MATHEMATICS OF OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0364-765X | EISSN 1526-5471 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Learning to Rank under Strategic Corruption: A Nonatomic Approach

Qinzhen Li[1] Yifan Feng[2] Hongfan (Kevin) Chen[3]

[1]Institute of Operations Research and Analytics, National University of Singapore, email: qinzhen@u.nus.edu
[2]NUS Business School and Institute of Operations Research and Analytics, National University of Singapore, email: yifan.feng@nus.edu.sg
[3]CUHK Business School, The Chinese University of Hong Kong, email: kevinchen@cuhk.edu.hk

We study a dynamic learning-to-rank problem in which N agents strategically inflate the reward signals that are used to rank them, with corruption budgets chosen endogenously. The learner observes noisy manipulated outcomes and seeks to maximize total true reward over T periods. Under a general supermodular ranking-reward model, we show that efficient learning remains achievable despite strategic manipulation. Specifically, a simple Experiment-Then-Commit policy attains regret bounded in T, namely $\mathcal{O}(\min\{N \log T, N^3\})$, under a strategy profile that forms an $o(1)$-equilibrium among agents as N and T grow.

Our analysis proceeds in three parts. First, we show that agents' strategic interactions converge to a static nonatomic game. Second, we find a nonatomic equilibrium in this limit, where agents' manipulation intensity increases with their types. Third, we transport the limit equilibrium to the finite-agent, finite-horizon setting, which yields non-asymptotic approximate-equilibrium and regret guarantees.

Key words: sequential learning, ranking, strategic data manipulation, nonatomic games, approximate equilibrium, regret analysis

1. Introduction We study the problem of sequential learning to rank a population of N strategic agents over T periods. In each period, the learner assigns the agents to ranked positions, observes a noisy signal of each agent's performance, and uses these signals to refine her future rankings so as to maximize cumulative reward. Each agent, in turn, can inflate his observed signal at a cost, anticipating that it will earn him a more valuable position later. As a result, the learner faces data corruption as an outcome of the collective strategic responses of the agents. Her task is to design a ranking policy that achieves low regret under such corruption.

Such settings arise naturally in many data-driven ranking environments. One main motivating example is product ranking on e-commerce platforms, where sellers are ranked based on consumer engagement signals (e.g., clicks and purchases) and can inflate these signals through "brushing", i.e., the systematic generation of fake orders to boost rankings, which is a prevalent practice documented

as an industry-wide phenomenon (see, e.g., Golrezaei et al. 2023, Jin et al. 2023). The same logic arises in search engine rankings, where websites compete for visibility in organic search results and can inflate the signals on which the search engine relies, using click farms, manufactured engagement, and other "black-hat" search engine optimization (SEO) tactics, to climb the rankings. In all such settings, the algorithm's policy shapes agents' inflation incentives, and the resulting inflation distorts the very data the algorithm relies on. This coupling poses challenges on two interlocking fronts: the algorithm's need to *learn* from data, and the *endogenous corruption* of that data by strategic agents.

From the learning side, the learner must contend not only with the combinatorial difficulty already inherent to ranking-bandit problems (e.g., Lagrée et al. 2016, Lattimore et al. 2018, Ermis et al. 2020, Wang et al. 2024), but with the fact that the signals she relies on to distinguish high-type from low-type agents are systematically distorted by the agents themselves. The existing literature on learning under corrupted data predominantly models corruptions as *adversarial*, i.e., those chosen by a worst-case adversary subject to a budget constraint (e.g., Lykouris et al. 2018, Golrezaei et al. 2023). While powerful, the adversarial framework can be too conservative: by treating corruption as unstructured worst-case disruptions, it forgoes the gains available when corruption carries concrete structure. For example, in our setting, the structure we wish to capture comes from the equilibrium behavior of self-interested agents. As we will show, recognizing this structure leads to performance guarantees substantially sharper than the adversarial framework would suggest.

On the game-theoretic side, each agent's inflation action depends on its effect on this agent's own future ranking and on the actions of all other agents. The resulting intertemporal strategic interaction can thus be highly intricate. Characterizing this equilibrium, as well as the learner's performance against it, becomes a nontrivial problem in its own right. Together, these two challenges are the central object of this paper. We thus ask: *does there exist a ranking algorithm that can achieve near-optimal performance under potentially pervasive strategic corruption?*

1.1. Overview of Main Results. In this paper, we show that a simple, model-free policy can make efficient learning remain achievable under strategic corruption. We study Experiment-Then-Commit (ETC), a policy that simply cycles uniformly through rankings during an experimentation phase before committing to the empirically best one. Under an appropriate equilibrium notion of agents' strategic behavior, the regret of ETC is of order $\mathcal{O}_{N,T}(\min\{N \log T, N^3\})$. This matches a universal lower bound up to logarithmic factors, and stands in sharp contrast to adversarial corruption models, where an $\Omega(T)$ corruption budget yields a vacuous $\Omega(T)$ regret.

We establish the results for a general class of supermodular ranking models that generalizes the classic position-based model (PBM). In this class, the organic reward is a supermodular function of the agent's type and his assigned position, so that the clairvoyant's optimal ranking is to simply rank

the agents by their types. We then consider a strategic environment, where agents can manipulate anonymously, at low cost, and without exogenous restrictions. The technical heavy lifting lies in the game-theoretic and performance analysis of ETC under this dynamic and strategic environment. To that end, we first show that under ETC, it is a weakly dominant strategy for every agent to exhaust his corruption budget during the experimentation phase (Proposition 4). This reduction collapses the dynamic game into a static one in which each agent's only strategic decision is the size of his corruption budget.

Despite the static reduction, substantial challenges remain: agents' utilities depend on high-dimensional stochastic processes, making them often intractable to evaluate, and the induced payoff structures are often ill-behaved (e.g., non-concave, multi-modal). To overcome this, we adopt a three-step strategy, as explained below.

1. *Step 1: Passing the game to the limit.* We pass the number of agents N and the time horizon T to infinity. In the limit, we derive a simple nonatomic game in the spirit of Schmeidler (1973), Mas-Colell (1984).

   The validity of the nonatomic limit rests on a general convergence theory we develop (Theorems 1 and 2), which identifies conditions under which the pre-limit ranking game converges to its nonatomic limit. This theory certifies that the game in the limit is the right object to study, and it serves as the engine for the finite-system guarantees in Step 3.
2. *Step 2: Equilibrium characterization in the limit.* In the nonatomic game, we find a nonatomic equilibrium (NAE) in closed form, where every agent's inflation intensity is given by an integral involving his marginal reward with respect to the ranking position; see Theorem 3.

   A key structural insight that arises from this characterization is that the inflation intensity is strictly increasing in type in this equilibrium. That represents a "*self-reinforcing*" property of the agents' strategic manipulation, which ultimately facilitates, rather than erodes, learning of the agents' ordinal information. A deeper intuition for this self-reinforcing property comes from the supermodularity of the reward function, which implies that higher-type agents are more sensitive to ranking positions and hence have stronger incentives to boost their positions.
3. *Step 3: Performance guarantees in the pre-limit.* Combining the convergence theory of Step 1 with the equilibrium characterization of Step 2, we obtain two non-asymptotic guarantees for the original N-agent, T-horizon game.
   - *Approximate equilibrium (Theorem 4).* The nonatomic equilibrium profile constitutes an $\mathcal{O}_{N,T}\Big(\frac{\log T}{T}+\frac{1}{N}\Big)$-equilibrium. As a result, every agent's incentive to deviate is on the order of $o(1)$ as N and T grow, regardless of the relative magnitudes of N and T.
   - *Regret bound (Theorem 5).* Under the nonatomic-equilibrium strategy profile, the regret of ETC grows on the order of $\mathcal{O}_{N,T}(\min\{N\log T,\, N^3\})$. At the same time, we derive a universal

regret lower bound that grows on the order of $\Omega_{N,T}(N)$; see Proposition 1. Together these imply that ETC is order-optimal in N and T separately and jointly near-order-optimal up to logarithmic factors. This regret order is decisively better than results under adversarial corruption, where an $\Omega(T)$ corruption budget generically leads to $\Omega(T)$ regret.

The main technical contribution of this paper is a nonatomic-convergence framework that turns an otherwise intractable dynamic strategic-learning problem into a tractable continuum game while providing non-asymptotic guarantees in the pre-limit market; see Remark 2 for further explanation and Figure 1 for an illustration. The framework is distinctively different from the usual mean-field game approaches, and it does not rely on a specific scaling between N and T. We believe it is also of independent interest and extends naturally to richer model variants, as we demonstrate in the extension where we study the cascade ranking model.

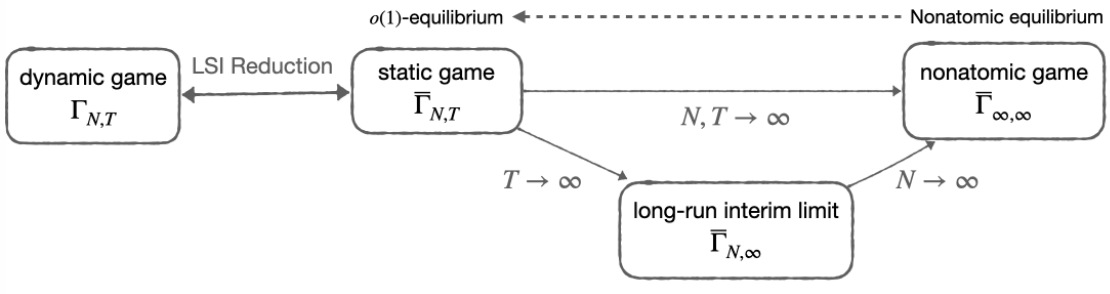


FIGURE 1. **Roadmap of the game formulations.**

1.2. Further Analysis and Extensions We conduct several additional analyses to deepen the understanding of our findings and evaluate their robustness.

Finite-agent game analysis. We analyze the existence of exact Nash equilibria in the finite-N game. We find that while desirable equilibria (e.g., the manipulation-free equilibrium) can emerge under short horizons, high noise, or high manipulation cost, Nash equilibria may fail to exist altogether otherwise (Propositions 5 and 6). Since we aim to cover settings in which T can be arbitrarily larger than N, our findings suggest that adopting relaxed equilibrium notions – such as the *approximate* equilibrium used in this paper – may be inevitable.

Cascade reward model. We extend the analysis to a cascade model, where each agent's reward depends on the agents ranked above him. We obtain analogues of all main results in the base model: LSI remains a dominant strategy (Proposition 7), the nonatomic equilibrium admits a closed-form solution (Theorem 6), the NAE profile serves as an approximate equilibrium with explicit convergence rates (Theorem 7), and the regret of ETC remains near order-optimal (Theorem 8).

2. Literature Review

Dynamic learning with corrupted/manipulated data. From a methodological perspective, our paper contributes to the growing literature on dynamic learning under corrupted or manipulated data. Most of this literature models corruption as adversarial; a representative example is Lykouris et al. (2018), with further developments in a variety of settings; see, for example, Gupta et al. (2019), Bogunovic et al. (2020), Zhong et al. (2021), Golrezaei et al. (2023), Chen and Wang (2023), Chen et al. (2023). In particular, Golrezaei et al. (2023) studies a learning and ranking problem under a cascade demand model with adversarial fake clicks. A limitation of the adversarial formulation, however, is that it can be overly pessimistic for settings in which data corruption is incentive-driven. The resulting regret guarantees typically scale *linearly* in the corruption budget.

Our work is motivated by the idea that stronger guarantees may be possible when corruption is structured rather than fully adversarial. To this end, we study strategic, *self-interested* manipulation, which remains relatively less explored; see Braverman et al. (2019), Feng et al. (2020), Birge et al. (2021), Dong et al. (2022), Chen et al. (2022), Harris et al. (2023), Kleine Buening et al. (2024b,a) for related examples. Our work is particularly inspired by Feng et al. (2020), which considers strategic corruption in the standard stochastic bandit problem under *exogenously* specified corruption budgets, where self-interested arms inflate rewards. Relative to that setting, our model differs in two important respects. First, we move from the classical bandit setting to a *ranking* problem, where *positions* enlarge the action space and alter both incentives and learning dynamics. Second, the corruption budgets in our model are determined *endogenously* by the agents rather than fixed ex ante. These differences are central to our analysis and to the stronger guarantees we obtain. This underscores the new structure induced by self-interested corruption with endogenous budget formation.

Our policy is also related to the broad class of "explore-then-exploit" algorithms and batched-learning algorithms, studied under various names and formulations; see, for example, Lattimore and Szepesvári (2020), Rusmevichientong et al. (2010), Broder and Rusmevichientong (2012), Perchet and Rigollet (2013), Sauré and Zeevi (2013), Garivier et al. (2016), Perchet et al. (2016), Cheung et al. (2017), Han et al. (2020), Birge et al. (2025), Yang and Feng (2023). In that regard, our contribution is to study how strategic data manipulation affects the performance of such policies.

Equilibrium analysis in dynamic and multiplayer systems. A central technical challenge in studying strategic manipulation is to characterize agents' equilibrium behavior. Our analysis is also related, at a broader technical level, to the literature on multi-agent learning and equilibrium approximation in large games; see, for example, Bravo et al. (2018), Mertikopoulos and Zhou (2019), Babichenko (2018). In our setting, this challenge arises because agents' manipulation incentives are induced endogenously by the learning policy. Characterizing such behavior is difficult even with a

single manipulator; see, for example, Birge et al. (2021), Keppo et al. (2022). Our technical contribution in this direction lies in the asymptotic analysis of the induced game and in quantifying convergence from the finite market to its nonatomic counterpart. Our equilibrium analysis differs from standard mean-field game formulations (Weintraub et al. 2008, Adlakha and Johari 2013, Balseiro and Gur 2019), because the learning policy in our model does not induce a (stationary) Markov game among agents. Instead, our nonatomic equilibrium is developed following the frameworks of Schmeidler (1973), Mas-Colell (1984). Moreover, we do not use the limiting game merely for asymptotic characterization; rather, we use it as a tool to obtain performance guarantees for the pre-limit market.

Ranking problems for online learners. From the application side, there is growing interest in ranking problems faced by online learners, where ranking serves as a key operational lever; see, for example, Chen et al. (2021), Niazadeh et al. (2021), Singh et al. (2021), Ferreira et al. (2022), Derakhshan et al. (2022), Gao et al. (2022), Lei et al. (2022), Asadpour et al. (2023), Zhang et al. (2022), Devic et al. (2024). More specifically, the impact of fake clicks or fake orders on ranking systems has drawn attention across operations management (Golrezaei et al. 2023), information systems (Jin et al. 2023), and marketing (Long and Liu 2022). Among these papers, Jin et al. (2023) is perhaps most closely related to ours. Under a stylized model in a two-agent-one-period environment with binary action set, they characterize equilibrium behavior under a Greedy ranking rule and identify the possibility of "defensive brushing." Our paper instead studies the *prescriptive* problem of designing a ranking policy in the presence of strategic manipulation in a multi-agent, multi-period environment under a general demand model.

Auctions, contests, and tournaments. Our model is also related, at a conceptual level, to the literature on contests, tournaments, and auctions; see, for example, Edelman et al. (2007), Varian (2007), Terwiesch and Xu (2008), Athey and Ellison (2011), Jerath et al. (2011), Körpeoğlu and Cho (2018), Ales et al. (2019), Mihm and Schlapp (2019), Bubeck et al. (2019), Conitzer et al. (2022), Liu et al. (2022), Feng et al. (2024). In our setting, agents exert costly effort to improve their relative standing, and after the dominant-strategy reduction, the ranking manipulation game resembles an all-pay position auction. The environments are nevertheless substantially different: the learner is not an auctioneer, does not observe inflation actions, and learns product quality from endogenous data generated over time. The connection is therefore primarily conceptual; it is also useful for interpreting low-noise regimes in which pure-strategy equilibria may fail to exist.

3. Model, ETC, and the Budget Competition Game

3.1. Problem Setup There are $N \geq 2$ agents, indexed by $i \in [N] := \{1, \ldots, N\}$. Each agent has a heterogeneous type, which we interpret as the agent's intrinsic quality. Qualities are normalized to lie on the uniform grid $\Theta_N := \{0, \frac{1}{N-1}, \frac{2}{N-1}, \ldots, 1\}$. (We will discuss this uniform-grid normalization further in Footnote 5.) Let $\sigma_* : [N] \to \Theta_N$ denote the true quality-based ranking, where a larger value of $\sigma_*(i)$ corresponds to a higher-quality agent. By relabeling agents in increasing order of quality, we may assume without loss of generality that $\sigma_* = \boldsymbol{e} := (0, \frac{1}{N-1}, \ldots, \frac{N-2}{N-1}, 1)$, which we refer to as the *identity ranking*. Under this labeling, agent i has quality $q^i := \frac{i-1}{N-1}$, and position under the identity ranking is $z^i := \boldsymbol{e}(i) = \frac{i-1}{N-1}$. Accordingly, the identity ranking $\boldsymbol{e}$ is also the quality-based ranking.

The learner interacts with the agents over T periods. In each period $t \in [T]$, she chooses a display ranking σ_t, represented by a bijection $\sigma_t : [N] \to \Theta_N$. Here, $\sigma_t(i)$ denotes the displayed position of agent i on the normalized grid Θ_N. Larger values of $\sigma_t(i)$ correspond to better positions, with $\sigma_t(i) = 1$ representing the top position and $\sigma_t(i) = 0$ the bottom position. Because σ_t is a bijection, each agent is assigned a distinct position, and each position is occupied by exactly one agent. Equivalently, σ_t induces a permutation of the N display positions, written in normalized form. Among all feasible rankings, the benchmark ranking is the quality-based ranking $\boldsymbol{e}$, which places higher-quality agents in higher positions.

True reward. Each agent's period-t reward depends on his intrinsic quality and displayed position. To model this dependence, let $F : [0,1]^2 \to [0, \infty)$ denote the mean reward function, where $F(q, z)$ is the expected reward of an agent with quality $q \in \Theta_N$ displayed at position $z \in \Theta_N$.[1] For agent $i \in [N]$ in period $t \in [T]$, the reward is given by

$$R_t(q^i, \sigma_t(i)) = F(q^i, \sigma_t(i)) + \epsilon_t^i. \tag{1}$$

Here, $\{\epsilon_t^i : t \in [T],\, i \in [N]\}$ are i.i.d. mean-zero sub-Gaussian random variables with variance proxy η^2, denoted by $\epsilon_t^i \sim \mathrm{subG}(0, \eta^2)$.[2] We do not impose specific functional forms on F beyond some mild structural assumptions; see Section 3.2 for details.

Observed signal. Agents may take costly actions to inflate their observed performance in order to obtain more favorable rankings and, in turn, higher rewards. Let $b_t^i \geq 0$ denote the inflation amount chosen by agent i in period t. Given displayed position $\sigma_t(i)$ and inflation amount b_t^i, the observed signal of agent i in period t is

$$D_t^i = R_t(q^i, \sigma_t(i)) + b_t^i = F(q^i, \sigma_t(i)) + b_t^i + \epsilon_t^i. \tag{2}$$

[1] Strictly speaking, F only needs to be defined on $\Theta_N \times \Theta_N$. We extend its domain to $[0,1]^2$ for notational convenience, especially for the asymptotic analysis as N grows.

[2] Specifically, a variance proxy η^2 requires that the noise term ϵ_t^i satisfies $\mathbb{E}[\exp(s\epsilon_t^i)] \leq \exp(\frac{1}{2}\eta^2 s^2)$ for every $s \in \mathbb{R}$. This standard assumption accommodates, for example, bounded and Gaussian noise distributions.

Thus, the learner observes the agent's reward only after it has been inflated by the agent's strategic action. For convenience, let $\boldsymbol{D}_t := (D_t^1, \ldots, D_t^N)$ denote the vector of observed signals in period t.

Learner and agent decisions. We consider an algorithmic learner that sequentially determines rankings based on historical observed-signal data. A learner policy is denoted by ψ and maps the public history $h_{t-1}^p := \{\sigma_1, \boldsymbol{D}_1, \ldots, \sigma_{t-1}, \boldsymbol{D}_{t-1}\}$ to the ranking σ_t in period t. The learner does not know the agents' intrinsic qualities or the reward function F. We assume that the learner commits to a ranking policy ψ at the outset and follows it throughout the process.[3]

Given ψ, each agent may strategically manipulate the data by "boosting" or inflating the observed signal D_t^i. Although costly, such manipulation shapes the data the learner observes, and can improve the agent's future ranking positions. Because higher positions yield higher rewards, agents trade a present cost for a future benefit. Formally, agent i's strategy Π^i has two components:

- At period 0, agent i chooses a corruption budget $B^i \geq 0$ at cost cB^i, which constrains his subsequent inflation.
- In each period $t \in [T]$, he chooses how much to "boost" the observed signal D_t^i within the remaining budget. We denote it by b_t^i and formally refer to it as the *inflation amount*. It is chosen adaptively as a function of the private history $h_{t-1}^i := (B^i, \sigma_1, D_1^i, b_1^i, \ldots, \sigma_{t-1}, D_{t-1}^i, b_{t-1}^i)$ taking values in the range $[0,\ B^i - \sum_{\ell=1}^{t-1} b_\ell^i]$.

