\theta}_i$ for $\Phi_i(\cdot;\tilde{\Theta})$,

$$\Phi_i(\tilde{\theta}_i;\tilde{\Theta}) \;\le\; \Phi_i(\theta^\star;\tilde{\Theta}).$$

Subtracting the two population CE terms at $W = \tilde{\theta}_i$ and $W = \theta^\star$, and using that

$$\tau\alpha_i \mathop{\mathbb{E}}_{\mathcal{P}_i} \ell_{\text{CE}}(\theta^\star) \;+\; (1-\tau)\,a_i(\tilde{\Theta}) \mathop{\mathbb{E}}_{\mathcal{D}_i(\tilde{\Theta})} \ell_{\text{CE}}(\theta^\star) \;\le\; \epsilon,$$

we obtain the stated bound. □

Lemma 23 (Norm bound and probability floor)**.** *We have* $\|\tilde{\theta}_i\|_F \le B_\theta = \sqrt{2(1+p)\log K/(\lambda\,p)}$*. Consequently, for any* x*,*

$$\min_{c\in[K]} q_{\tilde{\theta}_i}(x)_c \;\ge\; \gamma_B \;=\; \frac{1}{K}\exp\big(-2RB_\theta\big), \quad \Gamma_B \;=\; \log\frac{1}{\gamma_B} \;=\; \log K + 2RB_\theta.$$

Proof of Lemma 23. Compare $\Phi_i(\tilde{\theta}_i;\tilde{\Theta})$ to $\Phi_i(0;\tilde{\Theta})$. For $W = 0$, q_W is uniform and each CE term equals $\log K$. Thus

$$\frac{\lambda\,p}{2}\,\|\tilde{\theta}_i\|_F^2 \;\le\; (1+p)\,\log K \quad\Rightarrow\quad \|\tilde{\theta}_i\|_F \;\le\; \sqrt{\frac{2(1+p)\log K}{\lambda\,p}}.$$

For the floor, write for any c, c':

$$|z_c - z_{c'}| \;=\; |(W_c - W_{c'})^\top x| \;\le\; \|W_c - W_{c'}\|\,\|x\| \;\le\; 2\,\|W\|_F\,\|x\| \;\le\; 2R\|W\|_F.$$

Hence $q_W(x)_c \ge \frac{1}{K}\exp(-2R\|W\|_F)$. □

Lemma 24 (Pseudo-label discrepancy)**.** *Let* $\Delta_{1,n} = \frac{1}{n}\sum_{q=1}^n \|\tilde{y}_i^q - y_i^q\|_1$*. Under Assumption 6, for any* $\kappa \in (0,1)$*, with probability at least* $1-\kappa/3$*,*

$$\Delta_{1,n} \;\le\; \sqrt{2B_{\text{CE}}} \;+\; \sqrt{\frac{\log(3/\kappa)}{2n}}.$$

Proof of Lemma 24. For one-hot y, $\text{CE}(y,\tilde{y}) = \text{KL}(y\|\tilde{y})$ and Pinsker gives $\mathbb{E}\,\|y-\tilde{y}\|_1 \le \sqrt{2\,\mathbb{E}\,\text{KL}(y\|\tilde{y})} \le \sqrt{2B_{\text{CE}}}$. Since $\|y-\tilde{y}\|_1 \in [0,2]$, Hoeffding yields the stated bound. □

Lemma 25 (Concentration at $\theta^\star$)**.** *With probability at least* $1-\kappa/3$*,*

$$\left|\widehat{L}_{\rm true}(\theta^\star)-\epsilon\right| \;\le\; \Gamma_\star\sqrt{\frac{\log(3/\kappa)}{2n}},$$

where $\Gamma_\star=\log K+2R\|\theta^\star\|_F$.

Proof of Lemma 25. By Lemma 23 applied to $W=\theta^\star$, $\min_c q_{\theta^\star}(x)_c\ge\gamma_\star$, hence $\ell_{\rm CE}((x,y),\theta^\star)\in[0,\Gamma_\star]$. Hoeffding's inequality gives the result. ☐

Lemma 26 (Uniform convergence)**.** *With probability at least* $1-\kappa/3$*,*

$$\sup_{\|W\|_F\le B_\theta}\left|\widehat{L}_{\rm true}(W)-\mathcal{R}(W)\right| \;\le\; 2\sqrt{2}\,R\,B_\theta\sqrt{\frac{K}{n}} \;+\; \Gamma_B\sqrt{\frac{\log(3/\kappa)}{2n}}.$$