In short, each agent's decision separates the budget choice from the dynamic allocation problem. This structure mirrors a feature of practice, i.e., firms typically fix spending plans in advance and have limited flexibility to revise them mid-campaign. Earlier work also uses the same structure (e.g., Golrezaei et al. 2023, Feng et al. 2020) but treats the corruption budget as exogenous. Here, each agent chooses his budget endogenously.

Writing $\boldsymbol{\Pi} = \boldsymbol{\Pi}(\psi) := (\Pi^1, \ldots, \Pi^N)$ for the resulting strategy profile, the learner's policy and the agents' strategies jointly determine the system dynamics; see Figure 2 for an illustration. In later sections, we will characterize $\boldsymbol{\Pi}(\psi)$ using appropriate equilibrium concepts.

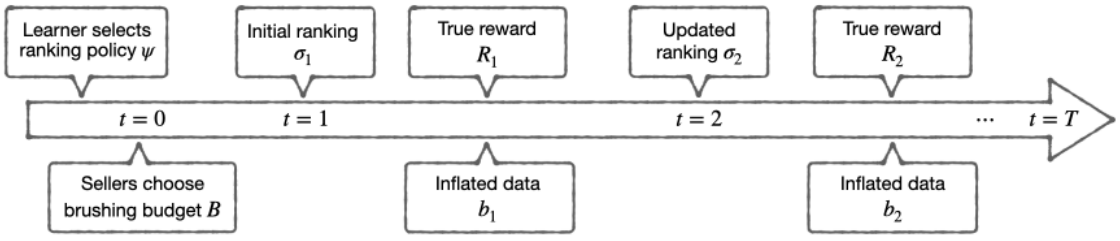


FIGURE 2. **Timeline.**

[3] To make the commitment credible, we require the ranking policy to be verifiable ex post by the agents. This condition is satisfied, for example, when the learner's ranking is a deterministic function of the public history, which we assume throughout the paper.

Learner's objective. The learner seeks to identify the quality-based benchmark ranking so as to maximize expected cumulative reward over the horizon. Under learner policy ψ and agents' strategy profile $\mathbf{\Pi}$, the learner's expected cumulative reward is $\mathbb{E}_{\psi,\mathbf{\Pi}}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} R_t(q^i,\sigma_t(i))\right]$. [4] To evaluate a learner policy ψ, we use regret, defined as the loss in expected cumulative reward relative to a clairvoyant learner that knows the agents' qualities and the reward function F. Under the structural assumptions on F introduced later, the clairvoyant learner selects the quality-based benchmark ranking $\boldsymbol{e}$ in every period. Accordingly, the regret is

$$\mathcal{R}_T \;=\; T\sum\nolimits_{i=1}^{N} F\left(q^i,z^i\right) - \mathbb{E}_{\psi,\mathbf{\Pi}}\left[\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{N} R_t(q^i,\sigma_t(i))\right].$$

Throughout the paper, we study the asymptotic growth of the learner's regret $\mathcal{R}_T$ as the horizon T increases, taking into account the agents' strategic responses. To describe these asymptotic rates, we use standard order notation. In particular, for two functions $f,g:\mathbb{Z}_+\to\mathbb{R}_+$, we write

$$f(T)=\begin{cases}\mathcal{O}(g(T)), & \text{if } \limsup_{T\to\infty} f(T)/g(T)<\infty;\\ o(g(T)), & \text{if } \limsup_{T\to\infty} f(T)/g(T)=0;\\ \Omega(g(T)), & \text{if } \liminf_{T\to\infty} f(T)/g(T)>0;\\ \omega(g(T)), & \text{if } \liminf_{T\to\infty} f(T)/g(T)=\infty.\end{cases}$$

In particular, if $\mathcal{R}_T=o(T)$, we say that the learner's ranking policy achieves *complete learning*.

Remark 1. In our convergence analysis for the nonatomic market model, we also consider a sequence of markets in which N grows. It is therefore useful to describe how regret scales when both T and N tend to infinity. To this end, we extend the order notation in the natural way: for example, we write $f(N,T)=\mathcal{O}_{N,T}(g(N,T))$ if $\limsup_{\min\{N,T\}\to\infty}\frac{f(N,T)}{g(N,T)}<\infty$. This definition allows N and T to grow jointly without imposing restriction on their relative rates. That said, in our setting the dependence on T is of primary interest, as illustrated by the following result.

Proposition 1. **(Regret order in N)** *Let $\Delta_F := F(1,1)-F(0,0)$. Under any learner ranking policy and any agent strategy profile, the learner's regret satisfies $\mathcal{R}_T\le\Delta_F NT=\mathcal{O}_{N,T}(NT)$. Conversely, there exists a problem instance for which $\mathcal{R}_T=\Omega_{N,T}(N)$.*

Proposition 1 provides universal upper and lower bounds on regret with respect to N. To streamline the presentation, we will not restate the universal upper bound in subsequent regret bounds.

[4] Throughout the paper, $\mathbb{E}_{\psi,\mathbf{\Pi}}$ and $\mathbb{P}_{\psi,\mathbf{\Pi}}$ denote expectation and probability under the system dynamics induced jointly by $(\psi,\mathbf{\Pi})$. When either ψ or $\mathbf{\Pi}$ is clear from the context, we use the abbreviated notation $\mathbb{E}_{\mathbf{\Pi}}$, $\mathbb{P}_{\mathbf{\Pi}}$, or simply $\mathbb{E}$, $\mathbb{P}$.

3.2. Assumptions and the Clairvoyant Benchmark We summarize our assumptions on the expected reward function F below.

ASSUMPTION 1. *The function F satisfies the following two properties:*

(i) MONOTONICITY: *$F(q,z)$ is strictly increasing in both quality q and ranking position z.*

(ii) SUPERMODULARITY: *$F(q,z)$ is supermodular on $\Theta_N \times \Theta_N$. That is, for all $i,j,m,n \in [N]$ such that $i>j$ and $m>n$, we have $F(q^i,z^m)-F(q^i,z^n) \geq F(q^j,z^m)-F(q^j,z^n)$.*

In the assumption above, Part (i) states that both quality and position positively affect reward. In our model, the notion of intrinsic quality enters only through its effect on the reward function F.[5] The positional effect is well documented in the literature; see, for example, Joachims et al. (2005), Craswell et al. (2008), Ursu (2018).

Part (ii) is the more substantive part of our model. One main reason we adopt this family of reward functions is that, under supermodularity, the clairvoyant's problem has a clean and intuitive solution, which is to rank the agents according to their intrinsic quality.

PROPOSITION 2. (**Clairvoyant's solution**) *If $F(q,z)$ is supermodular on $\Theta_N \times \Theta_N$, then the quality-based ranking maximizes the total expected true reward. That is,*

$$\boldsymbol{e} \in \arg\max_{\sigma} \sum\nolimits_i F(q^i, \sigma(i)).$$

When $N=2$, the converse is also true.

Under supermodular reward functions, the notion of "optimal" ranking can be naturally defined using the clairvoyant's solution, and is consistent with the usual intuitive goal of learning-to-rank problems. In addition, this result shows that, when $N=2$, the supermodularity of the reward function and the optimality of the quality-based ranking are *equivalent*; in Appendix A.1, we further extend the analysis to cases where $N>2$. The supermodularity of F also implies a useful insight. Roughly speaking, the sensitivity of high-quality agents' rewards to ranking position is greater than that of low-quality ones. This insight will be revisited throughout the paper.

The supermodular family of reward functions generalizes many widely studied models. The most prominent is the position-based model (PBM) in the ranking bandit literature, where $F(q,z) = f(q)g(z)$ with f,g positive and increasing, which makes F automatically supermodular; see Richardson et al. (2007), Lagrée et al. (2016), Ermis et al. (2020). Similar separable conditions also appear in click-through-rate models; see Aggarwal et al. (2006), Varian (2007), Edelman et al. (2007), L'Ecuyer et al. (2017). Our framework also connects to customer-search formulations. When $N=2$, it subsumes the cascade model and Weitzman-type search models (Jin et al. 2023). In Section 6.2, we will show that our results from the base model continue to hold under the cascade model for general N.

[5] The assumption that qualities lie on the uniform grid Θ_N is immaterial. Indeed, let $\tilde{\Theta}_N$ be an arbitrary grid containing the quality vector $\tilde{q}$, and let $\tilde{F}(\tilde{q}^i, z^j)$ be a reward function defined on $\tilde{\Theta}_N \times \Theta_N$. Define $F(q^i,z^j) := \tilde{F}(\tilde{q}^i, z^j)$ on $\Theta_N \times \Theta_N$. Then $\tilde{F}$ satisfies Assumption 1 if and only if F does.

3.3. The Experiment-Then-Commit Policy We next describe the learning algorithm as our proposed solution, which we call *Experiment-Then-Commit* (ETC). It is a simple policy that separates learning from exploitation. In the experimentation phase, the learner rotates the displayed ranking so that each agent occupies each position equally often. Specifically, fix N rankings $\{\kappa^j : j \in [N]\}$ defined as cyclic shifts of a baseline order $(z^1, \dots, z^N)$: $\kappa^j := (z^j, z^{j+1}, \dots, z^N, z^1, \dots, z^{j-1})$, for all $j \in [N]$. Under κ^j, agent 1 is assigned to position z^j, agent 2 to position z^{j+1}, and so on, with indices interpreted cyclically. The learner displays each ranking κ^j for $\tau_0 := \tau/N$ periods,[6] so that during the first τ periods each agent appears in every position exactly τ_0 times. Thus, this phase eliminates systematic differences in position exposure across agents.

Let $\widetilde{D}^i := \sum_{t=1}^{\tau} D_t^i$ denote agent i's total observed signal during Phase 1. The learner then forms the empirical ranking $\overline{\sigma}$ by ordering agents according to $\widetilde{D}^i$; equivalently, one may define

$$\overline{\sigma}(i) \;=\; \frac{1}{N-1} \sum_{j \in [N] \setminus \{i\}} \mathbb{I}\{\widetilde{D}^i > \widetilde{D}^j\}, \qquad i \in [N].$$

Hence, $\overline{\sigma}(i) > \overline{\sigma}(j)$ whenever $\widetilde{D}^i > \widetilde{D}^j$.[7] In the next phase, the learner displays the ranking $\overline{\sigma}$ in each of the remaining $T - \tau$ periods.

Algorithm 1 summarizes the ETC policy. It is a particularly simple policy for the dynamic learning problem studied here. It involves only a single tuning parameter, τ, uses minimal information about the model primitives, and is very easy to implement. The remainder of the paper investigates its performance in two different regimes, one in a low-manipulation regime, and the other one in a more interesting full-manipulation regime.

Algorithm 1 Experiment-Then-Commit (ETC)

INPUT: Length of the experimentation phase, τ.

PHASE 1: (*Experimentation*) Display all cyclic permutations $\{\kappa^j : j \in [N]\}$, each for $\tau_0 = \tau/N$ times.

PHASE 2: (*Commitment*) Obtain ranking $\overline{\sigma}$ by sorting the observed signals $\{\widetilde{D}^i\}$. Display ranking $\overline{\sigma}$ for the remaining $T - \tau$ periods.

[6] For the moment, we assume that τ is a multiple of N, so that $\tau_0 \in \mathbb{N}$. This assumption is inessential and will be relaxed later.

[7] If ties occur, they may be broken according to any fixed deterministic rule. In particular, when the aggregated signals admit continuous distributions, ties arise with probability zero.

3.4. The Low-Manipulation Regime We begin with the simpler case in which inflation can be effectively controlled. For each $M \geq 0$, define $\mathbf{\Pi}^M := \{\mathbf{\Pi} = (\Pi^1, \ldots, \Pi^N) : \Pi^i(h^i_{t-1}) \leq M \text{ for all } i \in [N],\ t \in [T], \text{ and histories } h^i_{t-1}\}$. Thus, $\mathbf{\Pi}^M$ is the class of strategy profiles for which every agent's per-period inflation amount is bounded by M at all times. The special case $M = 0$ recovers the manipulation-free setting. The next result shows that ETC performs well over $\mathbf{\Pi}^M$ for sufficiently small M.

PROPOSITION 3. **(Regret of ETC under low manipulation intensity)** *There exists a constant $M > 0$ such that the following holds. Suppose the agents use a strategy profile $\mathbf{\Pi} \in \mathbf{\Pi}^M$, and the learner implements ETC with experimentation length $\tau = N\lfloor C_0 \log T \rfloor$, where C_0 is a sufficiently large constant. Then $\mathcal{R}_T = \mathcal{O}(\log T)$.*

Proposition 3 shows that ETC achieves logarithmic regret whenever per-period inflation is sufficiently small. The intuition is that, if inflation is not strong enough to overturn the gaps in agents' rewards, then the cumulative observations collected during the experimentation phase preserve the correct ranking with high probability. Once this occurs, textbook concentration arguments imply that ETC commits to the optimal ranking after only $\mathcal{O}(\log T)$ experimentation periods. The logarithmic regret bound is also optimal in this manipulation-free setting. Indeed, when $N = 2$, the problem contains the classical two-armed bandit setting as a special case, so the $\Omega(\log T)$ lower bound of Lai and Robbins (1985) applies.

More broadly, Proposition 3 helps isolate the source of difficulty. Even when the aggregate amount of inflation over the T periods is linear in T, ETC achieves regret of order $\mathcal{O}(\log T)$, as long as the inflation amount in each period is controlled.[8] As a result, the proposition shows that the total volume of inflation is not by itself the relevant measure of difficulty. Rather, the difficulty arises when inflation can be temporally concentrated so as to distort the ranking precisely when the learner is attempting to identify it. We study this regime next.

3.5. The Full-Manipulation Regime and the Static Reduction We now turn to the main setting, in which agents face no exogenous per-period upper bound on inflation. This regime captures the more challenging case in which the learner has limited ability to control agents' manipulation activity.

The ranking manipulation game. In this setting, the agents' strategic responses to the learner's ranking policy form a dynamic game, which we call the *ranking manipulation game*. Formally, this is

[8] This contrasts with much of the stochastic-corruption literature, such as Lykouris et al. (2018), Golrezaei et al. (2023), Feng et al. (2020), where the regret bound typically depends directly on the corruption budget, and a corruption budget of order $\Omega(T)$ can therefore lead to regret of order $\Omega(T)$. It also contrasts with adversarial settings in which even sublinear total corruption can be concentrated in the early periods and thereby mislead the learner.

a T-period extensive-form game with N players, denoted by $\Gamma_{N,T}$. The agents' decisions follow the timing described in Section 3.1: each agent i first chooses a corruption budget B^i and then allocates this budget sequentially and adaptively over time. Agent i seeks to maximize his cumulative reward net of manipulation costs, i.e., $\sum_{t=1}^{T} R_t(q^i, \sigma_t(i)) - cB^i$. Given a strategy profile $\mathbf{\Pi}$, agent i's (ex-ante) expected utility is therefore

$$V^i(\Pi^i, \mathbf{\Pi}^{-i}) := \mathbb{E}_{\mathbf{\Pi}}\left[\sum_{t=1}^{T} R_t(q^i, \sigma_t(i))\right] - cB^i. \tag{3}$$

Similarly, for any period $t \geq 1$ and history $\boldsymbol{h}_{t-1} := \{h^i_{t-1}\}_{i\in[N]}$, let $\mathbf{\Pi}(\boldsymbol{h}_{t-1}) := \{\Pi^i(h^i_{t-1})\}_{i\in[N]}$ denote the continuation strategy profile induced by $\mathbf{\Pi}$ after history $\boldsymbol{h}_{t-1}$. Agent i's continuation value from period t onward is then $\mathbb{E}_{\mathbf{\Pi}(\boldsymbol{h}_{t-1})}\left[\sum_{\ell=t}^{T} R_\ell(q^i, \sigma_\ell(i)) \mid \boldsymbol{h}_{t-1}\right]$. This formulation allows us to apply standard dynamic equilibrium concepts, such as subgame-perfect equilibrium, to the agents' manipulation behavior.

A static reduction. Although the ranking manipulation game is dynamic, ETC induces a simple structure in agents' budget-spending decisions. Conditional on a chosen budget, an agent weakly benefits from moving any future inflation expenditure earlier in time. That renders the dynamic game essentially *static*. Formally, for agent i with remaining budget $B^i - \sum_{\ell=1}^{t-1} b^i_\ell$ after history h^i_{t-1}, define the *lump-sum inflation* (LSI) rule by $b^i_t = B^i - \sum_{\ell=1}^{t-1} b^i_\ell$. That is, after every history, LSI spends the entire remaining budget in the current period. We denote this rule by $\Pi^i_{\mathrm{LSI}}(h^i_{t-1})$. The proposition below shows that LSI is a (weakly) dominant budget-spending rule for every agent, regardless of the realized history and opponents' continuation strategies.

Proposition 4. **(LSI as a dominant budget-spending rule)** *Fix any period $t \in [T]$, history $\boldsymbol{h}_{t-1}$, continuation strategy profile $\mathbf{\Pi}(\boldsymbol{h}_{t-1})$, and budget profile $\{B^i\}$. Then the LSI rule weakly dominates any alternative continuation budget-spending rule. That is, for every agent $i \in [N]$,*

$$\mathbb{E}_{\Pi^i_{\mathrm{LSI}}(h^i_{t-1}), \mathbf{\Pi}^{-i}(\boldsymbol{h}^{-i}_{t-1})}\left[\sum_{\ell=t}^{T} R_\ell(q^i, \sigma_\ell(i)) \,\middle|\, \boldsymbol{h}_{t-1}\right] \geq \mathbb{E}_{\mathbf{\Pi}(\boldsymbol{h}_{t-1})}\left[\sum_{\ell=t}^{T} R_\ell(q^i, \sigma_\ell(i)) \,\middle|\, \boldsymbol{h}_{t-1}\right].$$

Under ETC, the experimentation rankings are fixed ex ante, while the commitment ranking depends only on the cumulative observed signals collected during experimentation. Thus, moving any inflation expenditure earlier can only weakly increase its effect on the committed ranking. Proposition 4 should be interpreted in this weak-dominance sense. LSI is defined over the full horizon: after any history, including histories after experimentation, it prescribes spending the remaining budget immediately. After experimentation, however, inflation no longer affects the committed ranking, so the agent is indifferent among all remaining spending rules. LSI is therefore one representative weakly dominant continuation rule.

Moreover, the proposition does not require all inflation to occur in the first period. Under ETC, only the total inflation accumulated during experimentation matters for the committed ranking. Hence, any rule that exhausts the chosen budget by the end of experimentation is payoff-equivalent and leads to the same static budget-competition game. The reduction therefore does not rely on a first-period spike: discarding the first-period data would merely induce agents to shift inflation to the remaining experimentation periods.

Given the preceding characterization, we can reduce the dynamic ranking manipulation game to a *static* game, denoted by $\overline{\Gamma}_{N,T}$, that focuses only on agents' budget choices. Every Nash equilibrium of this static game induces a dominant-strategy subgame-perfect equilibrium of the original dynamic game; see also Feng et al. (2020). We refer to $\overline{\Gamma}_{N,T}$ as the *budget competition game*.

In this game with N players, each agent i's action is to choose a total corruption budget B^i. For simplicity, we normalize the budget and represent it by agent $i \in [N]$'s *inflation intensity* $b^i := B^i/T$. Given an inflation-intensity profile $\boldsymbol{b} = (b^1, \ldots, b^N)$, we use $\boldsymbol{\Pi}_{T\boldsymbol{b}}$ to denote the LSI strategy profile induced by the total budgets $T\boldsymbol{b}$. Agent i's normalized utility is thus

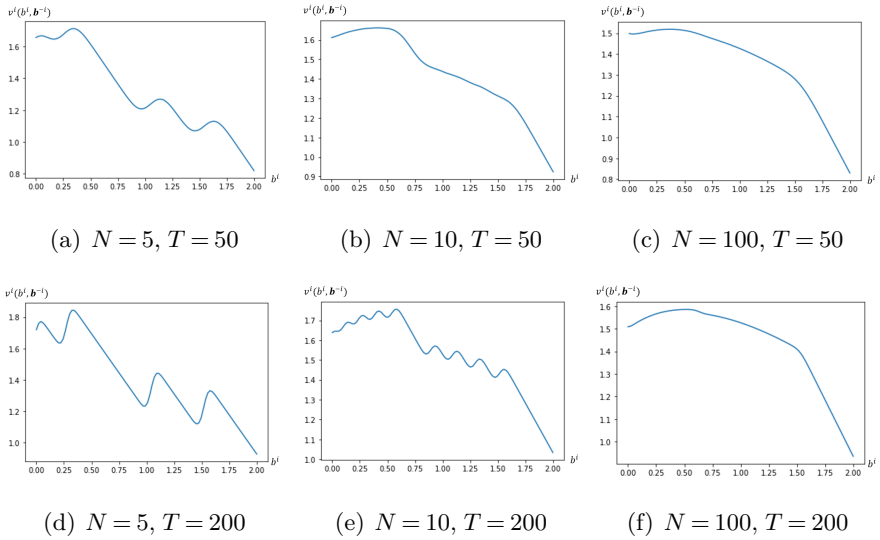
$$v^i(b^i, \boldsymbol{b}^{-i}) = \frac{1}{T} V^i(\Pi^i_{Tb^i}, \boldsymbol{\Pi}^{-i}_{T\boldsymbol{b}^{-i}}) = \frac{1}{T} \mathbb{E}_{\boldsymbol{\Pi}_{T\boldsymbol{b}}} \left[\sum_{t=1}^{T} F(q^i, \sigma_t(i)) \right] - cb^i. \tag{4}$$

The normalization by T places both actions and payoffs on a per-period scale, which makes it easier to compare across horizons.