Proof of Lemma 26. For any sample $\{(x_i,y_i)\}_{i=1}^n$ with $\|x_i\|\le R$, we bound the empirical Rademacher complexity of

$$\mathcal{F}_{B_\theta} \;:=\; \left\{(x,y)\mapsto\ell_{\rm CE}((x,y),W):\|W\|_F\le B_\theta\right\}.$$

The function $f(z,y)=-\log\operatorname{softmax}(z)_Y$ has gradient $\nabla_z f=q-y$ with $\|\nabla_z f\|_2=\|q-y\|_2\le\sqrt{2}$. Thus f is $\sqrt{2}$-Lipschitz in z. By the vector contraction lemma,

$$\mathfrak{R}_n(\mathcal{F}_{B_\theta}) \;\le\; \sqrt{2}\cdot\mathbb{E}\left[\sup_{\|W\|_F\le B_\theta}\frac{1}{n}\sum_{i=1}^n\sum_{k=1}^K\sigma_{i,k}\,(Wx_i)_k\right],$$

with $\sigma_{i,k}$ i.i.d. Rademacher. The inner supremum equals

$$\frac{B_\theta}{n}\,\mathbb{E}\left\|\sum_{i=1}^n\sum_{k=1}^K\sigma_{i,k}\,e_kx_i^\top\right\|_F \;\le\; \frac{B_\theta}{n}\sqrt{\mathbb{E}\left\|\sum_{i,k}\sigma_{i,k}\,e_kx_i^\top\right\|_F^2} \;=\; \frac{B_\theta}{n}\sqrt{\sum_{k=1}^K\sum_{i=1}^n\|x_i\|^2} \;\le\; B_\theta R\sqrt{\frac{K}{n}}.$$

Hence $\mathfrak{R}_n(\mathcal{F}_{B_\theta})\le\sqrt{2}\,B_\theta R\sqrt{K/n}$. A standard symmetrization and bounded-difference argument (the per-sample loss range over $\|W\|\le B_\theta$ is at most Γ_B by Lemma 23) yields, with prob. $\ge 1-\kappa/3$,

$$\sup_{\|W\|_F\le B_\theta}\left|\widehat{L}_{\rm true}(W)-\mathcal{R}(W)\right| \;\le\; 2\,\mathfrak{R}_n(\mathcal{F}_{B_\theta}) \;+\; \Gamma_B\sqrt{\frac{\log(3/\kappa)}{2n}},$$

which gives the stated bound. ☐

Lemma 27 (Logits bridge for cross-entropy; no probability floors)**.** *Let* $W\in\mathbb{R}^{K\times d}$*,* $z(x)=Wx\in\mathbb{R}^K$*, and* $q_W(x)=\operatorname{softmax}(z(x))$*. For any* $x\in\mathbb{R}^d$*, any one-hot label* $y\in\{e_1,\dots,e_K\}$*, and any soft pseudo-label* $\tilde{y}\in\Delta^{K-1}$ *(i.e.,* $\tilde{y}_c\ge0$*,* $\sum_c\tilde{y}_c=1$*), the cross-entropy difference satisfies the exact identity*

$$\mathrm{CE}(\tilde{y},q_W)-\mathrm{CE}(y,q_W) \;=\; -\langle\tilde{y}-y,\;z(x)\rangle.$$

Consequently,

$$\left|\mathrm{CE}(\tilde{y},q_W)-\mathrm{CE}(y,q_W)\right| \;\le\; \|\tilde{y}-y\|_1\cdot\|z(x)\|_\infty \;\le\; \|\tilde{y}-y\|_1\cdot R\,\|W\|_F.$$

Proof. Recall that $\mathrm{CE}(r,q)=-\sum_{c=1}^K r_c\log q_c$ for any probability vector r, and $q_c=\frac{e^{z_c}}{\sum_j e^{z_j}}$ with $z=Wx$. Hence

$$-\log q_c \;=\; \log\left(\sum_{j=1}^K e^{z_j}\right)-z_c \;=\; \operatorname{logsumexp}(z)-z_c.$$

Therefore, for any $r\in\Delta^{K-1}$,

$$\mathrm{CE}(r,q_W) \;=\; \sum_{c=1}^K r_c\left(\operatorname{logsumexp}(z)-z_c\right) \;=\; \operatorname{logsumexp}(z)\cdot\underbrace{\sum_{c=1}^K r_c}_{=1} \;-\; \sum_{c=1}^K r_c\,z_c \;=\; \operatorname{logsumexp}(z) \;-\; \langle r,z\rangle.$$

Applying this twice with $r = \tilde{y}$ and $r = y$ gives

$$\mathrm{CE}(\tilde{y}, q_W) - \mathrm{CE}(y, q_W) \;=\; \big(\mathrm{logsumexp}(z) - \langle \tilde{y}, z\rangle\big) \;-\; \big(\mathrm{logsumexp}(z) - \langle y, z\rangle\big) \;=\; -\langle \tilde{y} - y, z\rangle,$$

which is the claimed identity.

For the inequalities, use Hölder and the fact that $\|z\|_\infty = \max_c |z_c|$:

$$\big|\,\mathrm{CE}(\tilde{y}, q_W) - \mathrm{CE}(y, q_W)\big| \;=\; |\langle \tilde{y} - y, z\rangle| \;\le\; \|\tilde{y} - y\|_1 \, \|z\|_\infty.$$

Finally, since $z_c = W_c^\top x$ and $\|x\| \le R$,

$$\|z\|_\infty \;=\; \max_c |W_c^\top x| \;\le\; \max_c \|W_c\|_2 \cdot \|x\| \;\le\; \Big(\max_c \|W_c\|_2\Big) R \;\le\; R\,\|W\|_F,$$

because $\|W\|_F^2 = \sum_c \|W_c\|_2^2 \ge \max_c \|W_c\|_2^2$. Combining the bounds yields the result. □

Lemma 28 (Empirical bridge on the probing batch)**.** *On the probing dataset* $\{(\tilde{x}_i^q, \tilde{y}_i^q, y_i^q)\}_{q=1}^n$, *define*

$$\widehat{L}_{\mathrm{probe}}(W) \;=\; \frac{1}{n}\sum_{q=1}^n \mathrm{CE}(\tilde{y}_i^q, q_W(\tilde{x}_i^q)), \quad \widehat{L}_{\mathrm{true}}(W) \;=\; \frac{1}{n}\sum_{q=1}^n \mathrm{CE}(y_i^q, q_W(\tilde{x}_i^q)), \quad \Delta_{1,n} \;=\; \frac{1}{n}\sum_{q=1}^n \|\tilde{y}_i^q - y_i^q\|_1.$$

Then for any W,

$$\big|\widehat{L}_{\mathrm{probe}}(W) - \widehat{L}_{\mathrm{true}}(W)\big| \;\le\; R\,\|W\|_F \cdot \Delta_{1,n}.$$

Proof. Apply Lemma 27 termwise and average. □

G EXPERIMENTAL DETAILS AND ADDITIONAL RESULTS

G.1 DATASET AND PREPROCESSING DETAILS

MovieLens-10M [Harper and Konstan, 2015]. This dataset contains 10 million movie ratings from 70k users across 10k movies, providing a natural testbed for multi-learner competition in recommendation. Following Bose et al. [2024] and Su and Dean [2024], we extract $d = 16$ dimensional user embeddings via matrix factorization and retain ratings for the top 200 most-rated movies, yielding a population of 69,474 users. Each user's data consists of $z = (x, r)$ where $x \in \mathbb{R}^d$ is the embedding and r contains their ratings. Let Ω_x denote the set of movies rated by user x with $|\Omega_x|$ movies. Each learner fits a linear model $\theta \in \mathbb{R}^{d\times 200}$ using squared loss:

$$\ell(z;\theta) = \frac{1}{|\Omega_x|}\sum_{i\in\Omega_x} (\theta_i^\top x - r_i)^2.$$

ACS Employment [Ding et al., 2021]. We use the ACSEmployment task from `folktables`, where the goal is to predict employment status from demographic features. The population consists of 38,221 individuals from the 2018 Alabama census (ages 16–90), with $d = 16$ features describing age, education, marital status, etc. Each user's data is $z = (x, y)$ where $x \in \mathbb{R}^d$ (standardized to zero mean, unit variance) and $y \in \{0, 1\}$. Each learner uses logistic regression:

$$\ell(z;\theta) = -y\log(\sigma(\theta^\top x)) - (1-y)\log(1-\sigma(\theta^\top x)),$$

where σ is the sigmoid function. The model predicts $\hat{y} = \mathbf{1}[\theta^\top x > 0]$.