Remaining challenges. Although the dynamic-to-static reduction substantially simplifies the analysis, the resulting static game still presents two challenges. The first is analytical and computational: the utility function v^i is difficult to evaluate because it depends on the joint distribution of $\{D^i_t\}$, whose dimension grows with N and T.

The second challenge is structural: the utility functions may be highly irregular, displaying strong non-concavity and multiple local optima. To put it into perspective, consider a sequence of markets with the same primitives (e.g., reward function F and inflation cost c), except for market size N and horizon T. Figure 3 plots the utility of an agent with fixed quality as N and T vary. As can be seen, the utilities can be highly irregular in small markets (e.g., $N = 5$). In Section 6.1, we show that this irregularity can make Nash equilibria fragile, especially when the horizon is long, signal variance is small, and inflation costs are low.

4. Asymptotic Game Analysis To address these challenges, we now study the budget competition game in an asymptotic regime where both the number of agents N and the time horizon T are large. To build intuition, note that Figure 3 illustrates the two forces associated with N and T. A large T makes agents' utilities highly sensitive to accumulated signals, which creates nonconcavities and payoff irregularities in the finite-agent game. These irregularities tend to make pure-strategy



(a) $N = 5, T = 50$ (b) $N = 10, T = 50$ (c) $N = 100, T = 50$

(d) $N = 5, T = 200$ (e) $N = 10, T = 200$ (f) $N = 100, T = 200$

FIGURE 3. **Utility functions for different values of N and T.** The figure plots the utility of an agent with fixed quality $q = 0.5$ as a function of his inflation intensity. We set $F(q, z) = (q + 1)(z + 1)$, $c = 1$, $\tau = \lfloor \log T \rfloor$, and assume that the shocks $\{\epsilon_t^i\}$ are standard normal. The utility is $v^i(\cdot, \boldsymbol{b}^{-i})$, where the inflation intensities of all other agents are given by $\overline{\pi}(q) = \frac{q^2+2q}{2c}$.

equilibria fragile. By contrast, a large N smooths individual ranking externalities, suggesting that a continuum model may provide a tractable approximation to the original finite market.

To cleanly separate these forces and derive the limiting game, we develop the convergence theory in two steps. First, fixing N and letting $T \to \infty$ yields an interim long-run game $\overline{\Gamma}_{N,\infty}$ in which utilities admit closed-form, though discontinuous, expressions. Second, we let $N \to \infty$ and show that $\overline{\Gamma}_{N,\infty}$ converges to a continuum limiting game, denoted by $\overline{\Gamma}_{\infty,\infty}$. We then analyze this continuum game by finding an equilibrium in the limit. Table 2 in the appendix provides a summary of the games and a roadmap of the results.

4.1. The Long-Run Interim Limit ($\overline{\Gamma}_{N,\infty}$). We now fix N and let $T \to \infty$. In the limit, the model reduces to an N-player normal-form game. Agent i's action is denoted by b^i, which represents his long-run average inflation per period. We define an *inflation-intensity based ranking* among the N agents. For any profile $\boldsymbol{b} \in \mathbb{R}_+^N$, define agent i's ranking position by

$$H^N(b^i \mid \boldsymbol{b}^{-i}) = \tfrac{1}{N-1} \sum_{j \in [N] \setminus \{i\}} \mathbb{I}\{b^j < b^i\}. \tag{5}$$

Thus, $H^N(b^i \mid \boldsymbol{b}^{-i})$ is the fraction of opponents whose inflation intensities are below b^i. In particular, if $\boldsymbol{b}$ has no ties (i.e., $b^i \neq b^j$ for all $i \neq j$), then $H^N(b^i \mid \boldsymbol{b}^{-i}) > H^N(b^j \mid \boldsymbol{b}^{-j})$ whenever $b^i > b^j$. [9] Given

[9] The current formulation assigns tied agents the less favorable rank among the tied positions. In general, the induced ranking may depend on the tie-breaking rule. Appendix C.2 shows that our results are robust to a broad class of tie-breaking rules.

$\boldsymbol{b}^{-i}$, agent i's long-run utility is

$$\overline{v}^i(b^i, \boldsymbol{b}^{-i}) = F\big(q^i, H^N(b^i \mid \boldsymbol{b}^{-i})\big) - cb^i. \tag{6}$$

Having defined the game, we now build intuition for why it is a plausible formulation through a back-of-the-envelope argument. Consider an agent who chooses a positive inflation intensity $b > 0$ in the interim game $\overline{\Gamma}_{N,\infty}$. The corresponding total corruption budget in the finite game is of order $\Omega(T)$. Under LSI, this budget is allocated to the experimentation phase, whose length satisfies $\tau = o(T)$. Hence $\Omega(T) \gg o(T)$, and cumulative inflation dominates true reward during experimentation. The learner's ranking is therefore asymptotically determined by the agents' inflation intensities.

Theorem 1 below formalizes the convergence from the finite game $\overline{\Gamma}_{N,T}$ to the interim limit game $\overline{\Gamma}_{N,\infty}$ as the horizon T grows.

THEOREM 1. **(Convergence from $\overline{\Gamma}_{N,T}$ to $\overline{\Gamma}_{N,\infty}$)** *Fix $N \geq 2$ and consider ETC with experimentation length $\tau = o(\sqrt{T})$. Let $\boldsymbol{b} \in \mathbb{R}_+^N$ be any inflation-intensity profile with no ties.*

(i) Suppose that in the game $\overline{\Gamma}_{N,T}$, all agents use LSI with total budgets $T\boldsymbol{b}$. Then, as $T \to \infty$,

$$\mathbb{P}\left(\overline{\sigma}(i) \neq H^N(b^i \mid \boldsymbol{b}^{-i})\right) \to 0 \quad \text{and} \quad v^i(b^i, \boldsymbol{b}^{-i}) \to \overline{v}^i(b^i, \boldsymbol{b}^{-i}), \quad \forall i \in [N].$$

In other words, the committed ranking converges to the inflation-intensity based ranking, and the finite-game utilities converge to the long-run utilities at point $\boldsymbol{b}$.

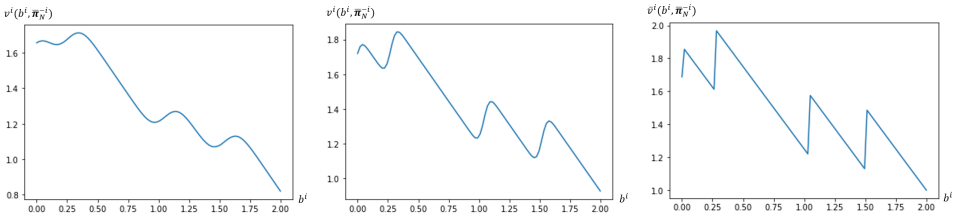
(ii) Suppose $\boldsymbol{b}$ is an ε'-equilibrium of $\overline{\Gamma}_{N,\infty}$ for some $\varepsilon' \geq 0$. That is, for every agent $i \in [N]$, and $b \geq 0$, $\overline{v}^i(b^i, \boldsymbol{b}^{-i}) \geq \overline{v}^i(b, \boldsymbol{b}^{-i}) - \varepsilon'$. Then for every $\varepsilon > 0$, $\boldsymbol{b}$ is an $(\varepsilon + \varepsilon')$-equilibrium of $\overline{\Gamma}_{N,T}$ for all sufficiently large T. That is, for every agent $i \in [N]$, and $b \geq 0$,

$$v^i(b^i, \boldsymbol{b}^{-i}) \geq v^i(b, \boldsymbol{b}^{-i}) - (\varepsilon + \varepsilon').$$

Theorem 1 establishes two complementary results. Part (i) formalizes the intuition that the learner's committed ranking is asymptotically determined by agents' inflation intensities; consequently, the pre-limit utility v^i converges pointwise to its long-run counterpart $\overline{v}^i$ at every tie-free profile. Part (ii) lifts this pointwise convergence to an equilibrium statement: any ε'-equilibrium of $\overline{\Gamma}_{N,\infty}$ remains an $(\varepsilon + \varepsilon')$-equilibrium of $\overline{\Gamma}_{N,T}$ for all sufficiently large T, with the additional ε vanishing as $T \to \infty$. The formulation covers both exact Nash equilibria ($\varepsilon' = 0$) and approximate ones, which serves as a bridge between $\overline{\Gamma}_{N,\infty}$ and the nonatomic market $\overline{\Gamma}_{\infty,\infty}$ that we will develop later.

A main effect of the long-run limit is that the role of stochastic fluctuations vanishes asymptotically. As a result, the utility functions admit simple closed-form expressions. At the same time, this leads to an important observation: as T increases, the finite-game utility converges to a jagged, step-like limit, with each jump corresponding to the agent overtaking one more opponent in inflation intensity; this

phenomenon reflects an intrinsic difficulty when $N \ll T$, and is illustrated in Figure 4. As established in Section 6.1, this ill-behaved feature eventually leads to fragility of Nash equilibria. That motivates the second step of our asymptotic analysis, where we let $N \to \infty$ and study the resulting continuum limit.



(a) $T = 50$ (b) $T = 200$ (c) $T = \infty$

FIGURE 4. **Utility functions for $N = 5$ and various T.** Under the same setting as in Figure 3, we focus on the agent with quality 0.5 and illustrate how this agent's utility function changes as T grows. In panels (a) and (b), we plot the utility function $v^i(\cdot, \overline{\boldsymbol{\pi}}_N^{-i})$. In panel (c), we plot the limiting function $\overline{v}^i(\cdot, \overline{\boldsymbol{\pi}}_N^{-i})$.

4.2. The Continuum Limit and the Nonatomic Model ($\overline{\Gamma}_{\infty,\infty}$) We now introduce the *nonatomic market* model, where a continuum of agents with total mass one play the manipulation game in the long run. Later, we show that this is the limit of the finite game $\overline{\Gamma}_{N,T}$ as both N and T grow large.

We now define each element of this nonatomic market. Recall that in the finite model, agents' qualities and ranking positions lie on the uniform mesh Θ_N. In the nonatomic model, they are described by continuous variables $q, z \in [0,1]$. Both variables have a quantile interpretation: an agent of quality q has higher quality than a mass q of agents, and an agent placed at position z is ranked above a mass z of agents. Thus, larger values of q and z correspond to higher quality and better ranking positions, respectively. Agents' qualities are uniformly distributed on $[0,1]$. The learner's reward from assigning quality q to position z is described by a function $F : [0,1]^2 \to [0,\infty)$. For tractability, we impose the following assumption.

ASSUMPTION 2. *The expected reward function $F(q,z)$ is continuously differentiable and supermodular on $[0,1]^2$. Moreover, $F_1(q,z) := \frac{\partial}{\partial q}F(q,z) > 0$ and $F_2(q,z) := \frac{\partial}{\partial z}F(q,z) > 0$ for all $(q,z) \in [0,1]^2$, so that F is strictly increasing in both arguments.*

We now formulate the nonatomic game, denoted by $\overline{\Gamma}_{\infty,\infty}$. As in the finite game $\overline{\Gamma}_{N,T}$, we specify the agents' action space, payoff functions, and equilibrium concept. An action profile in the nonatomic game is a measurable function $\pi : [0,1] \to \mathbb{R}_+$. For an agent with quality q, the action $\pi(q)$ represents his long-run inflation intensity. Following the interim long-run game, we define the nonatomic payoff

through the ranking induced by agents' inflation intensities. We call this ordering the *inflation-intensity based ranking*. Given an inflation profile π, let

$$H(b \mid \pi) := \int_0^1 \mathbb{I}\{\pi(q) \le b\}\, dq. \tag{7}$$

The function $H(\cdot \mid \pi)$ is the distribution function of inflation intensities in the nonatomic market. In particular, $H(b \mid \pi)$ is the mass of agents whose inflation intensity is no greater than b, and hence gives the position quantile associated with inflation intensity b. Thus, higher inflation corresponds to a weakly higher ranking position. Holding the population profile π fixed, an agent of quality q who chooses inflation intensity b obtains long-run utility

$$U_\pi(q, b) = F(q, H(b \mid \pi)) - cb. \tag{8}$$

The first term is the reward associated with the inflation-induced position quantile $H(b \mid \pi)$, while the second term is the per-period inflation cost. As is standard in nonatomic games, an individual agent is infinitesimal and therefore takes the aggregate distribution $H(\cdot \mid \pi)$ as given when choosing his own inflation intensity. We use the following equilibrium concept, in the spirit of Schmeidler (1973), Mas-Colell (1984).

Definition 1. An action profile π is a *nonatomic equilibrium* (NAE) if, for every $q \in [0,1]$,

$$\pi(q) \in \arg\max_{b \ge 0} U_\pi(q, b).$$

In an NAE, almost every agent chooses an inflation intensity that maximizes his long-run utility given the aggregate inflation profile. Thus, an NAE is the continuum-agent analogue of Nash equilibrium, with the distinction that each individual agent has zero mass and hence does not affect the aggregate ranking distribution.

The next result connects the nonatomic equilibrium to the finite-agent interim game. It has two parts. First, for any continuous and strictly increasing inflation profile, the discrete inflation-intensity based rankings and interim utilities in $\overline{\Gamma}_{N,\infty}$ uniformly approximate their nonatomic counterparts, as illustrated in Figure 5. Second, if the limiting profile is an NAE, then its discretization forms an approximate equilibrium of the large finite-agent interim game. Therefore, the nonatomic market is not only a tractable limiting model, but also an equilibrium approximation to $\overline{\Gamma}_{N,\infty}$ when N is large.

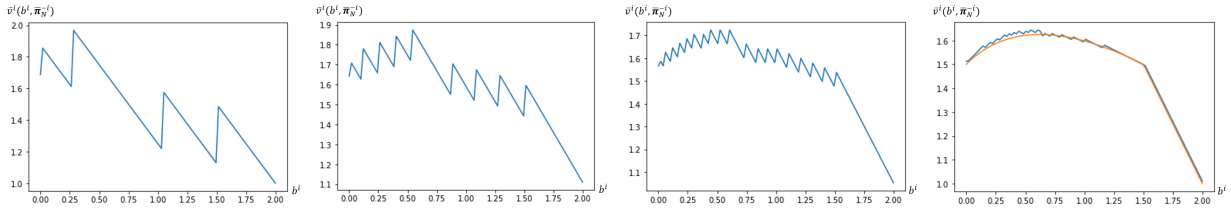
Theorem 2. **(Convergence from $\overline{\Gamma}_{N,\infty}$ to $\overline{\Gamma}_{\infty,\infty}$)** *Let $\pi : [0,1] \to [0,\infty)$ be any continuous and strictly increasing function. For each $N \ge 2$, define the induced inflation-intensity profile $\boldsymbol{\pi}_N := \{\pi(q^i) : i \in [N]\}$. The following statements hold.*

(i) There exists a constant $L_F > 0$, depending only on the reward function F, such that

$$\sup_{b \ge 0,\, i \in [N]} \left| H^N(b \mid \boldsymbol{\pi}_N^{-i}) - H(b \mid \pi) \right| \le \frac{1}{N-1}, \quad \text{and} \quad \sup_{b \ge 0,\, i \in [N]} \left| \overline{v}^i(b, \boldsymbol{\pi}_N^{-i}) - U_\pi(q^i, b) \right| \le \frac{L_F}{N-1}.$$

(ii) If π is an NAE, then for every $\varepsilon > 0$, $\boldsymbol{\pi}_N$ is an ε-equilibrium of $\overline{\Gamma}_{N,\infty}$ for all sufficiently large N.

Intuitively, $\overline{\Gamma}_{N,\infty}$ is a discretized version of $\overline{\Gamma}_{\infty,\infty}$. In the long-run game, agents lie on the finite grid Θ_N, while in the nonatomic game, they form a continuum on $[0,1]$. As N grows, the grid becomes denser, and the empirical distribution of inflation intensities converges to its continuum counterpart. Part (i) formalizes this idea by showing that the finite-agent inflation-intensity based ranking H^N uniformly approximates H, and that the corresponding long-run utilities uniformly approximate U_π. Figure 5 illustrates this convergence. As N increases, the jagged utility curves become smoother because the payoff gain from overtaking a single adjacent opponent becomes smaller in a larger population.



(a) $N = 5$ (b) $N = 10$ (c) $N = 20$ (d) $N = 100\,(\infty)$

FIGURE 5. **Utility functions for $T = \infty$ and various N.** Under the same setting as in Figure 3, we focus on the agent with quality 0.5 and illustrate how this agent's limiting utility function changes as N grows. Specifically, in panels (a) through (d), we plot the limiting utility function $\overline{v}^i(\cdot, \overline{\boldsymbol{\pi}}_N^{-i})$ in the blue curve. In panel (d), we also plot the nonatomic utility function $U_{\overline{\pi}}(0.5, \cdot)$ as the orange curve.

Part (ii) then translates this payoff convergence into an equilibrium statement. If no infinitesimal agent can profitably deviate in the nonatomic market, then, for sufficiently large N, no finite agent can gain more than ε by deviating in the interim game. In this sense, an NAE provides an approximate equilibrium characterization of large finite-agent long-run markets. Combined later with the finite-to-interim convergence result, this approximation will allow us to return to the original finite-(N,T) game and derive finite-market performance guarantees.

4.3. Solving the Nonatomic Equilibrium Thus far, we have introduced a parsimonious nonatomic market as a limiting object and shown that its equilibria are relevant in finite markets. It remains to be seen whether a strictly increasing and continuous NAE does exist. The following result provides a positive answer in a simple closed form.

THEOREM 3. **(Closed-form nonatomic equilibrium)** *Define*

$$\overline{\pi}(q) := \int_0^q \frac{F_2(t,t)}{c}\, dt, \qquad q \in [0,1],$$

where $F_2(q,z) = \partial F(q,z)/\partial z$. *Then* $\overline{\pi}$ *is a nonatomic equilibrium.*

Theorem 3 gives a simple solution to the nonatomic game, in contrast to the complexity of the finite game $\overline{\Gamma}_{N,T}$. The equilibrium also carries significant insight into agents' responses to ETC. First, although a higher inflation cost c reduces manipulation incentives, the equilibrium inflation intensity satisfies $\overline{\pi}(q) > 0$ for all $q > 0$. Thus, except for the lowest type, agents choose positive inflation in equilibrium. The reason is that the nonatomic game bases the commitment ranking on relative observed signals. Consequently, each agent has an incentive to raise his signal in order to secure a favorable relative position.

Second and more interestingly, the inflation intensity $\overline{\pi}(q)$ is strictly increasing in the agent's quality q. This means that, at equilibrium, ETC can induce the agents' manipulation actions to "reinforce" their quality-based ranking. Intuitively, the validity of the NAE is closely related to the supermodularity condition of F. As a consequence, the high-quality agents have "more to lose."

5. Performance Guarantees in the Pre-limit Having characterized a nonatomic equilibrium $\overline{\pi}$, we now show how it approximates the original finite-(N,T) budget-competition game. Fix N and T. If each finite-market agent follows the nonatomic prescription, the induced action profile is $\overline{\boldsymbol{\pi}}_N := \{\overline{\pi}(q^i) : i \in [N]\}$. We use it as a description of agents' strategic response to the learner's ETC policy in the finite game.

5.1. Equilibrium Guarantees The next theorem establishes two facts. First, when all other agents follow $\overline{\boldsymbol{\pi}}_N$, each agent's finite-game payoff converges uniformly to his nonatomic payoff. The budget competition game converges to the nonatomic market model with a convergence rate of $\mathcal{O}_{N,T}\Big(\frac{\log T}{T} + \frac{1}{N}\Big)$. Second, this convergence implies that $\overline{\boldsymbol{\pi}}_N$ is an $\mathcal{O}_{N,T}\Big(\frac{\log T}{T} + \frac{1}{N}\Big)$-equilibrium of the finite budget-competition game.[10]

THEOREM 4. **(Approximate equilibrium guarantee)** *The action profile* $\overline{\boldsymbol{\pi}}_N$ *satisfies the following approximation guarantee under ETC:*

$$\sup_{b \geq 0,\, i \in [N]} \left| v^i\big(b, \overline{\boldsymbol{\pi}}_N^{-i}\big) - U_{\overline{\pi}}(q^i, b) \right| \leq \frac{\varepsilon}{2},$$

where

$$\varepsilon = \delta_1 \tfrac{\tau}{T} + \delta_2 \tfrac{\log T}{T} + \delta_3 \exp\Big(-\tfrac{1}{4\eta^2} \tfrac{\log^2 T}{\tau}\Big) + \delta_4 \tfrac{1}{N-1},$$

[10] Theorem 4 allows arbitrary N, T, and τ, including cases in which τ is not a multiple of N (e.g., $\tau \leq N$). Accordingly, in this section we use the natural extension of ETC that rotates rankings until period τ and then commits. The rounding error is negligible in large markets, as shown in the proof.

and $\delta_1, \delta_2, \delta_3, \delta_4$ are positive constants independent of N and T, and specified in the proof. Consequently, $\overline{\boldsymbol{\pi}}_N$ is an ε-equilibrium of the finite budget-competition game $\overline{\Gamma}_{N,T}$. That is,

$$\sup_{b\geq 0,\, i\in[N]} \left\{ v^i\big(b, \overline{\boldsymbol{\pi}}_N^{-i}\big) - v^i\big(\overline{\pi}(q^i), \overline{\boldsymbol{\pi}}_N^{-i}\big) \right\} \leq \varepsilon.$$

As a result, under the choice of $\tau = \mathcal{O}_{N,T}(1)$, $\overline{\boldsymbol{\pi}}_N$ forms an $\mathcal{O}_{N,T}\Big(\frac{\log T}{T} + \frac{1}{N}\Big)$-equilibrium.

Theorem 4 provides a *non-asymptotic* and *uniform* link between the finite game and the nonatomic game. The first part shows that, when all other agents follow the nonatomic prescription, agent i's finite-game utility $v^i(b, \overline{\boldsymbol{\pi}}_N^{-i})$ is uniformly close to the limiting utility $U_{\overline{\pi}}(q^i, b)$ over all agents and deviations b. Thus, it implies that the budget competition game converges to the nonatomic market model with a rate of $\mathcal{O}_{N,T}\Big(\frac{\log T}{T} + \frac{1}{N}\Big)$. The second part translates this payoff convergence into an equilibrium approximation. Since $\overline{\pi}$ is optimal in the nonatomic game, the finite profile $\overline{\boldsymbol{\pi}}_N$ leaves each agent with at most a utility gain of ε from unilateral deviation. Hence, agents' incentives to deviate from $\overline{\boldsymbol{\pi}}_N$ vanish as N and T become large. In this sense, $\overline{\boldsymbol{\pi}}_N$ provides a justified approximation to equilibrium behavior in the finite game $\overline{\Gamma}_{N,T}$. Numerically, even a moderate N already provides a high-quality approximation; see Table 1.

Proof sketch and key intuition. We briefly explain why the finite budget-competition game converges to the nonatomic market game. The convergence is driven by three forces. The first two forces are reflected in the previous convergence theory: as T grows, cumulative inflation becomes increasingly influential relative to reward, whereas as N grows, each ranking position becomes infinitesimal, which creates a smoothing effect.

These two forces may interact in a nontrivial way when agents' inflation intensities are close: a small difference in inflation can determine the ordering, while finite-sample noise may still perturb the realized ranking. The third force addresses this issue. The equilibrium $\overline{\pi}$ characterized in Theorem 3 separates agents' inflation intensities in a sufficiently regular way. This separation ensures that, in the joint limit as N and T grow, the realized ranking remains inflation-intensity based with high probability, which yields the convergence result.

Remark 2. Theorems 1 and 2 provide a two-step convergence methodology from the finite game $\overline{\Gamma}_{N,T}$ to the nonatomic market $\overline{\Gamma}_{\infty,\infty}$. The first step (Theorem 1, $T \to \infty$ at fixed N) shows almost-everywhere pointwise utility convergence from $\overline{\Gamma}_{N,T}$ to the interim game $\overline{\Gamma}_{N,\infty}$ and lifts any ε'-equilibrium of $\overline{\Gamma}_{N,\infty}$ into an $(\varepsilon+\varepsilon')$-equilibrium of $\overline{\Gamma}_{N,T}$ under the stated conditions, with ε vanishing as $T \to \infty$. The second step (Theorem 2, $N \to \infty$) shows uniform utility convergence from $\overline{\Gamma}_{N,\infty}$ to $\overline{\Gamma}_{\infty,\infty}$, and lifts any continuous and strictly increasing NAE into an approximate equilibrium of $\overline{\Gamma}_{N,\infty}$ for sufficiently large N. Composing the two steps, any continuous and strictly increasing NAE of $\overline{\Gamma}_{\infty,\infty}$ prescribes an approximate equilibrium of the original finite game $\overline{\Gamma}_{N,T}$ when both N and T are large.

N	Time Horizon T				
	20	50	100	500	1000
20	1.84%	2.32%	2.41%	2.27%	2.23%
50	0.46%	0.70%	0.83%	0.88%	0.91%
100	0.11%	0.35%	0.44%	0.45%	0.46%
500	0.02%	0.05%	0.06%	0.08%	0.09%

Table 1. Relative sub-optimality gap (%) for the mid-quality agent ($q = 0.5$) under NAE. We consider the specification $F(q, z) = (q + 1)(z + 1)$ and $c = 1$, and assume that the noise terms $\{\epsilon_t^i\}$ are i.i.d. and normally distributed with mean zero and standard deviation 0.5. Smaller values correspond to weaker incentives to deviate.

In comparison, Theorem 4 sharpens this qualitative convergence into a non-asymptotic guarantee. By exploiting the closed-form NAE profile $\overline{\pi}$, it directly compares $\overline{\Gamma}_{N,T}$ with $\overline{\Gamma}_{\infty,\infty}$ and yields an explicit uniform error bound of $\mathcal{O}_{N,T}(\log T/T + 1/N)$, without requiring a specific order of the limits in N and T. In short, the two-step results lead to the nonatomic formulation, whereas Theorem 4 delivers the quantitative pre-limit guarantee used in the equilibrium and regret analysis. The same methodology is applied again in Section 6.2 to the cascade reward model.

5.2. Regret Guarantees. The nonatomic equilibrium analysis provides a tractable characterization of agents' limiting behavior. We now translate this characterization into a non-asymptotic regret guarantee for ETC in the original finite system.

THEOREM 5. **(Regret of ETC under strategic response)** *Consider ETC with a fixed experimentation length τ. Its regret satisfies*

$$\mathcal{R}_T \le \min\left\{\phi_1 N + \phi_2 cN \log T,\, \phi_3 N^3\right\}$$

under agents' LSI strategy response with total budgets $T\overline{\boldsymbol{\pi}}_N$, where the constants ϕ_1, ϕ_2, and ϕ_3 are independent of N and T and given in closed form in the proof. As a result, under the choice of $\tau = \mathcal{O}_{N,T}(1)$, the regret of ETC grows on the order of $\mathcal{O}_{N,T}(N \log T)$.

Theorem 5 characterizes the regret performance of ETC in the finite ranking manipulation game when agents follow the equilibrium inflation profile induced by the nonatomic analysis. For fixed N, the regret is bounded in T; that is, $\mathcal{R}_T = \mathcal{O}(1)$. This dependence on T is order-optimal. Interestingly, the equilibrium inflation generated by ETC can make the commitment ranking more sharply aligned with quality than in the manipulation-free setting, where ETC incurs regret of order $\mathcal{O}(\log T)$.

The theorem also gives a joint bound in N and T, i.e., it achieves $\mathcal{R}_T \le \phi_1 N + \phi_2 cN \log T = \mathcal{O}_{N,T}(N \log T)$. This matches the lower bound $\Omega_{N,T}(N)$ up to a logarithmic factor. Thus, ETC learns the optimal ranking effectively using only aggregate observed signals, despite being agnostic to the reward model and requiring no additional algorithmic sophistication (such as the "debiasing" algorithms in Golrezaei et al. 2023), or operational/informational levers (such as fraud or anomaly detection). Section 6.2 shows that the same ETC policy continues to perform well when rewards are generated through a cascade structure.

REMARK 3. The regret bound also clarifies the role of the inflation cost c. One might expect learning to become more difficult as c decreases, since lower costs make manipulation more attractive. In contrast, our analysis shows that sufficiently low inflation costs can improve the performance of ETC. In particular, if $c = \mathcal{O}_{N,T}(1/\log T)$, then the regret bound improves to $\mathcal{O}_{N,T}(N)$, which is jointly order-optimal in N and T. Thus, ETC can learn the optimal ranking not only when inflation is tightly controlled, but also in the opposite regime where manipulation-blocking measures are ineffective.

6. Further Analysis and Extensions This section complements the main results from two perspectives. Section 6.1 studies equilibria in finite-player games and shows that *exact* Nash equilibria can be fragile, thereby motivating the use of approximate equilibria. Section 6.2 extends the analysis to a cascade reward model and establishes analogues of the main results.

6.1. Nash Equilibria with Finite Agents The preceding analysis shows that the NAE induces an $o_{N,T}(1)$-approximate Nash equilibrium in the finite game. A natural question is whether one can strengthen this conclusion to exact Nash equilibrium. In this section, we show that pure-strategy equilibria may fail to exist in the hard regimes of the model. Thus, approximate equilibrium is not merely an analytical convenience. It is the relevant finite-game benchmark for describing behavior in regimes where exact pure-strategy equilibria are fragile or nonexistent.

Equilibrium in $\overline{\Gamma}_{2,T}$. We begin with the two-agent finite-horizon game. The next proposition characterizes when pure-strategy Nash equilibria exist and when they fail to exist.

PROPOSITION 5. *Let τ be an even integer. Suppose the noise terms $\{\epsilon_t^i\}$ are i.i.d. normal with variance η^2, and the reward function F is strictly increasing and strictly supermodular.*

(i) *No pure-strategy Nash equilibrium (b^1, b^2) exists with $b^1 > 0$. Consequently, any pure-strategy equilibrium, if it exists, must take the form $(0, b^2)$.*

(ii) *The manipulation-free action profile $(b^1, b^2) = (0, 0)$ is a Nash equilibrium if $c > \frac{F(1,1)(T-\tau)}{\eta\sqrt{2\tau}}$.*

(iii) *No pure-strategy Nash equilibrium exists if any of the following conditions holds: (1) T is sufficiently large; (2) c is sufficiently small; or (3) $(T-\tau)\frac{F(0,1)-F(0,0)}{F(1,1)+F(1,0)-F(0,1)-F(0,0)} > c\tau$, and η is sufficiently small.*

Proposition 5 offers three insights. First, Part (i) implies that, in any pure-strategy equilibrium, the lower-quality agent chooses zero inflation. Thus, whenever a pure-strategy equilibrium exists, the higher-quality agent inflates weakly more than the lower-quality agent. This is consistent with the monotone inflation pattern identified in the nonatomic analysis. Second, Part (ii) gives a sufficient condition under which the manipulation-free action profile is an equilibrium. Hence, the no- or low-manipulation regime studied in Section 3.4 can arise endogenously under appropriate parameter

values. This equilibrium is also *socially efficient*: the learner identifies the optimal ranking without any agent incurring manipulation costs. The result formalizes the intuitive "easier" case in which sufficiently high manipulation costs deter agents from distorting observed signals.

Third, Part (iii) shows that pure-strategy equilibria can fail to exist in "harder" regimes, including long horizons, low-noise environments, and low-cost manipulation. In these regimes, small changes in inflation can decisively affect the committed ranking, which destabilizes pure-strategy equilibrium behavior. These theoretical findings are consistent with our numerical experiments, which are reported in Figure 6.

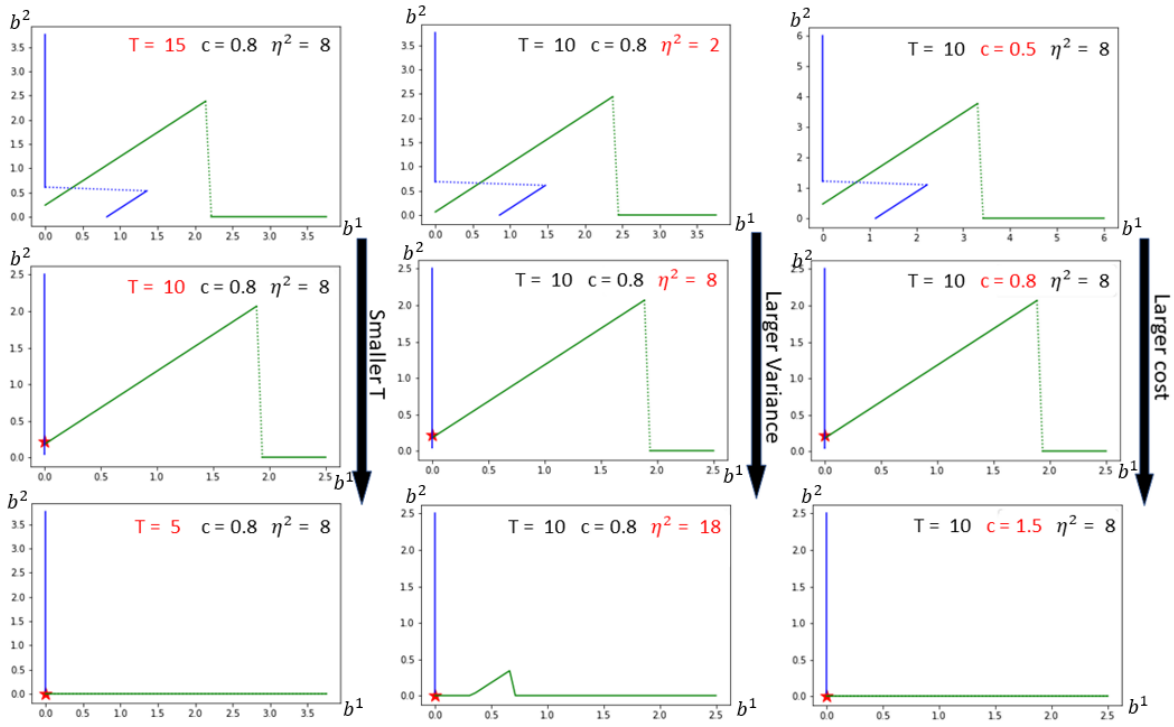


FIGURE 6. **Best Response Plots for $\overline{\Gamma}_{2,T}$.** We pick $F(1,1)=6, F(1,0)=3, F(0,1)=4, F(0,0)=2$, and assume that $\{\epsilon_t^i\}$ follow the normal distribution with mean zero and variance η^2. The x-axis (resp. y-axis) represents the inflation intensity of the low-quality (resp. high-quality) agent. The blue (resp. green) curve represents the best response of the low-quality (resp. high-quality) agent. The intersection of the best response curves represents a pure-strategy Nash equilibrium (red star).

Equilibrium in $\overline{\Gamma}_{N,\infty}$. We next consider the case with more than two agents. Although the finite game is generally difficult to analyze directly, the interim long-run game $\overline{\Gamma}_{N,\infty}$ introduced in Section 4.1 has closed-form utilities and therefore provides a tractable way to study equilibrium structure. The next result shows that pure-strategy equilibria fail to exist in this limit.

PROPOSITION 6. *No pure-strategy Nash equilibrium exists in $\overline{\Gamma}_{N,\infty}$.*

The intuition is as follows. As $T \to \infty$, stochastic fluctuations in the accumulated signals vanish, and rankings become determined by inflation intensities. This creates discontinuous incentives around ranking position changes. If an agent is ranked above another by a positive margin, he can reduce his inflation slightly and save cost without changing his rank. Conversely, an agent just below another may profitably increase inflation slightly to overtake that agent. When agents tie, the outcome depends on the tie-breaking rule, and any agent disadvantaged by the rule has an incentive to increase inflation marginally. These local incentives prevent any pure action profile from being stable. The same logic is not specific to ETC; it also arises more broadly in ranking-based learning environments when noise and manipulation costs are sufficiently small.

This nonexistence result is consistent with our earlier findings. In the two-agent finite game, the existence of pure-strategy equilibria depends on the horizon, signal variance, and manipulation cost. In the long-run limit, where rankings are deterministic and ranking position changes occur discontinuously, pure-strategy equilibria disappear altogether. This reinforces our choice of solution concept: in hard regimes such as when T is large relative to N, relaxed notions such as ε-equilibria may be inevitable.

6.2. The Cascade Reward Model A simplifying feature of the main model is that an agent's reward depends only on his own quality and position. In this subsection, we extend the analysis to a cascade reward model, in which an agent's reward may also depend on the qualities of agents ranked above him.

Consider a learner with N agents over a horizon of T periods. In each period, user arrivals follow a Poisson process with mean λN, where λ represents the average arrival rate per agent. Each arriving user examines agents sequentially from the top-ranked position to the bottom-ranked one. Upon examining agent i, the user generates a reward with probability $r(q^i)$, where q^i is agent i's quality and $r : [0,1] \to [0,1]$ is increasing. If no reward is generated, the user either proceeds to the next position with probability $1-\zeta$ or exits with probability ζ, where $\zeta \in (0,1)$ captures limited attention or patience. If the user reaches the end of the ranking without generating a reward, the user exits.

Under this cascade model, the probability that a given user generates a reward for agent i under ranking $\sigma : [N] \to \Theta_N$ is

$$\mathbb{P}_i(\sigma) = r(q^i) \prod_{j:\sigma(j)>\sigma(i)} [(1-r(q^j))(1-\zeta)].$$

Thus, with a slight abuse of notation, the expected reward of agent i at time t can be written as $\mathbb{E}[R_t^i] = F(q^i, \sigma_t)$, where

$$F(q^i, \sigma) := \lambda N r(q^i) \prod_{j:\sigma(j)>\sigma(i)} [(1-r(q^j))(1-\zeta)]. \tag{9}$$

Here, F is a reward function under cascade reward formation. Unlike in the main model, the expected reward of agent i depends not only on his own quality and position, but also on the qualities of all agents ranked above the agent. Accordingly, the reward function is written as $F(q^i, \sigma)$, depending on the full ranking σ, rather than only on the position $\sigma(i)$.

Except for the reward model, we keep all other settings and notation as in the base setting. In particular, we consider the same ETC policy in Algorithm 1. Our goal is to replicate the analytical framework and main results of the base model, and show that this "naive" design, which is agnostic to the cascade model structure, still achieves good performance.

LSI responses. We first extend Proposition 4 to the cascade reward model. The same dominant-strategy logic applies: conditional on a chosen budget, LSI is weakly dominant for every agent, regardless of the realized history and opponents' continuation strategies.

PROPOSITION 7. **(LSI as a dominant strategy under cascade rewards)** *Fix any period $t \in [T]$, history $\boldsymbol{h}_{t-1}$, and continuation strategy profile $\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})$. For every agent $i \in [N]$,*

$$\mathbb{E}_{\Pi^i_{\mathrm{LSI}}(h^i_{t-1}),\, \boldsymbol{\Pi}^{-i}(\boldsymbol{h}^{-i}_{t-1})} \left[\sum_{\ell=t}^{T} R_\ell(q^i, \sigma_\ell) \,\middle|\, \boldsymbol{h}_{t-1} \right] \geq \mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})} \left[\sum_{\ell=t}^{T} R_\ell(q^i, \sigma_\ell) \,\middle|\, \boldsymbol{h}_{t-1} \right].$$

Proposition 7 implies that the dynamic ranking manipulation game again reduces to a static budget-competition game, as in Section 3.5. In this static game, agent i chooses an inflation intensity b^i, and his normalized utility is $v^i(b^i, \boldsymbol{b}^{-i}) = \frac{1}{T}\mathbb{E}_{\boldsymbol{\Pi}_{T\boldsymbol{b}}}\left[\sum_{t=1}^{T} F(q^i, \sigma_t)\right] - cb^i$.

The nonatomic market under cascade rewards. We next formulate the nonatomic version of the cascade game and relate it to the finite-N game. To ensure a nontrivial scaling regime, we take the reward-generation probability as $r(q^i) = q^i/N$ and the exit probability as $\zeta = 1/N$. This scaling factor of $1/N$ yields a nondegenerate large-market limit: as N grows, the aggregate arrival volume scales with market size on the order of $\Theta(N)$, and each individual agent with a fixed quality quantile receives an effective reward on the order of $\Theta(1)$, which is the same as the base model.

We now derive the limiting utility function and characterize the corresponding nonatomic equilibrium. The main technical challenge is that, under cascade rewards, an agent's payoff depends not only on his own position but also on the *entire* ranking σ, through the agents ranked above him. Nevertheless, the cascade structure leads to a well-defined limiting payoff. Specifically, fix a strictly increasing inflation profile π and consider an agent with quality $q \in (0, 1)$. If the agent unilaterally deviates to inflation intensity b, his long-run utility is

$$U_\pi(q, b) = \lambda q \exp\left(\frac{H^2(b \mid \pi)}{2} + H(b \mid \pi) - \frac{3}{2} \right) - cb, \tag{10}$$

where $H(b \mid \pi)$ denotes the inflation-intensity based ranking position defined in (7). Theorem 7 below formally justifies this limiting utility representation.