Amazon Reviews 2023. We use the `McAuley-Lab/Amazon-Reviews-2023` corpus (via HuggingFace), constructing a binary sentiment task from review text and star ratings. We sample up to 30,000 reviews from nine product categories and define labels by $y = \mathbf{1}[\text{rating} \ge 4]$ (so 1–3 stars are negative, 4–5 stars are positive). Features are $d = 384$ dimensional sentence embeddings of the review text produced by `all-MiniLM-L6-v2`, with a stratified 95/5 train-test split. Each learner uses the same logistic regression setup as Census.

User Preferences. For Census and MovieLens, we simulate a market with $m = 5$ learners and model inherent user preferences $\pi(z)$ via K-means clustering ($K = 5$) on user features, assigning each user's preferred platform based on their cluster membership. For Amazon, we use category-based partitions ($m = 9$ learners, one per product category): each review is assigned to its category group, and this group index induces the preference partition $\{S_i\}_{i=1}^m$. In all cases, the partition captures the intuition that users from different demographic or behavioral segments may have systematic affinities for different platforms.

Train/test splits. Each experiment constructs one held-out test set at the start and uses the remaining users as the shared training pool for all learners. Census uses a fixed 99/1 split with random seed 0; Amazon uses a stratified 95/5 split over sentiment labels; MovieLens uses a 90/10 user split. We do not use a validation split because the experimental hyperparameters are fixed in advance rather than tuned per run. Dataset-specific hyperparameters are summarized in Table 3.

Census neural experiment. The neural Census experiment in Figure 5 changes only the learner family relative to the linear Census experiments. Each learner is a two-layer ReLU MLP with architecture $d \to 64 \to 1$, trained with SGD on binary cross-entropy. The held-out test split, K-means preference clusters, and cluster-induced learner rankings are reused unchanged. In the bad-outcome run, learners are randomly initialized, $\tau = 0.3$, and no learner probes. In the good-outcome run, learners are initialized by partition pretraining on the users who rank them first, $\tau = 0.7$, and Learner 2 probes. Probing is offline: pseudo-labels are collected once from the initialized peer selected by the preference-aware rule $T_i(x) = \{\pi(x)\}$, with $\kappa = 0$, and the probing learner then trains on the weighted combination of organic and pseudo-labeled losses. We use $T = 1000$ MSGD iterations, three random seeds, and the probing weights shown in Figure 5.

G.2 ADDITIONAL EXPERIMENTS

In this section, we provide additional experiments, which elaborate on Section 5.

Expt 1 for other scenarios We replicate Expt 1 from Section 5 in the two globally-good settings from Definition 2: market-leader and majority-good. As in the preference-aware case, standard MSGD without probing ($p = 0$, $\tau = 0.3$) converges to equilibria with large full-population gaps to the dashed black baseline (Figures 7 and 9). In the market-leader setting (Figure 7), the most overspecialized learner is about 0.32 below baseline on Census (≈ 0.47 vs ≈ 0.79), and more than 2.6 MSE above baseline on MovieLens (> 5.6 vs ≈ 2.95). In the majority-good setting (Figure 9), two learners remain overspecialized with sizable baseline gaps (Census around 0.22–0.25 below baseline; MovieLens around 2.0–2.9 above baseline). These results show that poor global equilibria are not unique to the preference-aware scenario.

Expt 2 for other scenarios We next replicate Expt 2 for the other scenarios. In the market-leader scenario, Learner 4 probes the known leader (Learner 0), and its final Census accuracy improves from about 0.55 to about 0.75, while its MovieLens loss drops from about 5.1 to about 3.1 as p increases (Figure 8). Relative to the dashed baseline, this closes most of the initial gap (roughly from 0.24 to 0.04 on Census, and from about 2.2 to about 0.1 on MovieLens). In the majority-good scenario, where Learner 4 probes via median aggregation over all peers, we observe a similar pattern: Census accuracy rises from about 0.52 to about 0.77, and MovieLens loss decreases from about 4.7 to about 3.1 (Figure 10). This again closes a large fraction of the baseline gap (roughly from 0.27 to 0.02 on Census, and from about 1.7 to about 0.1 on MovieLens). Well-performing learners change only modestly, indicating that probing primarily benefits the underperforming learner and mitigates overspecialization.