Given the limiting utility in (10), the notion of nonatomic equilibrium is defined analogously to Definition 1. The next result provides a closed-form equilibrium for the cascade reward model.

THEOREM 6. **(NAE under cascade rewards)** *Define* $\widetilde{\pi}:[0,1]\to\mathbb{R}_+$ *by*

$$\widetilde{\pi}(q) := \frac{\lambda}{c}\int_0^q t(t+1)\exp\left(\frac{t^2}{2}+t-\frac{3}{2}\right)dt.$$

Then $\widetilde{\pi}$ *is an NAE of the nonatomic cascade game.*

Theorem 6 is the cascade analogue of Theorem 3. Although the equilibrium profile takes a different functional form because rewards are generated through a cascade structure, it preserves the same qualitative property as in the base model: $\widetilde{\pi}$ is smooth and strictly increasing. Thus, equilibrium inflation remains aligned with quality. The same logic also yields the cascade analogue of Theorem 4. In particular, the finite cascade game converges to the nonatomic cascade game as N and T grow, and the NAE induces an approximate equilibrium in the finite game. Formally, for each fixed N, let $\widetilde{\boldsymbol{\pi}}_N := \{\widetilde{\pi}(q^i) : i\in[N]\}$ denote the finite-agent profile obtained by applying the nonatomic equilibrium prescription to each agent. The next result shows that $\widetilde{\boldsymbol{\pi}}_N$ provides a high-quality approximation to equilibrium behavior in the finite cascade game when N and T are large.

THEOREM 7. **(Connection to the finite cascade game)** *Fix* $T\in\mathbb{N}_+$ *and* $N\geq 3$*, and set* $\tau = 3\log(N^2T)/\lambda$*. Then*

$$\sup_{b\geq 0,\, i\in[N]} \left|v^i\big(b,\widetilde{\boldsymbol{\pi}}_N^{-i}\big) - U_{\widetilde{\pi}}(q^i,b)\right| \leq \frac{\varepsilon}{2},$$

where

$$\varepsilon = \delta_1\frac{\log(N^2T)}{T} + \delta_2\frac{1}{N^{5/4}T^{9/8}} + \delta_3\sqrt{\frac{\log(N^2T)}{T}} + \delta_4\frac{1}{N} = \mathcal{O}_{N,T}\left(\sqrt{\frac{\log(NT)}{T}}+\frac{1}{N}\right).$$

and $\delta_1,\delta_2,\delta_3,\delta_4$ *are positive constants specified in the proof. Consequently,* $\widetilde{\boldsymbol{\pi}}_N$ *is an* ε*-equilibrium of the finite cascade game:*

$$\sup_{b\geq 0,\, i\in[N]} \left\{v^i\big(b,\widetilde{\boldsymbol{\pi}}_N^{-i}\big) - v^i\big(\widetilde{\pi}(q^i),\widetilde{\boldsymbol{\pi}}_N^{-i}\big)\right\} \leq \varepsilon.$$

Theorem 7 is the cascade analogue of Theorem 4. It shows that, when the experimentation length is chosen appropriately, the finite-game utility $v^i(\cdot,\widetilde{\boldsymbol{\pi}}_N^{-i})$ converges uniformly to the nonatomic utility $U_{\widetilde{\pi}}(q^i,\cdot)$, with an explicit non-asymptotic error bound. Up to logarithmic factors, the convergence rate is $\widetilde{\mathcal{O}}_{N,T}(1/\sqrt{T}+1/N)$. As a consequence, the profile $\widetilde{\boldsymbol{\pi}}_N$ forms an $\widetilde{\mathcal{O}}_{N,T}(1/\sqrt{T}+1/N)$-equilibrium of the finite cascade game.

We are now ready to analyze the performance of ETC in the finite cascade system. The equilibrium characterization above mirrors a key insight from the main model: higher-quality agents choose higher inflation intensities in equilibrium. This monotonicity aligns the inflation-induced ranking with the quality-based ranking and leads to the following regret guarantee.

THEOREM 8. **(Regret of ETC under cascade rewards)** *Let the experimentation length be* $\tau = 3\log(N^2T)/\lambda$. *Under the learner's ETC policy and the agents' LSI strategies with total budgets* $T\widetilde{\boldsymbol{\pi}}_N$, *the learner's regret satisfies*

$$\mathcal{R}_T \le \phi_1 N \log(N^2T) + \phi_2 = \mathcal{O}_{N,T}\big(N\log(NT)\big),$$

where the constants ϕ_1 *and* ϕ_2 *are given in closed form in the proof.*

Analogous to Theorem 5, Theorem 8 provides a non-asymptotic regret guarantee for ETC under cascade rewards. The regret is of order $\mathcal{O}_{N,T}(N\log(NT))$. Although this bound differs slightly from the base case due to the additional dependence structure induced by the cascade model, it matches the global lower bound $\Omega_{N,T}(N)$ up to logarithmic factors. Thus, the performance guarantee for ETC extends beyond the base reward model to environments in which rewards are generated through a cascade structure. This is notable because ETC does not use knowledge of either the reward function or the cascade structure; it relies only on aggregated observed signals.

7. Concluding Remarks This paper studies a learning and ranking problem under strategic data manipulation by multiple agents. We show that learning need not take the form of an arms race between the learner and the agents through detection and evasion. Instead, the learner can sometimes leverage agents' strategic incentives, allowing a simple policy to perform well. We establish this point in a sequential learning model with heterogeneous agents and a general reward structure. In particular, we analyze ETC, a simple Experiment-Then-Commit policy, and show that under agents' strategic responses, inflation need not prevent the learner from identifying the correct ranking. The key step is the characterization of a nonatomic market that emerges when the number of agents and the time horizon are large. We further provide convergence results that connect this limiting market to the original finite system and yield non-asymptotic performance guarantees.

Several directions remain for future work. First, this paper focuses on the inflationary side of manipulation, where agents increase their own observed signals. A natural extension is to study negative manipulation, where agents may instead harm competitors' observed signals, for example through negative reviews or other forms of sabotage. Second, our results show that the learner can recover the correct ranking under manipulation with limited algorithmic sophistication and few operational levers. It would be useful to understand the additional value of richer informational or operational tools, such as fraud detection, signal filtering, or personalized ranking, especially in more complex environments. Third, our convergence theory currently focuses on strictly increasing NAEs; the existence of non-monotone NAEs, as well as their connection to the pre-limit game, is left for future research.

Appendix

TABLE 2. **Summary of the game formulations.**

	$\Gamma_{N,T}$ *(original)*	$\overline{\Gamma}_{N,T}$ *budget competition*	$\overline{\Gamma}_{N,\infty}$ *interim long-run limit*	$\overline{\Gamma}_{\infty,\infty}$ *nonatomic limit*
Regime	dynamic; finite N,T	static; finite N,T	$T\to\infty$; finite N	$N,T\to\infty$
Action/Strategy	Π^i	$b^i = B^i/T$	b^i	$\pi(\cdot)$
Utility	$V^i(\Pi^i, \boldsymbol{\Pi}^{-i})$	$v^i(b^i, \boldsymbol{b}^{-i})$	$\overline{v}^i(b^i, \boldsymbol{b}^{-i})$	$U_\pi(q,b)$
Ranking position	$\sigma_t(i)$	$\overline{\sigma}(i)$	$H^N(b^i \mid \boldsymbol{b}^{-i})$	$H(b \mid \pi)$
Solution concept	dominant-strategy SPE	ε-Nash	ε-Nash	nonatomic (NAE)
Defined/Characterized in	Prop. 4	Thms. 4, 5; Prop. 5	Thm. 1, Prop. 6	Thms. 2, 3

References

Adlakha, Sachin, Ramesh Johari. 2013. Mean field equilibrium in dynamic games with strategic complementarities. *Operations Research* **61**(4) 971–989.

Aggarwal, Gagan, Ashish Goel, Rajeev Motwani. 2006. Truthful auctions for pricing search keywords. *ACM Conference on Electronic Commerce*. 1–7. URL http://doi.acm.org/10.1145/1134707.1134708.

Ales, Laurence, Soo-Haeng Cho, Ersin Körpeoğlu. 2019. *Innovation and Crowdsourcing Contests*. Springer International Publishing, Cham, 379–406. doi:10.1007/978-3-030-01863-4_16. URL https://doi.org/10.1007/978-3-030-01863-4_16.

Asadpour, Arash, Rad Niazadeh, Amin Saberi, Ali Shameli. 2023. Sequential submodular maximization and applications to ranking an assortment of products. *Operations Research* **71**(4) 1154–1170.

Athey, Susan, Glenn Ellison. 2011. Position Auctions with Consumer Search*. *The Quarterly Journal of Economics* **126**(3) 1213–1270. doi:10.1093/qje/qjr028. URL https://doi.org/10.1093/qje/qjr028.

Babichenko, Yakov. 2018. Fast convergence of best-reply dynamics in aggregative games. *Mathematics of Operations Research* **43**(1) 333–346. doi:10.1287/moor.2017.0868. URL https://doi.org/10.1287/moor.2017.0868.

Balseiro, Santiago R., Yonatan Gur. 2019. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science* **65**(9) 3952–3968. doi:10.1287/mnsc.2018.3174.

Birge, John R., Hongfan Chen, N.Bora Keskin. 2025. Markdown policies for demand learning with forward-looking customers. *Operations Research* **73** 2550–2566.

Birge, John R., Yifan Feng, N. Bora Keskin, Adam Schultz. 2021. Dynamic learning and market making in spread betting markets with informed bettors. *Operations Research* **69**(6) 1746–1766. doi:10.1287/opre.2021.2109. URL https://doi.org/10.1287/opre.2021.2109.

Bogunovic, Ilija, Andreas Krause, Jonathan Scarlett. 2020. Corruption-tolerant gaussian process bandit optimization. *International Conference on Artificial Intelligence and Statistics*. PMLR, 1071–1081.

Boucheron, Stéphane, Gábor Lugosi, Pascal Massart. 2013. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press. doi:10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Braverman, Mark, Jieming Mao, Jon Schneider, S. Matthew Weinberg. 2019. Multi-armed bandit problems with strategic arms. Alina Beygelzimer, Daniel Hsu, eds., *Proceedings of the Thirty-Second Conference on Learning Theory*, *Proceedings of Machine Learning Research*, vol. 99. PMLR, 383–416. URL https://proceedings.mlr.press/v99/braverman19b.html.

Bravo, Mario, David Leslie, Panayotis Mertikopoulos. 2018. Bandit learning in concave n-person games. *Advances in Neural Information Processing Systems* **31**.

Broder, Josef, Paat Rusmevichientong. 2012. Dynamic pricing under a general parametric choice model. *Operations Research* **60**(4) 965–980.

Bubeck, Sébastien, Nikhil R. Devanur, Zhiyi Huang, Rad Niazadeh. 2019. Multi-scale online learning: Theory and applications to online auctions and pricing. *Journal of Machine Learning Research* **20**(62) 1–37. URL http://jmlr.org/papers/v20/17-498.html.

Chen, Ningyuan, Anran Li, Shuoguang Yang. 2021. Revenue maximization and learning in products ranking. *Proceedings of the 22nd ACM Conference on Economics and Computation*. 316–317.

Chen, Xi, Jianjun Gao, Dongdong Ge, Zizhuo Wang. 2022. Bayesian dynamic learning and pricing with strategic customers. *Production and Operations Management* **31**(8) 3125–3142. doi:https://doi.org/10.1111/poms.13741. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.13741`.

Chen, Xi, Akshay Krishnamurthy, Yining Wang. 2023. Robust dynamic assortment optimization in the presence of outlier customers. *Operations Research* .

Chen, Xi, Yining Wang. 2023. Robust dynamic pricing with demand learning in the presence of outlier customers. *Operations Research* **71**(4) 1362–1386. doi:10.1287/opre.2022.2280. URL `https://doi.org/10.1287/opre.2022.2280`.

Cheung, Wang Chi, David Simchi-Levi, He Wang. 2017. Technical note-Dynamic pricing and demand learning with limited price experimentation. *Operations Research* **65**(6) 1722–1731.

Conitzer, Vincent, Christian Kroer, Eric Sodomka, Nicolas E Stier-Moses. 2022. Multiplicative pacing equilibria in auction markets. *Operations Research* **70**(2) 963–989.

Craswell, Nick, Onno Zoeter, Michael Taylor, Bill Ramsey. 2008. An experimental comparison of click position-bias models. *Proceedings of the 2008 international conference on web search and data mining*. 87–94.

Derakhshan, Mahsa, Negin Golrezaei, Vahideh Manshadi, Vahab Mirrokni. 2022. Product ranking on online platforms. *Management Science* **68**(6) 4024–4041. doi:10.1287/mnsc.2021.4044. URL `https://doi.org/10.1287/mnsc.2021.4044`.

Devic, Siddartha, Aleksandra Korolova, David Kempe, Vatsal Sharan. 2024. Stability and multigroup fairness in ranking with uncertain predictions. *Proceedings of the 41st International Conference on Machine Learning*. ICML'24, JMLR.org.

Dong, Jing, Ke Li, Shuai Li, Baoxiang Wang. 2022. Combinatorial bandits under strategic manipulations. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 219–229.

Edelman, Benjamin, Michael Ostrovsky, Michael Schwarz. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review* **97**(1) 242–259. doi:10.1257/aer.97.1.242. URL `https://www.aeaweb.org/articles?id=10.1257/aer.97.1.242`.

Ermis, Beyza, Patrick Ernst, Yannik Stein, Giovanni Zappella. 2020. Learning to rank in the position based model with bandit feedback. *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM)*, vol. 48. 2405 – 2412.

Feng, Yiding, Brendan Lucier, Aleksandrs Slivkins. 2024. Strategic budget selection in a competitive autobidding world. *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 213–224.

Feng, Zhe, David Parkes, Haifeng Xu. 2020. The intrinsic robustness of stochastic bandits to strategic manipulation. *International Conference on Machine Learning*. PMLR, 3092–3101.

Ferreira, Kris J., Sunanda Parthasarathy, Shreyas Sekar. 2022. Learning to rank an assortment of products. *Management Science* **68**(3) 1828–1848. doi:10.1287/mnsc.2021.4130. URL `https://doi.org/10.1287/mnsc.2021.4130`.

Folland, Gerald B. 1999. *Real analysis: modern techniques and their applications*. John Wiley & Sons.

Gao, Xiangyu, Stefanus Jasin, Sajjad Najafi, Huanan Zhang. 2022. Joint learning and optimization for multi-product pricing (and ranking) under a general cascade click model. *Management Science* **68**(10) 7362–7382.

Garivier, Aurelien, Tor Lattimore, Emilie Kaufmann. 2016. On explore-then-commit strategies. D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett, eds., *Advances in Neural Information Processing Systems*, vol. 29. URL `https://proceedings.neurips.cc/paper_files/paper/2016/file/ef575e8837d065a1683c022d2077d342-Paper.pdf`.

Golrezaei, Negin, Vahideh Manshadi, Jon Schneider, Shreyas Sekar. 2023. Learning product rankings robust to fake users. *Operations Research* **71**(4) 1171–1196. doi:10.1287/opre.2022.2380. URL `https://doi.org/10.1287/opre.2022.2380`.

Gupta, Anupam, Tomer Koren, Kunal Talwar. 2019. Better algorithms for stochastic bandits with adversarial corruptions. *Conference on Learning Theory*. PMLR, 1562–1578.

Han, Yanjun, Zhengqing Zhou, Zhengyuan Zhou, Jose Blanchet, Peter W Glynn, Yinyu Ye. 2020. Sequential batch learning in finite-action linear contextual bandits. *arXiv preprint arXiv:2004.06321* .

Harris, Keegan, Chara Podimata, Steven Z Wu. 2023. Strategic apple tasting. *Advances in Neural Information Processing Systems* **36** 79918–79945.

Jerath, Kinshuk, Liye Ma, Young-Hoon Park, Kannan Srinivasan. 2011. A "position paradox" in sponsored search auctions. *Marketing Science* **30**(4) 612–627. doi:10.1287/mksc.1110.0645. URL `https://doi.org/10.1287/mksc.1110.0645`.

Jin, Chen, Luyi Yang, Kartik Hosanagar. 2023. To brush or not to brush: Product rankings, consumer search, and fake orders. *Information Systems Research* **34**(2) 532–552. doi:10.1287/isre.2022.1128. URL `https://doi.org/10.1287/isre.2022.1128`.

Joachims, Thorsten, Laura Granka, Bing Pan, Helene Hembrooke, Geri Gay. 2005. Accurately interpreting clickthrough data as implicit feedback. *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '05, Association for Computing Machinery, New York, NY, USA, 154–161. doi:10.1145/1076034.1076063. URL `https://doi.org/10.1145/1076034.1076063`.

Keppo, Jussi, Michael Jong Kim, Xinyuan Zhang. 2022. Learning manipulation through information dissemination. *Operations Research* **70**(6) 3490–3510. doi:10.1287/opre.2021.2209. URL `https://doi.org/10.1287/opre.2021.2209`.

Kleine Buening, Thomas, Aadirupa Saha, Christos Dimitrakakis, Haifeng Xu. 2024a. Bandits meet mechanism design to combat clickbait in online recommendation. *International Conference on Learning Representations*, vol. 2024. 10275–10303.

Kleine Buening, Thomas, Aadirupa Saha, Christos Dimitrakakis, Haifeng Xu. 2024b. Strategic linear contextual bandits. *Advances in Neural Information Processing Systems* **37** 116638–116675.

Körpeoğlu, Ersin, Soo-Haeng Cho. 2018. Incentives in contests with heterogeneous solvers. *Management Science* **64**(6) 2709–2715.

Lagrée, Paul, Claire Vernade, Olivier Cappé. 2016. Multiple-play bandits in the position-based model. *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS'16, Curran Associates Inc., Red Hook, NY, USA, 1605–1613.

Lai, T.L., Herbert Robbins. 1985. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics* **6** 4–22.

Lattimore, Tor, Branislav Kveton, Shuai Li, Csaba Szepesvari. 2018. Toprank: A practical algorithm for online stochastic ranking. *Advances in Neural Information Processing Systems* **31**.

Lattimore, Tor, Csaba Szepesvári. 2020. *Stochastic Bandits with Finitely Many Arms*. Cambridge University Press, 87–92.

Lei, Yanzhe, Stefanus Jasin, Joline Uichanco, Andrew Vakhutinsky. 2022. Joint product framing (display, ranking, pricing) and order fulfillment under the multinomial logit model for e-commerce retailers. *Manufacturing & Service Operations Management* **24**(3) 1529–1546.

Liu, Lydia T, Nikhil Garg, Christian Borgs. 2022. Strategic ranking. *International Conference on Artificial Intelligence and Statistics*. PMLR, 2489–2518.

Long, Fei, Yunchuan Liu. 2022. Fake sales and ranking algorithms in online retail marketplace with sponsored advertising. *Available at SSRN 4009128* .

Lykouris, Thodoris, Vahab Mirrokni, Renato Paes Leme. 2018. Stochastic bandits robust to adversarial corruptions. *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 114–122.

L'Ecuyer, Pierre, Patrick Maillé, Nicolás E. Stier-Moses, Bruno Tuffin. 2017. Revenue-maximizing rankings for online platforms with quality-sensitive consumers. *Operations Research* **65**(2) 408–423. doi:10.1287/opre.2016.1569. URL `https://doi.org/10.1287/opre.2016.1569`.

Mas-Colell, Andreu. 1984. On a theorem of schmeidler. *Journal of Mathematical Economics* **13**(3) 201–206.

Mas-Colell, Andreu, Michael Dennis Whinston, Jerry R Green, et al. 1995. *Microeconomic theory*, vol. 1. Oxford university press New York.

Mertikopoulos, Panayotis, Zhengyuan Zhou. 2019. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming* **173**(1) 465–507.

Mihm, Jürgen, Jochen Schlapp. 2019. Sourcing innovation: On feedback in contests. *Management Science* **65**(2) 559–576. doi:10.1287/mnsc.2017.2955. URL `https://doi.org/10.1287/mnsc.2017.2955`.

Niazadeh, Rad, Negin Golrezaei, Joshua R Wang, Fransisca Susan, Ashwinkumar Badanidiyuru. 2021. Online learning via offline greedy algorithms: Applications in market design and optimization. *Proceedings of the 22nd ACM Conference on Economics and Computation*. 737–738.

Perchet, Vianney, Philippe Rigollet. 2013. The multi-armed bandit problem with covariates. *The Annals of Statistics* **41** 693–721.

Perchet, Vianney, Philippe Rigollet, Sylvain Chassang, Erik Snowberg. 2016. Batched bandit problems. *The Annals of Statistics* **44** 660–681.