Expt 4: Impact of noise in selection of probed labels Finally, we test robustness to noisy probing-source selection. For each probe query x, the probing learner queries $\pi(x)$ with probability $1 - \kappa$, and with probability κ it queries a random other learner. Across Census, Amazon, and MovieLens (Figure 4), increasing κ causes only mild changes in the probing learner's final performance relative to the low-noise case, while preserving strong gains from probing. Thus, our method is robust to imperfect estimates of the ranking function.

Expt 5: What happens when multiple learners probe? We also evaluate a preference-aware setting where multiple learners probe simultaneously. In Figure 11, the triangle-marked learners (Learners 2 and 3) both probe while the dashed black line indicates the full-data baseline. As p increases, Learner 2 improves from about 0.60 to about 0.78, and Learner 3 improves from about 0.66 to about 0.79. Equivalently, their baseline gaps shrink from roughly 0.19 and 0.13 at $p = 0$ to about 0.01 and near zero at $p = 0.8$. The non-probing learners move only slightly, indicating that simultaneous probing remains stable and still helps underperforming learners recover most of the overspecialization gap.

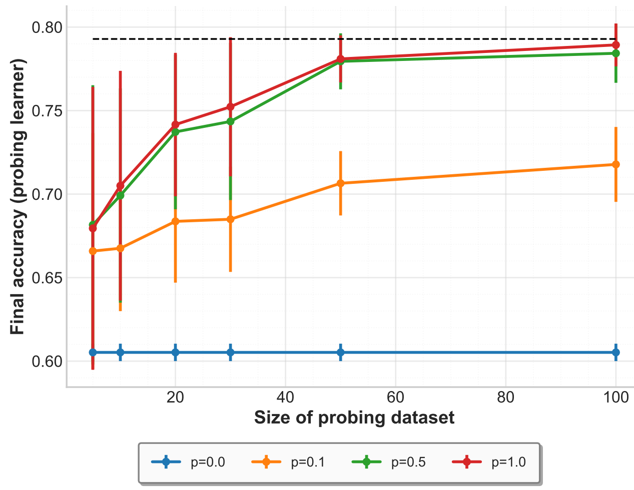

Figure 6: **Performance of probing learner on Census as a function of n.** Error bars show one standard deviation over 10 random seeds.

Parameter	Description	Census	MovieLens	Amazon
m	Number of learners	5	5	9
T	Total rounds	4000	4000	20000
λ	L2 regularization	10^{-9}	10^{-3}	10^{-9}
n	Offline probe dataset size	100	1000	500

Table 3: Hyperparameters by dataset. Shared values are merged across columns.

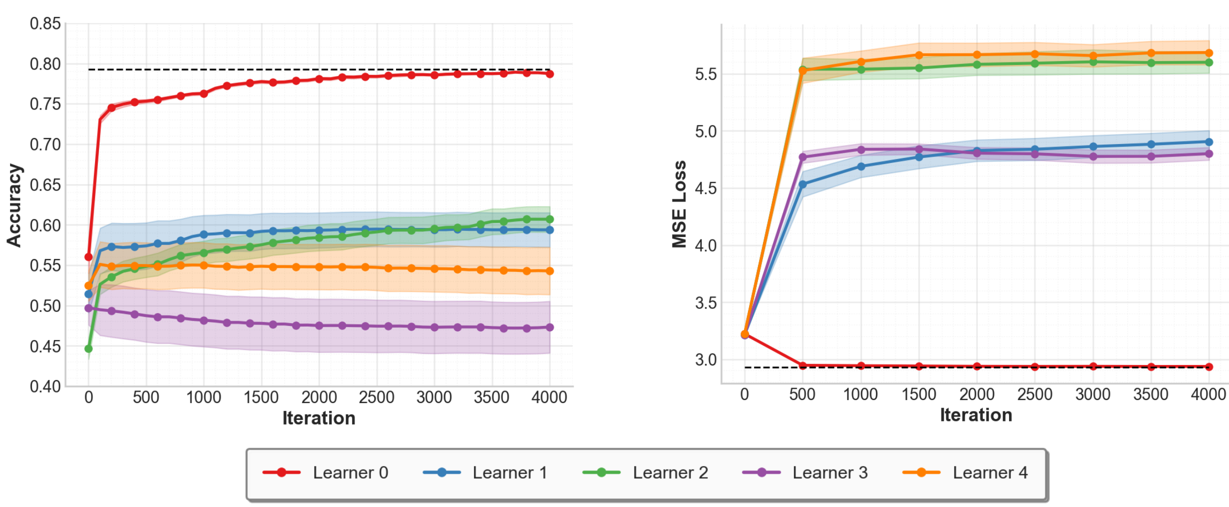

Figure 7: **MSGD full-population performance with random initialization (Market-leader scenario).** Left: Census test accuracy. Right: MovieLens test loss. Here $\tau = 0.3$.