Richardson, Matthew, Ewa Dominowska, Robert Ragno. 2007. Predicting clicks: estimating the click-through rate for new ads. *Proceedings of the 16th international conference on World Wide Web*. 521–530.

Rusmevichientong, Paat, Zuo-Jun Max Shen, David B Shmoys. 2010. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations research* **58**(6) 1666–1680.

Sauré, Denis, Assaf Zeevi. 2013. Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management* **15**(3) 387–404.

Schmeidler, David. 1973. Equilibrium points of nonatomic games. *Journal of statistical Physics* **7**(4) 295–300.

Singh, Ashudeep, David Kempe, Thorsten Joachims. 2021. Fairness in ranking under uncertainty. *Advances in Neural Information Processing Systems* **34** 11896–11908.

Terwiesch, Christian, Yi Xu. 2008. Innovation contests, open innovation, and multiagent problem solving. *Management Science* **54**(9) 1529–1543. doi:10.1287/mnsc.1080.0884.

Ursu, Raluca M. 2018. The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science* **37**(4) 530–552. doi:10.1287/mksc.2017.1072. URL `https://doi.org/10.1287/mksc.2017.1072`.

Varian, Hal R. 2007. Position auctions. *International Journal of Industrial Organization* **25**(6) 1163–1178. doi:https://doi.org/10.1016/j.ijindorg.2006.10.002.

Wang, Jingyuan, Perry Dong, Ying Jin, Ruohan Zhan, Zhengyuan Zhou. 2024. Adaptively learning to select-rank in online platforms. *arXiv preprint arXiv:2406.05017* .

Weintraub, Gabriel Y, C Lanier Benkard, Benjamin Van Roy. 2008. Markov perfect industry dynamics with many firms. *Econometrica* **76**(6) 1375–1411.

Yang, Junwen, Yifan Feng. 2023. Nested elimination: a simple algorithm for best-item identification from choice-based feedback. *International Conference on Machine Learning*. PMLR, 39205–39233.

Zhang, Zijin, Hyun-Soo Ahn, Lennart Baardman. 2022. Ordering and ranking products for an online retailer. *Available at SSRN 4061071* .

Zhong, Zixin, Wang Chi Cheung, Vincent Tan. 2021. Probabilistic sequential shrinking: A best arm identification algorithm for stochastic bandits with corruptions. *International Conference on Machine Learning*. PMLR, 12772–12781.

Online Appendix

A. Further Analysis of the Reward Model (Assumption 1)

A.1. Relationship with Supermodularity Here, we will explore the connection between supermodularity and optimality of the quality-based ranking more deeply. We say that the reward function F is *diagonally supermodular* on $\Theta_N \times \Theta_N$ if supermodularity only needs to hold on diagonal entries. That is, $F(q^i, z^i) - F(q^i, z^j) \geq F(q^j, z^i) - F(q^j, z^j)$. The result below is a strengthened version of Proposition 2.

PROPOSITION EC.1. *If $F(q, z)$ is supermodular on $\Theta_N \times \Theta_N$, then the quality-based ranking $\boldsymbol{e}$ is optimal. Conversely, if $\boldsymbol{e}$ is optimal, then $F(q, z)$ is diagonally supermodular on $\Theta_N \times \Theta_N$. When $N = 2$, all three properties are equivalent.*

Proof of Proposition EC.1. First, suppose $F(q, z)$ is supermodular on $\Theta_N \times \Theta_N$. Pick an arbitrary ranking $\sigma \neq \boldsymbol{e}$, and agents i, j such that $i > j$ while $\sigma(i) < \sigma(j)$. We can construct a new ranking σ' by exchanging the positions of agents i and j while keeping all other agents' positions the same as in σ. That is, $\sigma'(i) = \sigma(j)$, $\sigma'(j) = \sigma(i)$ and $\sigma'(m) = \sigma(m)$ for all $m \neq i, j$. It holds that

$$\begin{aligned}
\textstyle\sum_{s=1}^N F(q^s, \sigma'(s)) &= \textstyle\sum_{s\in\{1,\dots,N\}\setminus\{i,j\}} F(q^s, \sigma'(s)) + F(q^i, \sigma'(i)) + F(q^j, \sigma'(j)) \\
&\overset{(a)}{=} \textstyle\sum_{s\in\{1,\dots,N\}\setminus\{i,j\}} F(q^s, \sigma(s)) + F(q^i, \sigma(j)) + F(q^j, \sigma(i)) \\
&\overset{(b)}{\geq} \textstyle\sum_{s\in\{1,\dots,N\}\setminus\{i,j\}} F(q^s, \sigma(s)) + F(q^i, \sigma(i)) + F(q^j, \sigma(j)) \\
&= \textstyle\sum_{s=1}^N F(q^s, \sigma(s)).
\end{aligned}$$

Therefore, the total expected reward is increased. In the derivations above, part (a) is by the construction of σ'. Part (b) is by the supermodularity of $F(q, z)$ after taking $m = \sigma(j)$ and $n = \sigma(i)$, which implies that $F(q^i, \sigma(j)) + F(q^j, \sigma(i)) \geq F(q^i, \sigma(i)) + F(q^j, \sigma(j))$. This "swapping" argument can be used repeatedly until $\sigma = \boldsymbol{e}$. In other words, $\boldsymbol{e} \in \arg\max_\sigma \sum_{s=1}^N F(q^s, \sigma(s))$.

We next prove that optimality of $\boldsymbol{e}$ implies diagonal supermodularity. Let an arbitrary $i, j \in [N]$ be such that $i > j$. Let σ be a ranking that swaps the positions of agent i and j, i.e., $\sigma(i) = z^j$, $\sigma(j) = z^i$ and $\sigma(s) = z^s$ for all $s \neq i, j$. Optimality of $\boldsymbol{e}$ implies that $\sum_{s=1}^N F(q^s, z^s) \geq \sum_{s=1}^N F(q^s, \sigma(s))$. Cancel out terms on both sides that correspond to $s \notin \{i, j\}$, and we have $F(q^i, z^i) + F(q^j, z^j) \geq F(q^j, z^i) + F(q^i, z^j)$. That is precisely what we need for diagonal supermodularity.

Finally, it suffices to check that when $N = 2$, all three conditions are equivalent to $F(1, 1) + F(0, 0) \geq F(1, 0) + F(0, 1)$. ■

A.2. Universal Regret Bounds **Proof of Proposition 1.** Notice that a universal upper bound of the regret, regardless of the learner's policy and the agents' strategy profile, is given by

$$\mathcal{R}_T = \mathbb{E}\Big[\sum_{t=1}^T \Big(\sum_{i=1}^N F(q^i, z^i) - \sum_{i=1}^N F(q^i, \sigma_t(i))\Big)\Big] \leq \Delta_F T N,$$

which is linear in N. Conversely, note that the learner needs to pick the initial ranking σ_1 *prior to* seeing any data in the first period. Without loss, let it be $\sigma_1 = (1, \frac{N-2}{N-1}, \cdots, 0)$ based on a relabeling argument. That is, agent 1 is ranked at the top, agent 2 is ranked at the second, and so on. However,

suppose it happens that the *ground truth* quality-based ranking is $\boldsymbol{e} = (0, \frac{1}{N-1}, \dots, 1)$. That is, agent 1 has the lowest quality, agent 2 has the second lowest, and so on. Then

$$\begin{aligned}
\mathcal{R}_T &= \mathbb{E}\Big[\sum_{t=1}^{T}\Big(\sum_{i=1}^{N} F(q^i, z^i) - \sum_{i=1}^{N} F(q^i, \sigma_t(i))\Big)\Big] \\
&\geq \sum_{i=1}^{N} F(q^i, z^i) - \sum_{i=1}^{N} F(q^i, \sigma_1(i)) \\
&= \sum_{i=1}^{N} F(\tfrac{i-1}{N-1}, \tfrac{i-1}{N-1}) - \sum_{i=1}^{N} F(\tfrac{i-1}{N-1}, 1 - \tfrac{i-1}{N-1}) \\
&= N\Big[\underbrace{\frac{1}{N}\sum_{i=1}^{N} F(\tfrac{i-1}{N-1}, \tfrac{i-1}{N-1})}_{*} - \underbrace{\frac{1}{N}\sum_{i=1}^{N} F(\tfrac{i-1}{N-1}, 1 - \tfrac{i-1}{N-1})}_{**}\Big],
\end{aligned}$$

where terms $(*)$ and $(**)$ converge to $\int_0^1 F(x,x)dx$ and $\int_0^1 F(x,1-x)dx$, respectively, as N grows. At the same time, note that

$$\begin{aligned}
&\int_0^1 \big[F(x,x) - F(x,1-x)\big]\, dx \\
&= \int_0^{1/2} \big[F(x,x) - F(x,1-x)\big]\, dx + \int_{1/2}^1 \big[F(x,x) - F(x,1-x)\big]\, dx \\
&\overset{(a)}{=} \int_0^{1/2} \big[F(x,x) - F(x,1-x)\big]\, dx + \int_0^{1/2} \big[F(1-x,1-x) - F(1-x,x)\big]\, dx \\
&= \int_0^{1/2} \Big\{F(x,x) + F(1-x,1-x) - F(x,1-x) - F(1-x,x)\Big\}\, dx \\
&\overset{(b)}{>} 0,
\end{aligned}$$

where part (a) is from substituting x by $1-x$ in the second integral, and part (b) is because the integrand is *strictly* positive for all $x \in [0, \frac{1}{2}]$ due to the strict supermodularity of F. As a result, it is straightforward to verify that $\mathcal{R}_T = \Omega_{N,T}(N)$. ■

B. Preliminary Analysis of ETC

B.1. Performance of ETC in the Low-Manipulation Regime (Proposition 3) Preliminaries At the end of the experimentation phase under the learner's ETC policy, every agent appears in every position for τ_0 periods. Therefore, for every $i \in [N]$, the total observed signal for agent i during the experimentation phase can be represented by:

$$\widetilde{D}^i = \tau_0 \sum_{j=1}^{N} F(q^i, z^j) + \sum_{t=1}^{\tau} (b_t^i + \epsilon_t^i) = N\tau_0 \mu^i + \sum_{t=1}^{\tau} (b_t^i + \epsilon_t^i). \tag{EC.1}$$

Here we have used the shorthand notation $\mu^i = \frac{1}{N}\sum_{k=1}^{N} F(q^i, z^k)$ to represent agent i's average expected reward per period after one "cycle" of ranking rotation.

We also introduce a few constants to facilitate our analysis. Recall that $\Delta_F = F(1,1) - F(0,0)$ is a strictly positive constant that only depends on the problem input F. Further, let $\delta_F := \min\big\{i \neq j : \big|\sum_{\ell=1}^{N} F(q^i, z^\ell) - \sum_{\ell=1}^{N} F(q^j, z^\ell)\big|\big\}$ be a positive separation parameter independent of T. Finally, recall that η^2 is the variance proxy for the random shocks $\{\epsilon_t^i\}$.

Intuitively, during the experimentation phase, if inflation is "light," then the reward will "overwhelm" inflation. As a result, the total observed signal $\{\widetilde{D}^i\}$ can still be indicative regarding the

agents' intrinsic qualities. The following auxiliary result formalizes this intuition by quantifying the concentration properties of $\{\widetilde{D}^i\}$ after carefully accounting for the inflation amounts and random shocks.

LEMMA EC.1. **(experimentation phase of ETC)** *There exists a constant $M>0$ such that under the learner's ETC and agents' strategy profile $\mathbf{\Pi}\in\mathbf{\Pi}^M$, the ranking in the commitment phase $\overline{\sigma}$ satisfies*

$$\mathbb{P}(\overline{\sigma}=\boldsymbol{e})\geq 1-2N\exp\Big(-\frac{\delta_F^2}{32\eta^2 N}\tau_0\Big).$$

Proof of Lemma EC.1. Pick $M:=\frac{\delta_F}{4N}$. To prove the claim, we divide the proof into the following steps.

Step 1. Show that $\mathbb{P}(\overline{\sigma}=\boldsymbol{e})\geq\mathbb{P}\Big[\Big|\frac{\widetilde{D}^i}{\tau_0}-N\mu^i\Big|<\frac{\delta_F}{2},\ \forall i\in[N]\Big]$. Recall that the ranking in the commitment phase $\overline{\sigma}$ is obtained by simply sorting $\{\widetilde{D}^i\}$. Consider the event $\Big\{\Big|\frac{\widetilde{D}^i}{\tau_0}-N\mu^i\Big|<\frac{\delta_F}{2},\ \forall i\in[N]\Big\}$. Under this event, for every $i>j$,

$$\frac{\widetilde{D}^i}{\tau_0}-\frac{\widetilde{D}^j}{\tau_0}=\left[\frac{\widetilde{D}^i}{\tau_0}-N\mu^i\right]+\left[N\mu^i-N\mu^j\right]+\left[N\mu^j-\frac{\widetilde{D}^j}{\tau_0}\right]>-\delta_F+\left[N\mu^i-N\mu^j\right]\geq 0.$$

Therefore, $\overline{\sigma}=\boldsymbol{e}$, the identity ranking. In other words, the event $\Big\{\Big|\frac{\widetilde{D}^i}{\tau_0}-N\mu^i\Big|<\frac{\delta_F}{2},\ \forall i\in[N]\Big\}$ implies the event $\{\overline{\sigma}=\boldsymbol{e}\}$.

Step 2. Control $\mathbb{P}\big\{|\frac{\widetilde{D}^i}{\tau_0}-N\mu^i|\geq\frac{\delta_F}{2}\big\}$ for every fixed $i\in[N]$. We apply the generalized Hoeffding's inequality to finish this step. More specifically, for every $i\in[N]$,

$$\begin{aligned}
\mathbb{P}\Big[\Big|\frac{\widetilde{D}^i}{\tau_0}-N\mu^i\Big|\geq\frac{1}{2}\delta_F\Big] &\overset{(a)}{=}\mathbb{P}\Big[\Big|\sum_{t=1}^{\tau}(b_t^i+\epsilon_t^i)\Big|\geq\frac{1}{2}\delta_F\tau_0\Big]\\
&\leq\mathbb{P}\Big[\Big|\sum_{t=1}^{\tau}\epsilon_t^i\Big|\geq\frac{1}{2}\delta_F\tau_0-\sum_{t=1}^{\tau}b_t^i\Big]\\
&\overset{(b)}{\leq}\mathbb{P}\Big[\Big|\sum_{t=1}^{\tau}\epsilon_t^i\Big|\geq\frac{1}{2}\delta_F\tau_0-M\tau\Big]\\
&\overset{(c)}{=}\mathbb{P}\Big[\Big|\sum_{t=1}^{\tau}\epsilon_t^i\Big|\geq\frac{1}{2}\delta_F\tau_0-\frac{1}{4}\delta_F\tau_0\Big]\\
&=\mathbb{P}\Big[\Big|\sum_{t=1}^{\tau}\epsilon_t^i\Big|\geq\frac{1}{4}\delta_F\tau_0\Big]\\
&\overset{(d)}{\leq}2\exp\Big(-\frac{1}{2N\tau_0\eta^2}\frac{\delta_F^2\tau_0^2}{16}\Big)\\
&=2\exp\Big(-\frac{\delta_F^2}{32\eta^2N}\tau_0\Big)
\end{aligned}\tag{EC.2}$$

In (a), we invoke the definition of $\widetilde{D}^i$ in (EC.1) and then multiply both sides of the inequality by τ_0; (b) follows from the fact that $b_t^i\leq M$ for all $t\in[T]$ and $i\in[N]$; (c) is because $M=\delta_F/(4N)$ and $\tau=\tau_0 N$; in (d), the inequality follows directly from the generalized Hoeffding's inequality for the i.i.d sub-Gaussian noise ϵ_t^i.

Step 3. Conclude the claim through the union bound. From the observation in (EC.2), we can conclude that

$$\mathbb{P}(\overline{\sigma}=\boldsymbol{e})\geq\mathbb{P}\Big[\Big|\frac{\widetilde{D}^i}{\tau_0}-N\mu^i\Big|<\frac{\delta_F}{2},\ \forall i\in[N]\Big]\ =\ 1-\mathbb{P}\Big[\Big|\frac{\widetilde{D}^i}{\tau_0}-N\mu^i\Big|\geq\frac{\delta_F}{2},\ \text{ for some } i\in[N]\Big]$$

$$\geq 1-\sum_{i=1}^{N}\mathbb{P}\Big[\Big|\frac{\widetilde{D}^i}{\tau_0}-N\mu^i\Big|\geq\frac{\delta_F}{2}\Big]\ \geq\ 1-2N\exp\Big(-\frac{\delta_F^2}{32\eta^2N}\tau_0\Big).$$

That finishes the proof. ■

Proof of Proposition 3. Recall that the learner's ETC policy has two phases: the experimentation phase of the first τ periods and the commitment phase of the last $T-\tau$ periods. Hence, in the experimentation phase, we have

$$\begin{aligned}\mathcal{R}^E(T):=&\tau\sum\nolimits_{i=1}^{N}F(q^i,z^i)-\mathbb{E}\left[\sum\nolimits_{t=1}^{\tau}\sum\nolimits_{i=1}^{N}R_t(q^i,\sigma_t(i))\right]\\ =&\sum\nolimits_{t=1}^{\tau}\sum\nolimits_{i=1}^{N}[F(q^i,z^i)-F(q^i,\sigma_t(i))]\ \leq\ \Delta_F\tau N=\Delta_F\tau_0N^2.\end{aligned}$$

In the commitment phase, the ranking is fixed to be $\overline{\sigma}$ for the last $T-\tau$ periods. Therefore,

$$\begin{aligned}\mathcal{R}^C(T):=&(T-\tau)\sum\nolimits_{i=1}^{N}F(q^i,z^i)-\mathbb{E}\left[\sum\nolimits_{t=\tau+1}^{T}\sum\nolimits_{i=1}^{N}R_t(q^i,\sigma_t(i))\right]\\ \leq&(T-\tau)\sum\nolimits_{i=1}^{N}F(q^i,z^i)-\mathbb{E}\left[\sum\nolimits_{t=\tau+1}^{T}\sum\nolimits_{i=1}^{N}R_t(q^i,\sigma_t(i))\mid\overline{\sigma}=\boldsymbol{e}\right]\mathbb{P}(\overline{\sigma}=\boldsymbol{e})\\ \leq&(T-\tau)\sum\nolimits_{i=1}^{N}F(q^i,z^i)-\mathbb{E}\left[\sum\nolimits_{t=\tau+1}^{T}\sum\nolimits_{i=1}^{N}R_t(q^i,z^i)\right]\mathbb{P}(\overline{\sigma}=\boldsymbol{e})\\ \overset{(a)}{=}&(T-\tau)\sum\nolimits_{i=1}^{N}F(q^i,z^i)-\mathbb{E}\left[\sum\nolimits_{t=\tau+1}^{T}\sum\nolimits_{i=1}^{N}F(q^i,z^i)+\epsilon_t^i\right]\mathbb{P}(\overline{\sigma}=\boldsymbol{e})\\ =&(T-\tau)\sum\nolimits_{i=1}^{N}F(q^i,z^i)-\mathbb{P}(\overline{\sigma}=\boldsymbol{e})(T-\tau)\sum\nolimits_{i=1}^{N}F(q^i,z^i)\\ \overset{(b)}{\leq}&2\Delta_FN^2(T-\tau)\exp\Big(-\frac{\delta_F^2}{32\eta^2N}\tau_0\Big)\ \leq\ 2\Delta_FN^2T\exp\Big(-\frac{\delta_F^2}{32\eta^2N}\tau_0\Big),\end{aligned}$$

where (a) is because $R_t(q^i,z^i)=F(q^i,z^i)+\epsilon_t^i$, and (b) is due to Lemma EC.1. Combining both pieces, it holds that

$$\mathcal{R}_T=\mathcal{R}^E(T)+\mathcal{R}^C(T)\leq\Delta_F\tau_0N^2+2\Delta_FN^2T\exp\Big(-\frac{\delta_F^2}{32\eta^2N}\tau_0\Big).$$

Now, one can pick $\tau_0=C_0\log T$ as long as $C_0>\frac{32\eta^2}{\delta_F^2}N$, say, $C_0=\frac{64\eta^2}{\delta_F^2}N$. In that case,

$$\begin{aligned}\mathcal{R}_T=\mathcal{R}^E(T)+\mathcal{R}^C(T)\leq&\Delta_F\cdot\tfrac{64\eta^2}{\delta_F^2}N\log T\cdot N^2+2\Delta_FN^2T\exp\Big(-2\log T\Big)\\ =&\tfrac{64\eta^2\Delta_F}{\delta_F^2}N^3\log T+2\Delta_FN^2T^{-1}.\end{aligned}$$

It is thus clear that $\mathcal{R}_T=\mathcal{O}(\log T)$. ■

REMARK EC.1. We can make the choice of τ_0 agnostic of the problem inputs, such as the separation parameter δ_F. To achieve this, pick an arbitrary coercive function $f(x):(0,+\infty)\to(1,+\infty)$ so that $f(x)\uparrow\infty$ as $x\uparrow\infty$, where the divergence rate can be arbitrarily slow. It is straightforward to verify that as long as one picks $\tau_0=f(T)\log T$, the same proof applies and we have $\mathcal{R}_T=\mathcal{O}(f(T)\log T)$.