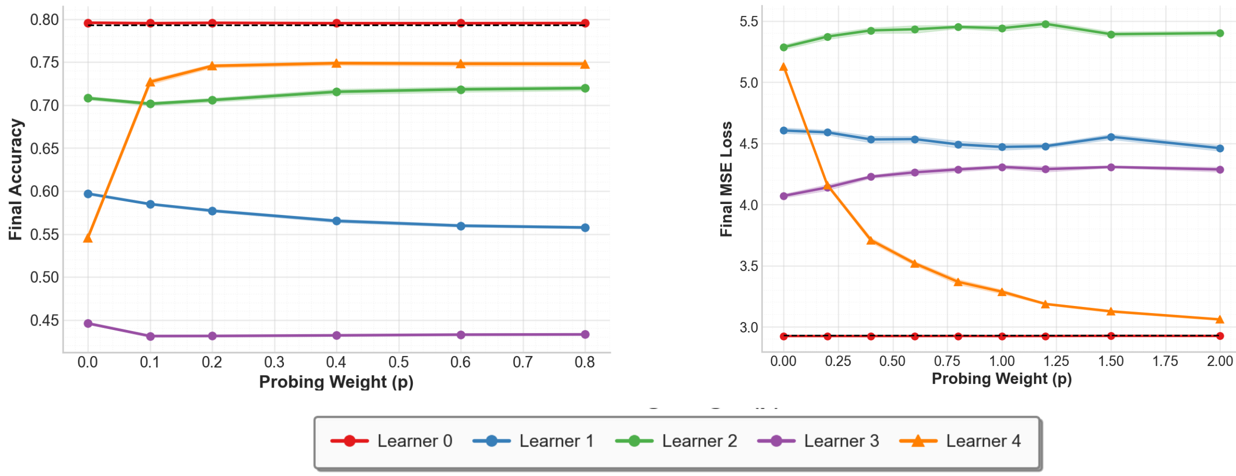

Figure 8: **Effect of probing on full-population performance (Market-leader scenario)** Left: Final accuracy vs probing weight p on Census. Right: MovieLens final loss vs p. The triangle markers indicate Learner 4 probes the market leader, learner 0. Here $\tau = 0.7$.

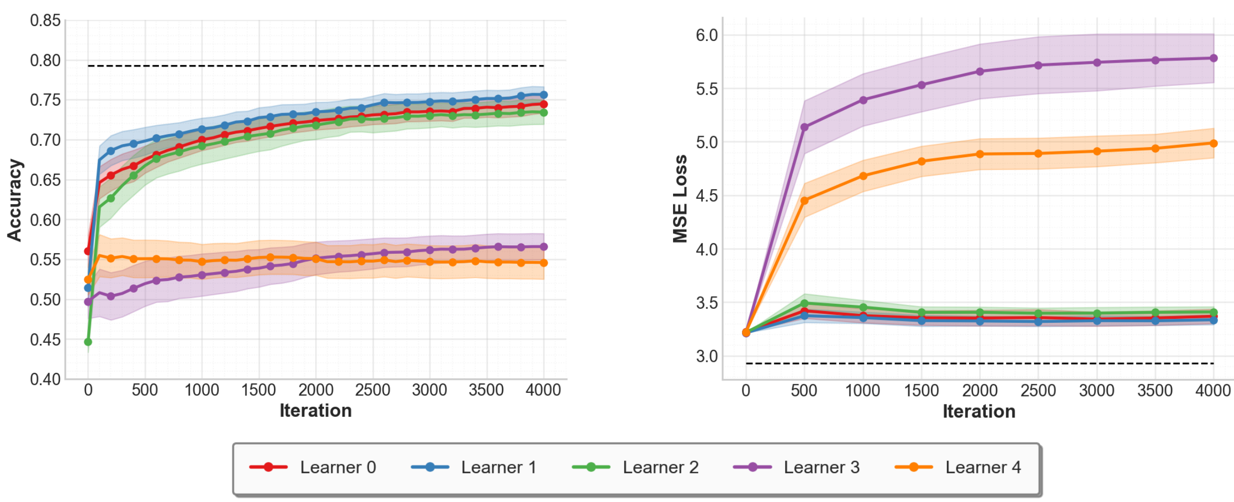

Figure 9: **MSGD full-population performance with random initialization (Majority good scenario).** Left: Census test accuracy. Right: MovieLens test loss. Here $\tau = 0.3$.