REMARK EC.2. One could also characterize the order of the regret $\mathcal{R}_T$ *jointly* in N and T. In fact, suppose the reward function F is smooth and $F_2(q,z)=\frac{\partial}{\partial z}F(q,z)$ is strictly positive. Then $\delta_F^*:=\min_{t\in[0,1]}F_2(t,t)$ is a strictly positive constant independent of both N and T. In addition, $\delta_F\geq\sum_{\ell=1}^{N}\frac{\delta_F^*}{N-1}\geq\delta_F^*>0$. Hence $\mathcal{R}_T\leq\frac{64\eta^2\Delta_F}{(\delta_F^*)^2}N^3\log T+2\Delta_FN^2T^{-1}=\mathcal{O}_{N,T}(N^3\log T)$.

B.2. LSI as Dominant Strategy for Budget Spending (Proposition 4) Proof of Proposition 4. Fix an arbitrary agent $i \in [N]$, time period $t \in [T]$, history $\boldsymbol{h}_{t-1} = \{\boldsymbol{B}, \sigma_1, \boldsymbol{D}_1, \boldsymbol{b}_1, \ldots, \sigma_{t-1}, \boldsymbol{D}_{t-1}, \boldsymbol{b}_{t-1}\}$, and agents' continuation strategy profiles $\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})$. Agent i's continuation value can be written as

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[\sum_{\ell=t}^{T} R_\ell(q^i, \sigma_\ell(i)) \mid \boldsymbol{h}_{t-1}\right] = \mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[\sum_{\ell=t}^{T} F(q^i, \sigma_\ell(i)) + \epsilon_\ell^i \mid \boldsymbol{h}_{t-1}\right] \\
&= \mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[\sum_{\ell=t}^{T} F(q^i, \sigma_\ell(i)) \mid \boldsymbol{h}_{t-1}\right] \\
&= \begin{cases} \sum_{\ell=t}^{\tau} F(q^i, \sigma_\ell(i)) + (T-\tau)\mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[F(q^i, \overline{\sigma}(i)) \mid \boldsymbol{h}_{t-1}\right] & \text{if } t \le \tau \\ (T-t+1)F(q^i, \overline{\sigma}(i)) & \text{if } t \ge \tau+1 \end{cases}.
\end{aligned}$$

In the derivations above, we have used the form of $\{R_t\}$ and the definition of the ETC policy; see (1) and Algorithm 1. In particular, if $t \ge \tau+1$, the observed signals during the experimentation phase, $\{\widetilde{D}^i\}$, purely depend on the history $\boldsymbol{h}_{t-1}$; see (EC.1). Therefore, the ranking during the commitment phase, $\overline{\sigma}$, is deterministic given $\boldsymbol{h}_{t-1}$. That makes agent i's continuation value independent of $\Pi^i(h^i_{t-1})$. Therefore, agent i is indifferent between *any* strategy and the proposition is trivially true.

Next, we will assume that $t \le \tau$. Note that $\sum_{\ell=t}^{\tau} F(q^i, \sigma_\ell(i))$ is also independent of $\Pi^i(h^i_{t-1})$ since the rankings $\{\sigma_t : t = 1, \ldots, \tau\}$ are determined through a fixed schedule. Therefore, it suffices to investigate $\mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[F(q^i, \overline{\sigma}(i)) \mid \boldsymbol{h}_{t-1}\right]$. At the same time, notice that

$$\begin{aligned}
\overline{\sigma}(i) &= \tfrac{1}{N-1}\sum_{j\neq i} \mathbb{I}\{\widetilde{D}^i > \widetilde{D}^j\} \\
&= \tfrac{1}{N-1}\sum_{j\neq i} \mathbb{I}\left\{\tau_0 \sum_{k=1}^{N} F(q^i, z^k) + \sum_{\ell=1}^{\tau}(b^i_\ell + \epsilon^i_\ell) > \tau_0 \sum_{k=1}^{N} F(q^j, z^k) + \sum_{\ell=1}^{\tau}(b^j_\ell + \epsilon^j_\ell)\right\} \\
&= \tfrac{1}{N-1}\sum_{j\neq i} \mathbb{I}\left\{\underbrace{\tau_0 \sum_{k=1}^{N}[F(q^i, z^k) - F(q^j, z^k)] + \sum_{\ell=1}^{t-1}(b^i_\ell + \epsilon^i_\ell - b^j_\ell - \epsilon^j_\ell)}_{A} + \sum_{\ell=t}^{\tau}(b^i_\ell + \epsilon^i_\ell) > \underbrace{\sum_{\ell=t}^{\tau}(b^j_\ell + \epsilon^j_\ell)}_{B}\right\}.
\end{aligned}$$

In the expression above, Term A is deterministic conditional on $\boldsymbol{h}_{t-1}$, and Term B is independent of agent i's continuation strategy $\Pi^i(h^i_{t-1})$. Therefore, conditional on $\boldsymbol{h}_{t-1}$, $\overline{\sigma}(i)$ increases in $\sum_{\ell=t}^{\tau} b^i_\ell$ in the stochastic dominance sense. The latter quantity, which represents the total inflation amount up to period τ, attains its maximum as long as $\sum_{\ell=t}^{\tau} b^i_\ell = B^i - \sum_{\ell=1}^{t-1} b^i_\ell$, i.e., the budget is used up during the experimentation phase. That is particularly the case under LSI. Since $F(q, z)$ is strictly increasing in z (see Assumption (i)),

$$\mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\Big[F(q^i, \overline{\sigma}(i)) \mid \boldsymbol{h}_{t-1}\Big] \le \mathbb{E}_{\Pi^i_{LSI}(h^i_{t-1}), \boldsymbol{\Pi}^{-i}(\boldsymbol{h}^{-i}_{t-1})}\Big[F(q^i, \overline{\sigma}(i)) \mid \boldsymbol{h}_{t-1}\Big].$$

That finishes the proof. ■

C. The General Convergence Theory (Analysis of $\overline{\Gamma}_{N,\infty}$)

C.1. Further Details of Game Terminology For completeness and ease of reference, let us follow the standard convention and formally define the pure- and mixed-strategy Nash equilibrium for the game $\overline{\Gamma}_{N,\infty}$.

DEFINITION EC.1. **(Nash equilibrium for $\overline{\Gamma}_{N,\infty}$)** An action profile $\boldsymbol{b}_* = \{b_*^i\}$ is a (pure-strategy) Nash equilibrium of $\overline{\Gamma}_{N,\infty}$ if $b_*^i \in \arg\max_{b\geq 0} \overline{v}^i(b, \boldsymbol{b}_*^{-i})$ for every $i \in [N]$.

With a slight abuse of notation, for any mixed strategy profile $\boldsymbol{G} = \{G^i\}$ and agent i's strategy G^i, we may write agent i's expected utility as

$$\overline{v}^i(G^i, \boldsymbol{G}^{-i}) = \int_{\mathbb{R}_+} \int_{\mathbb{R}_+^{N-1}} \overline{v}^i(x, \boldsymbol{y})\ d\boldsymbol{G}^{-i}(\boldsymbol{y})\, dG^i(x). \tag{EC.3}$$

DEFINITION EC.2. **(mixed strategy equilibrium for $\overline{\Gamma}_{N,\infty}$)** A mixed strategy profile $\boldsymbol{G}_* = \{G_*^i\}$ is a (mixed strategy) Nash equilibrium of $\overline{\Gamma}_{N,\infty}$ if $\overline{v}^i(G_*^i, \boldsymbol{G}_*^{-i}) \geq \overline{v}^i(b^i, \boldsymbol{G}_*^{-i})$ for all $i \in [N]$ and $b^i \geq 0$.

We also find the following characterization of mixed strategy NE helpful. For any $i \in [N]$, we denote by $S(G^i) \subset \mathbb{R}_+$ the support of a mixed strategy G^i such that $\mathbb{P}_G(b^i \in S(G^i)) = 1$. The equivalent conditions for mixed strategy NE are stated below.

FACT 1. **(Proposition 8.D.1 in Mas-Colell et al. 1995)** A strategy profile $\{G^i\}_{i=1}^N$ is a mixed strategy equilibrium of $\overline{\Gamma}_{N,\infty}$ if and only if for any $i \in [N]$,

1. (Global optimality) $\overline{v}^i(b_*^i, \boldsymbol{G}^{-i}) \geq \overline{v}^i(b^i, \boldsymbol{G}^{-i})$ for all $b_*^i \in S(G^i)$ and $b^i \notin S(G^i)$;
2. (Indifference) $\overline{v}^i(b_*^i, \boldsymbol{G}^{-i}) = \overline{v}^i(b_{**}^i, \boldsymbol{G}^{-i})$ for all $b_*^i, b_{**}^i \in S(G^i)$.

C.2. On Robustness for Tie-breaking Rule Strictly speaking, there can be some ambiguity regarding the tie-breaking rule in defining the utility functions for the game $\overline{\Gamma}_{N,\infty}$. In this subsection, we provide a separate treatment of this ambiguity and show that under a large class of tie-breaking rules, all these game variants are actually equivalent in terms of the equilibrium outcomes.

Let us introduce a set of matrices $\mathcal{M} = \{M \in [0,1]^{N\times N} \mid M_{ii} = 0,\ 0 \leq M_{ij} + M_{ji} \leq 1,\ \text{for } i, j \in [N],\ i \neq j\}$. Here, every matrix $M \in \mathcal{M}$ represents a potentially different tie-breaking rule. Specifically, given the action profile $\boldsymbol{b}$, the "inflation-intensity based" position of agent $i \in [N]$ under the tie-breaking rule M is defined as

$$H_M^N(b^i \,|\, \boldsymbol{b}^{-i}) \ = \ \frac{1}{N-1} \sum_{j\neq i} \left(\mathbb{I}\{b^j < b^i\} + M_{ij} \times \mathbb{I}\{b^j = b^i\}\right). \tag{EC.4}$$

One could loosely interpret M_{ij} as the probability of agent i beating agent j if their inflation intensities are equal. The tie-breaking rule family above can include a wide range of tie-breaking conventions. Let us list a few special examples below.

- Suppose $M \equiv 0$, the all-zero matrix. Then we recover the original definition of $H^N(b^i \,|\, \boldsymbol{b}^{-i})$ in (5). This is the least preferable tie-breaking rule for all agents simultaneously.
- Suppose M is a constant matrix equal to $1/2$ (except for diagonal entries). Then M corresponds to the uniform random tie-breaking rule, and $H_M^N(b^i \,|\, \boldsymbol{b}^{-i})$ is the resulting (expected) ranking position.
- Suppose M is a permutation matrix, i.e., for some permutation $\mathfrak{p} : [N] \to [N]$, it holds that $M_{ij} = \mathbb{I}\{\mathfrak{p}(i) > \mathfrak{p}(j)\}$. Then the agents will be ranked based on the lexicographic order of $(\boldsymbol{b}, \mathfrak{p})$. In other words, M refers to the tie-breaking rule according to a pre-fixed order of clearance, regardless of the agents' qualities, observed signal, etc.

Every choice of the matrix M results in a slightly different variation of the original game $\overline{\Gamma}_{N,\infty}$, where the difference lies in the ranking position definition and thus utility function. We will refer to the game governed by tie-breaking M as the *M-limiting game*. Given a pure-strategy profile $\boldsymbol{b}$, agent i's utility is given by $\overline{v}_M^i(b^i, \boldsymbol{b}^{-i}) := F\big(q^i, H_M^N(b^i \,|\, \boldsymbol{b}^{-i})\big) - cb^i$. Given any mixed strategy profile $\boldsymbol{G} = (G^1, \ldots, G^N)$, agent i's utility is given by

$$\overline{v}_M^i(G^i, \boldsymbol{G}^{-i}) = \int_{\mathbb{R}_+} \int_{\mathbb{R}_+^{N-1}} \overline{v}_M^i(x, \boldsymbol{y})\, d\boldsymbol{G}^{-i}(\boldsymbol{y})\, dG^i(x).$$

It can be verified that if we take $M \equiv 0$, we recover the original utility defined in (6) and (EC.3). In the context of the (N, ∞) game, the only source of randomness is the agents' mixed strategies. Therefore, for convenience of notation, we use $\mathbb{P}_{\boldsymbol{G}}(\cdot)$ to represent the probability distribution over $\boldsymbol{b}$ under the budget strategy profile $\boldsymbol{G}$. Similarly, we use $\mathbb{P}_{G^i}(\cdot)$ to represent the (marginal) probability distribution over b^i under his mixed strategy G^i.

Our main results regarding the robustness of the tie-breaking rule in formulating the (N, ∞) game are summarized below.

PROPOSITION EC.2. **(Tie-breaking rule)** *If $\boldsymbol{G}$ is a mixed strategy Nash equilibrium for the M-limiting game for some $M \in \mathcal{M}$. Then the following facts hold.*

(1) $\mathbb{P}_{\boldsymbol{G}}(\{\boldsymbol{b} : \ b^i = b^j \text{ for some } i \neq j\}) = 0$. In other words, ties occur with zero probability in equilibrium.

(2) $\boldsymbol{G}$ is Nash equilibrium in the M'-limiting game for all $M' \in \mathcal{M}$. In other words, Nash equilibria are robust to the choice of tie-breaking rules.

Proof of Proposition EC.2. Pick an arbitrary mixed strategy Nash equilibrium $\boldsymbol{G}$ in the M-limiting game. For every agent i, let $E^i := \{b : \mathbb{P}_{G^i}\{b\} > 0\}$ be the singular set for the agent i, i.e., the set of positive point measures under the probability distribution G^i. Notice that E^i could be empty for some agents. We prove each of the claims below.

Verifying Claim (1). It suffices to show that all the singular sets are disjoint, i.e., $E^i \cap E^j = \emptyset$ for every $i \neq j$. Suppose towards a contradiction that for some distinct $i, j \in [N]$, there exists an inflation intensity $b^i_* \in E^i \cap E^j$. Since $0 \leq M_{ij} + M_{ji} \leq 1$, we may suppose that $M_{ij} < 1$ without loss of generality. Roughly speaking, it means that agent i does not "perfectly beat" agent j if their inflation intensities are tied in equilibrium; otherwise, we must have $M_{ji} = 0$. In that case, we can switch i and j and repeat the argument below.

Since $\boldsymbol{G}$ is a mixed strategy Nash equilibrium and b^i_* is supported by G^i, b^i_* must be the best-response pure strategy for the agent i. That is, $b^i_* \in \arg\max_{b \geq 0} \overline{v}^i_M(b, \boldsymbol{G}^{-i})$. At the same time, we will show that agent i can be strictly better off by choosing a slightly higher inflation intensity, thus leading to the contradiction. Recall that $H^N_M(b^i \,|\, \boldsymbol{b}^{-i}) = \frac{1}{N-1} \sum_{j \neq i} (\mathbb{I}\{b^j < b^i\} + M_{ij} \times \mathbb{I}\{b^j = b^i\})$ for any $\boldsymbol{b} \in \mathbb{R}^N_+$, where $0 \leq M_{ij} \leq 1$. Therefore, agent i's position increases in his inflation intensity. Particularly, for every $\varepsilon > 0$, $H^N_M(b^i_* + \varepsilon \,|\, \boldsymbol{b}^{-i}) \geq H^N_M(b^i_* \,|\, \boldsymbol{b}^{-i})$. Due to the strict monotonicity of F (see Assumption (i)), it further holds that

$$F\big(q^i, H^N_M(b^i_* + \varepsilon \,|\, \boldsymbol{b}^{-i})\big) - F\big(q^i, H^N_M(b^i_* \,|\, \boldsymbol{b}^{-i})\big) \ \geq \ 0.$$

Suppose in addition that $\boldsymbol{b}^{-i}$ is such that $b^j = b^i_*$, then

$$H^N_M(b^i_* + \varepsilon \,|\, b^i_*, \boldsymbol{b}^{-(i,j)}) - H^N_M(b^i_* \,|\, b^i_*, \boldsymbol{b}^{-(i,j)}) \geq \tfrac{1 - M_{ij}}{N-1} > 0$$

In other words, agent i's position is strictly "boosted" no matter how small ε is. As a result,

$$F\big(q^i, H^N_M(b^i_* + \varepsilon \,|\, \boldsymbol{b}^{-i})\big) - F\big(q^i, H^N_M(b^i_* \,|\, \boldsymbol{b}^{-i})\big) \ \geq \min_{i \in [N], z \in \Theta_N} [F(q^i, z + \tfrac{1 - M_{ij}}{N-1}) - F(q^i, z)] =: c_0 > 0.$$

In other words, agent i's expected reward is also strictly "boosted." Therefore, for sufficiently small $\varepsilon > 0$,

$$\begin{aligned}
& \overline{v}^i_M(b^i_* + \varepsilon, \boldsymbol{G}^{-i}) - \overline{v}^i_M(b^i_*, \boldsymbol{G}^{-i}) \\
= & \int_{\mathbb{R}^{N-1}_+} [\overline{v}^i_M(b^i_* + \varepsilon, \boldsymbol{b}^{-i}) - \overline{v}^i_M(b^i_*, \boldsymbol{b}^{-i})] \ d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i}) \\
= & \int_{\mathbb{R}^{N-1}_+} [F\big(q^i, H^N_M(b^i_* + \varepsilon \,|\, \boldsymbol{b}^{-i})\big) - F\big(q^i, H^N_M(b^i_* \,|\, \boldsymbol{b}^{-i})\big)] \ d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i}) - c\varepsilon \\
\overset{(a)}{\geq} & \int_{\{\boldsymbol{b}^{-i} : b^j = b^i_*\}} [F\big(q^i, H^N_M(b^i_* + \varepsilon \,|\, \boldsymbol{b}^{-i})\big) - F\big(q^i, H^N_M(b^i_* \,|\, \boldsymbol{b}^{-i})\big)] \ d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i}) - c\varepsilon \\
\overset{(b)}{\geq} & \int_{\{\boldsymbol{b}^{-i} : b^j = b^i_*\}} c_0 \ d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i}) - c\varepsilon = c_0 \mathbb{P}_{G^j}(\{b^i_*\}) - c\varepsilon > 0.
\end{aligned}$$

In the derivation above, part (a) is due to the nonnegativity of the integrand; part (b) is due to the lower bound of the integrand when $b^j = b^i_*$. We have thus proved Claim (1).

Verifying Claim (2). Pick an arbitrary $M' \in \mathcal{M}$, agent $i \in [N]$ and $b^i \geq 0$. We will verify that

$$\overline{v}^i_{M'}(b^i, \boldsymbol{G}^{-i}) \;\leq\; \overline{v}^i_{M'}(G^i, \boldsymbol{G}^{-i}),$$

thus verifying that $\boldsymbol{G}$ is also a Nash equilibrium in the M'-limiting game.

Suppose $b^i \notin E^{-i} := \cup_{j\neq i} E^j$. Then with probability one, the inflation intensities of agent i's opponents will not tie with b^i. Therefore, $H^N_{M'}(b^i \,|\, \boldsymbol{G}^{-i}) = H^N_M(b^i \,|\, \boldsymbol{G}^{-i})$, and thus

$$\overline{v}^i_{M'}(b^i, \boldsymbol{G}^{-i}) = \overline{v}^i_M(b^i, \boldsymbol{G}^{-i}) \overset{(a)}{\leq} \overline{v}^i_M(G^i, \boldsymbol{G}^{-i}) \overset{(b)}{=} \overline{v}^i_{M'}(G^i, \boldsymbol{G}^{-i}).$$

In the derivations above, (a) is because $\boldsymbol{G}$ is a Nash equilibrium in the M-limiting game; (b) is because of Claim (1), which implies that the probability of inflation ties is zero.