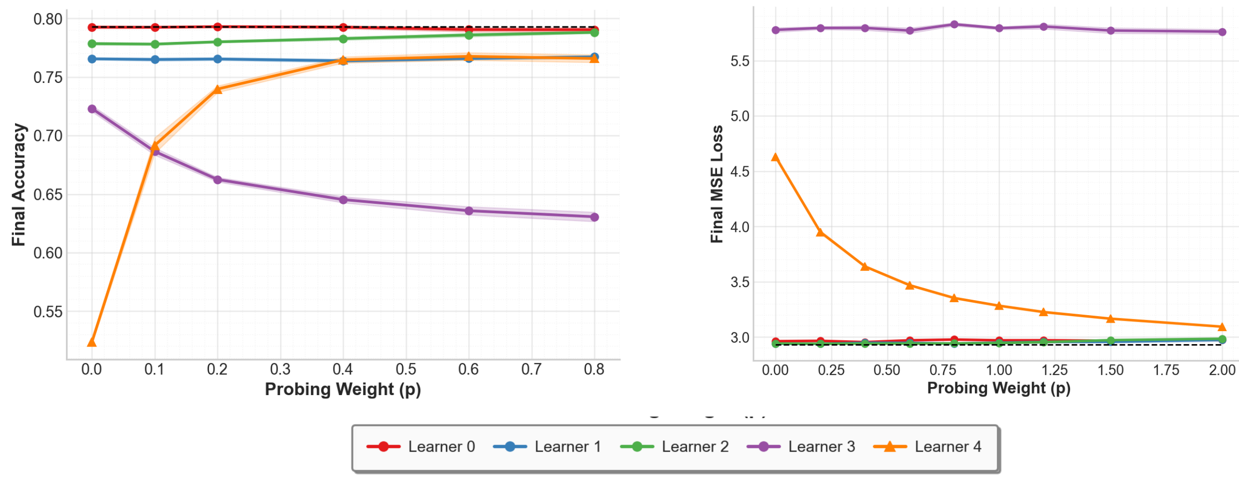

Figure 10: **Effect of probing on full-population performance (Majority Good Scenario)** Left: Final accuracy vs probing weight p on Census. Right: MovieLens final loss vs p. Triangle markers indicate Learner 4 is probing via median aggregation over all peers. Here $\tau = 0.7$.

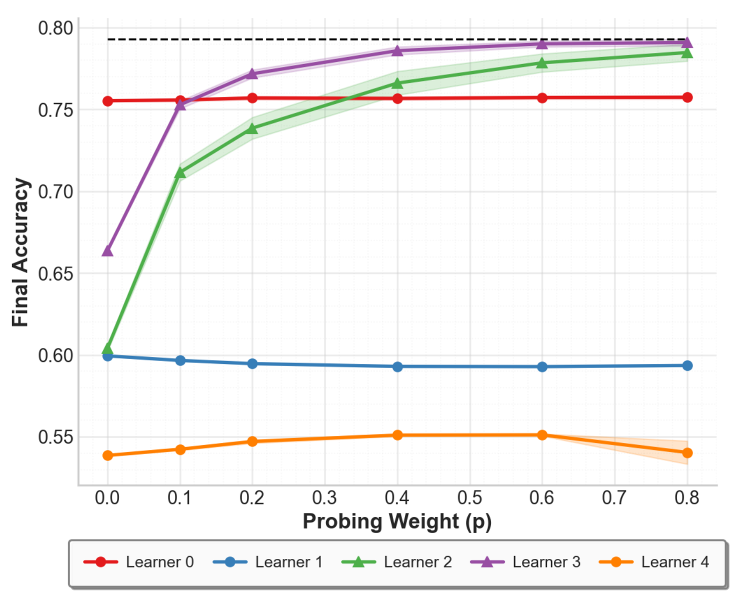

Figure 11: **Effect of probing on full-population performance when multiple learners probe (Preference-aware scenario).** Census final accuracy vs probing weight p. Triangle markers indicate the probing learners (Learners 2 and 3). The dashed black line denotes the full-data baseline.