Suppose $b^i \in E^{-i}$. Because of Claim (1), there is a unique agent j whose inflation intensity may tie with b^i with positive probability, i.e., $b^i \in E^j$. We further consider two cases. If $M'_{ij} \leq M_{ij}$, i.e., M' is a less preferable tie-breaking rule than M for agent i, it holds that

$$H^N_{M'}(b^i \,|\, \boldsymbol{G}^{-i}) - H^N_M(b^i \,|\, \boldsymbol{G}^{-i}) \;=\; 0 \times \mathbb{P}_{G^j}(\{b^i\}^c) + \tfrac{1}{N-1}\Big(M'_{ij} - M_{ij}\Big) \times \mathbb{P}_{G^j}\{b^i\} \leq 0.$$

As a result, $\overline{v}^i_{M'}(b^i, \boldsymbol{G}^{-i}) \leq \overline{v}^i_M(b^i, \boldsymbol{G}^{-i}) \leq \overline{v}^i_M(G^i, \boldsymbol{G}^{-i}) = \overline{v}^i_{M'}(G^i, \boldsymbol{G}^{-i})$. The other case where $M'_{ij} > M_{ij}$ is slightly more involved. We make a chain of derivations below. For every $\varepsilon > 0$,

$$\begin{aligned}
&\overline{v}^i_{M'}(b^i, \boldsymbol{G}^{-i}) \\
&\overset{(a)}{=} \overline{v}^i_{M'}(b^i, G^j, \boldsymbol{G}^{-(i,j)}) \\
&= \int_{\mathbb{R}_+\setminus\{b^i\}} \overline{v}^i_{M'}(b^i, x, \boldsymbol{G}^{-(i,j)})\; dG^j(x) + \mathbb{P}_{G^j}\{b^i\}\, \overline{v}^i_{M'}(b^i, b^i, \boldsymbol{G}^{-(i,j)}) \\
&\overset{(b)}{=} \int_{\mathbb{R}_+\setminus\{b^i\}} \overline{v}^i_M(b^i, x, \boldsymbol{G}^{-(i,j)})\; dG^j(x) + \mathbb{P}_{G^j}\{b^i\}\, \overline{v}^i_M(b^i, b^i, \boldsymbol{G}^{-(i,j)}) + \mathbb{P}_{G^j}\{b^i\}\Big[\overline{v}^i_{M'}(b^i, b^i, \boldsymbol{G}^{-(i,j)}) - \overline{v}^i_M(b^i, b^i, \boldsymbol{G}^{-(i,j)})\Big] \\
&\overset{(c)}{\leq} \Big[\int_{\mathbb{R}_+\setminus\{b\}} \overline{v}^i_M(b^i, x, \boldsymbol{G}^{-(i,j)})\; dG^j(x) + \mathbb{P}_{G^j}\{b^i\}\overline{v}^i_M(b^i, b^i, \boldsymbol{G}^{-(i,j)})\Big] \\
&\qquad + \mathbb{P}_{G^j}\{b^i\}\Big[\overline{v}^i_M(b^i+\varepsilon, b^i, \boldsymbol{G}^{-(i,j)}) + c\varepsilon - \overline{v}^i_M(b^i, b^i, \boldsymbol{G}^{-(i,j)})\Big] \\
&\overset{(d)}{=} \int_{\mathbb{R}_+\setminus\{b\}} \overline{v}^i_M(b^i, x, \boldsymbol{G}^{-(i,j)})\; dG^j(x) + \mathbb{P}_{G^j}\{b^i\}\Big[\overline{v}^i_M(b^i+\varepsilon, b^i, \boldsymbol{G}^{-(i,j)}) + c\varepsilon\Big] \\
&\overset{(e)}{\leq} \int_{\mathbb{R}_+\setminus\{b^i\}} \Big[\overline{v}^i_M(b^i+\varepsilon, x, \boldsymbol{G}^{-(i,j)}) + c\varepsilon\Big]\; dG^j(x) + \mathbb{P}_{G^j}\{b^i\}\Big[\overline{v}^i_M(b^i+\varepsilon, b^i, \boldsymbol{G}^{-(i,j)}) + c\varepsilon\Big] \\
&= \overline{v}^i_M(b^i+\varepsilon, G^j, \boldsymbol{G}^{-(i,j)}) + c\varepsilon \\
&\overset{(f)}{\leq} \overline{v}^i_M(G^i, G^j, \boldsymbol{G}^{-(i,j)}) + c\varepsilon \\
&= \overline{v}^i_M(G^i, \boldsymbol{G}^{-i}) + c\varepsilon \;=\; \overline{v}^i_{M'}(G^i, \boldsymbol{G}^{-i}) + c\varepsilon.
\end{aligned}$$

Let us explain more details of the derivations above:

- Pt. (a). We have adopted a slight abuse of notation by using $\overline{v}^i_{M'}(b^i, G^j, \boldsymbol{G}^{-(i,j)})$ to represent agent j's strategy G^j explicitly.
- Pt. (b). We have made two changes. First, we replace M' with M in the integrand. This is legitimate because only agent j's inflation can tie with b^i with positive probability. Second, we add and subtract the same term $\mathbb{P}_{G^j}\{b\}\overline{v}^i_M(b^i, b^i, \boldsymbol{G}^{-(i,j)})$.

Pt. (c). We replaced the term $\overline{v}^i_{M'}(b^i, b^i, \boldsymbol{G}^{-(i,j)})$ with the terms $\overline{v}^i_M(b^i+\varepsilon, b^i, \boldsymbol{G}^{-(i,j)}) + c\varepsilon$. We know that the former is smaller since by increasing the inflation from the tied level (i.e., b^i) to a winning level (i.e., $b^i+\varepsilon$), agent i's position will always be weakly better regardless of the transition of tie-breaking rule.

Pt. (d). We have canceled out the term $\mathbb{P}_{G^j}\{b^i\}\,\overline{v}^i_M(b^i, b^i, \boldsymbol{G}^{-(i,j)})$.

Pt. (e). We use the fact that $\overline{v}^i_M(b^i, b^i, \boldsymbol{G}^{-(i,j)}) \le \overline{v}^i_M(b^i+\varepsilon, b^i, \boldsymbol{G}^{-(i,j)}) + c\varepsilon$, which is similar to part (c) above.

Pt. (f). We use the fact that $\overline{v}^i_M(b^i+\varepsilon, G^j, \boldsymbol{G}^{-(i,j)}) \le \overline{v}^i_M(G^i, G^j, \boldsymbol{G}^{-(i,j)})$ since $\boldsymbol{G}$ is a Nash equilibrium.

The proof is finished because $\overline{v}^i_{M'}(b^i, \boldsymbol{G}^{-i}) \le \overline{v}^i_{M'}(G^i, \boldsymbol{G}^{-i}) + c\varepsilon$ for all $\varepsilon > 0$, which means that $\overline{v}^i_{M'}(b^i, \boldsymbol{G}^{-i}) \le \overline{v}^i_{M'}(G^i, \boldsymbol{G}^{-i})$. ■

For ease of future reference, let us also state a corollary here.

COROLLARY EC.1. *If $\boldsymbol{G} = (G^1, \dots, G^N)$ is an ε-equilibrium for $\overline{M}$-limiting game for some $\overline{M} \in \mathcal{M}$ and $\mathbb{P}_{\boldsymbol{G}}\{b^j = b^i\} = 0$ for any $i, j \in [N]$ with $i \neq j$. Then, $\boldsymbol{G}$ is an ε-equilibrium in M-limiting game for all $M \in \mathcal{M}$.*

Proof of Corollary EC.1. The proof can be repeated verbatim as the proof of part (2) of Proposition EC.2. ■

C.3. Convergence Results (Theorems 1 and 2) We split the proof of Theorem 1 into two parts. We will start from the first statement.

Proof of Part (i) of Theorem 1. For the learner's ETC policy, we set $\tau = o(\sqrt{T})$. In this proof, to highlight the dependence of v^i on the time horizon T in the budget competition game $\overline{\Gamma}_{N,T}$, we will denote v^i as v^i_T with a slight abuse of notation.

Step 1: Analyze $\overline{\sigma}(i)$ for every agent $i \in [N]$. Fix an arbitrary inflation intensity profile $\boldsymbol{b}$ throughout the step. Consider the pre-limit game $\overline{\Gamma}_{N,T}$ and suppose that all agents employ LSI with budgets $T\boldsymbol{b}$. For every two agents $i, j \in [N]$ such that $b^i > b^j$, let $\{\widetilde{D}^j > \widetilde{D}^i\}$ be the event that the learner observes more observed signal for agent j than agent i during the experimentation phase. It holds that if $\tau(\mu^i - \mu^j) + T(b^i - b^j) \ge 0$,

$$\begin{aligned}
\mathbb{P}(\widetilde{D}^j > \widetilde{D}^i) &\overset{(a)}{=} \mathbb{P}\Big\{\tau\mu^j + \textstyle\sum_{t=1}^{\tau}\epsilon^j_t + Tb^j > \tau\mu^i + \sum_{t=1}^{\tau}\epsilon^i_t + Tb^i\Big\} \\
&= \mathbb{P}\Big(\textstyle\sum_{t=1}^{\tau}(\epsilon^j_t - \epsilon^i_t) > \tau(\mu^i - \mu^j) + T(b^i - b^j)\Big) \\
&\overset{(b)}{\le} \exp\Big(-\tfrac{[\tau(\mu^i-\mu^j)+T(b^i-b^j)]^2}{4\tau\eta^2}\Big) \\
&= \exp\Big(-\tfrac{1}{4\tau\eta^2}|(\mu^i-\mu^j)\tau + (b^i-b^j)T|^2\Big).
\end{aligned} \tag{EC.5}$$

In the derivations above, part (a) uses the following fact: for every $l \in [N]$, (i) $\widetilde{D}^l = \tau\mu^l + \sum_{t=1}^{\tau}(b^l_t + \epsilon^l_t)$ from (EC.1); and (ii) LSI is the dominant strategy by Proposition 4, which further implies that $\sum_{t=1}^{\tau} b^l_t = Tb^l$. In part (b), the inequality follows from the generalized Hoeffding's inequality for sub-Gaussian noise $\{\epsilon^i_t\}$.

We now fix any $i \in [N]$ for the remaining proofs. Motivated by the analysis above, given the agent number N and the inflation profile $\boldsymbol{b}$, let us introduce (the minimum) $\overline{T}_i(\boldsymbol{b}) < +\infty$ so that as long as $T > \overline{T}_i(\boldsymbol{b})$, the concentration inequalities above are applicable for all $j \in \{j : b^j > b^i\}$; that is,

$$\tau(\mu^j - \mu^i) + T(b^j - b^i) \ge 0 \text{ for all } j \in \{j : b^j > b^i\}. \tag{EC.6}$$

Such $\overline{T}_i(\boldsymbol{b})$ exists because $\tau = o(\sqrt{T})$. Recall that $\overline{\sigma}(i) = \frac{1}{N-1}\sum_{j\in[N]\setminus\{i\}} \mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\}$ is agent i's ranking position based on $\{\widetilde{D}^i\}$. We will compare it with the following two quantities:

$$H^N(b^i|\boldsymbol{b}^{-i}) = \tfrac{1}{N-1}\sum_{j\in[N]\setminus\{i\}} \mathbb{I}\{b^j < b^i\} \quad \text{and} \quad H^N_{M^i}(b^i|\boldsymbol{b}^{-i}) = \tfrac{1}{N-1}\sum_{j\in[N]\setminus\{i\}} \mathbb{I}\{b^j \le b^i\}.$$

Those are both agent i's ranking position based on $\{b^i\}$ but with different tie-breaking rules. The former is based on (5) and his least preferred tie-breaking rule. The latter is based on (EC.4) where M^i is a tie-breaking matrix where $M^i_{im}=1$ for all $m\in[N]$. Thus, the latter is agent i's most favorable tie-breaking rule.

We next show that $\overline{\sigma}(i)\le H^N_{M^i}(b^i|\boldsymbol{b}^{-i})$ with high probability. In fact, it holds that

$$\begin{aligned} H^N_{M^i}(b^i|\boldsymbol{b}^{-i})<\overline{\sigma}(i) &\Rightarrow |\{j:b^j\le b^i\}|<|\{j:\widetilde{D}^j<\widetilde{D}^i\}| \\ &\Rightarrow \{j:\widetilde{D}^j<\widetilde{D}^i\}\not\subseteq\{j:b^j\le b^i\} \\ &\Rightarrow \text{there exists } j \text{ such that } \widetilde{D}^j<\widetilde{D}^i \text{ and } b^j>b^i. \end{aligned}$$

Therefore, for all $T>\overline{T}_i(\boldsymbol{b})$, due to the union bound,

$$\begin{aligned} \mathbb{P}\Big\{H^N_{M^i}(b^i|\boldsymbol{b}^{-i})<\overline{\sigma}(i)\Big\} &\le \textstyle\sum_{j:b^j>b^i}\mathbb{P}(\widetilde{D}^j<\widetilde{D}^i) \\ &= \textstyle\sum_{j:b^j>b^i}\exp\Big(-\frac{1}{4\tau\eta^2}|(\mu^j-\mu^i)\tau+(b^j-b^i)T|^2\Big) \\ &\le |\{j:b^j>b^i\}|\exp\Big(-\frac{1}{4\tau\eta^2}\min_{j:b^j>b^i}|(\mu^i-\mu^j)\tau+(b^i-b^j)T|^2\Big) \\ &\le N\exp\Big(-\frac{1}{4\tau\eta^2}\min_{j:b^j>b^i}|(\mu^i-\mu^j)\tau+(b^i-b^j)T|^2\Big) \end{aligned} \tag{EC.7}$$

Similarly, given N, we introduce (the minimum) $\widetilde{T}_i(\boldsymbol{b})<+\infty$ so that as long as $T>\widetilde{T}_i(\boldsymbol{b})$, the concentration inequality (EC.5) is applicable for all $j\in\{j:b^j>b^i\}$. That is, $\tau(\mu^j-\mu^i)+T(b^j-b^i)\ge 0$ for all $j\in\{j:b^j<b^i\}$. We will show that $\overline{\sigma}(i)\ge H^N(b^i|\boldsymbol{b}^{-i})$ with high probability. We observe that

$$H^N(b^i|\boldsymbol{b}^{-i})>\overline{\sigma}(i)\Rightarrow|\{j:b^j<b^i\}|>|\{j:\widetilde{D}^j\le\widetilde{D}^i\}|\Rightarrow\text{there exists } j \text{ such that } \widetilde{D}^j>\widetilde{D}^i \text{ and } b^j<b^i.$$

As a result, for all $T>\widetilde{T}_i(\boldsymbol{b})$,

$$\begin{aligned} &\mathbb{P}\Big\{H^N(b^i|\boldsymbol{b}^{-i})>\overline{\sigma}(i)\Big\} \\ &\le |\{j:b^j<b^i\}|\exp\Big(-\frac{1}{4\tau\eta^2}\min_{j:b^j<b^i}|(\mu^i-\mu^j)\tau+(b^i-b^j)T|^2\Big) \\ &\le N\exp\Big(-\frac{1}{4\tau\eta^2}\min_{j:b^j<b^i}|(\mu^i-\mu^j)\tau+(b^i-b^j)T|^2\Big). \end{aligned} \tag{EC.8}$$

In addition, when $\boldsymbol{b}$ has no ties, it holds that $H^N(b^i|\boldsymbol{b}^{-i})=H^N_{M^i}(b^i|\boldsymbol{b}^{-i})$. We can therefore invoke (EC.7) and (EC.8) and conclude that when $T>\max\{\overline{T}_i(\boldsymbol{b}),\widetilde{T}_i(\boldsymbol{b})\}$,

$$\begin{aligned} \mathbb{P}\Big\{H^N(b^i|\boldsymbol{b}^{-i})\ne\overline{\sigma}(i)\Big\} &= \mathbb{P}\Big\{H^N(b^i|\boldsymbol{b}^{-i})>\overline{\sigma}(i)\Big\}+\mathbb{P}\Big\{H^N(b^i|\boldsymbol{b}^{-i})<\overline{\sigma}(i)\Big\} \\ &= \mathbb{P}\Big\{H^N(b^i|\boldsymbol{b}^{-i})>\overline{\sigma}(i)\Big\}+\mathbb{P}\Big\{H^N_{M^i}(b^i|\boldsymbol{b}^{-i})<\overline{\sigma}(i)\Big\} \\ &\le N\exp\Big(-\frac{1}{4\tau\eta^2}\min_{j\in[N]\setminus\{i\}}|(\mu^i-\mu^j)\tau+(b^i-b^j)T|^2\Big). \end{aligned} \tag{EC.9}$$

Step 2: Bound $v^i_T(b^i,\boldsymbol{b}^{-i})$ for every $\boldsymbol{b}$. Fix any $\boldsymbol{b}\in\mathbb{R}^N_+$ for this step. Recall from (6) that $\overline{v}^i(b^i,\boldsymbol{b}^{-i})=F\big(q^i,H^N(b^i|\boldsymbol{b}^{-i})\big)-cb^i$. Similarly, $\overline{v}^i_{M^i}(b^i,\boldsymbol{b}^{-i})=F\big(q^i,H^N_{M^i}(b^i|\boldsymbol{b}^{-i})\big)-cb^i$. These are agent i's utilities based on agent i's least preferred and most preferred tie-breaking rules, respectively.

We are going to provide bounds of $v_T^i(b^i, \boldsymbol{b}^{-i})$ in terms of $\overline{v}^i(b^i, \boldsymbol{b}^{-i})$ and $\overline{v}^i_{M^i}(b^i, \boldsymbol{b}^{-i})$ separately. For $T > \overline{T}_i(\boldsymbol{b})$,

$$\begin{aligned}
& v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}^i_{M^i}(b^i, \boldsymbol{b}^{-i}) \\
=\ & \tfrac{1}{T}\Big\{ \textstyle\sum_{t=1}^{T} \mathbb{E}[R_t(q^i, \sigma_t(i))] - TF(q^i, H^N_{M^i}(b^i|\boldsymbol{b}^{-i}))\Big\} \\
=\ & \tfrac{1}{T}\Big\{ \textstyle\sum_{t=1}^{T} \mathbb{E}[F(q^i, \sigma_t(i))] - TF(q^i, H^N_{M^i}(b^i|\boldsymbol{b}^{-i}))\Big\} \\
\overset{(a)}{\le}\ & \tfrac{1}{T}\Big\{ \textstyle\sum_{t=1}^{\tau} \Big(F(q^i, \sigma_t(i)) - \tau F(q^i, H^N_{M^i}(b^i|\boldsymbol{b}^{-i}))\Big) + (T-\tau)\Big(\mathbb{E}[F(q^i, \overline{\sigma}(i))] - F(q^i, H^N_{M^i}(b^i|\boldsymbol{b}^{-i}))\Big)\Big\} \\
\overset{(b)}{\le}\ & \tfrac{\Delta_F}{T}\left(\tau + (T-\tau)\mathbb{P}\{\overline{\sigma}(i) > H^N_{M^i}(b^i|\boldsymbol{b}^{-i})\}\right). \\
\overset{(c)}{\le}\ & \tfrac{\Delta_F}{T}\Big[\tau + (T-\tau)N\exp\Big(-\tfrac{1}{4\tau\eta^2}\min_{j:b^j>b^i}|(\mu^i-\mu^j)\tau + (b^i-b^j)T|^2\Big)\Big].
\end{aligned} \tag{EC.10}$$

In the derivations above, (a) follows from the definition of ETC in Algorithm 1; (b) holds because $|F(q^i, z) - F(q^i, z')| \le \Delta_F = F(1,1) - F(0,0)$ for every ranking positions $z, z' \in [0,1]$; (c) follows from (EC.7).

Suppose further that $\boldsymbol{b}$ has no ties. Then (EC.9) implies that for $T > \max\{\overline{T}_i(\boldsymbol{b}), \widetilde{T}_i(\boldsymbol{b})\}$,

$$\begin{aligned}
& |v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}^i(b^i, \boldsymbol{b}^{-i})| \\
\le\ & \tfrac{1}{T}\Big\{ \textstyle\sum_{t=1}^{\tau} \Big|F(q^i, \sigma_t(i)) - \tau F(q^i, H^N(b^i|\boldsymbol{b}^{-i}))\Big| + (T-\tau)\Big|\mathbb{E}[F(q^i, \overline{\sigma}(i))] - F(q^i, H^N(b^i|\boldsymbol{b}^{-i}))\Big|\Big\} \\
\overset{(b)}{\le}\ & \tfrac{\Delta_F}{T}\Big(\tau + (T-\tau)\mathbb{P}\{\overline{\sigma}(i) \ne H^N(b^i|\boldsymbol{b}^{-i})\}\Big). \\
\overset{(c)}{\le}\ & \tfrac{\Delta_F}{T}\Big[\tau + (T-\tau)N\exp\Big(-\tfrac{1}{4\tau\eta^2}\min_{j\in[N]\setminus\{i\}}|(\mu^i-\mu^j)\tau + (b^i-b^j)T|^2\Big)\Big].
\end{aligned} \tag{EC.11}$$

Note that (EC.9) and (EC.11) correspond to Part (i) of the theorem. We derive the one-sided bound (EC.10) for ease of future reference in later steps of this proof. ■

Next, we will prove a strengthened version of Theorem 1's Part (ii) by generalizing it to mixed strategies. Formally speaking, for each agent $i \in [N]$, let G^i denote the cumulative distribution function of his (possibly mixed) strategy. Our convergence results apply to a broad class of mixed strategies satisfying the following regularity condition.

DEFINITION EC.3. A mixed-strategy profile $\boldsymbol{G} = \{G^i\}$ is *regular* if, for each $i \in [N]$, G^i can be decomposed into a continuous component with bounded density and a finite set of atoms. That is, there exists a finite set $E^i \subset \mathbb{R}_+$ and a bounded density function $\widetilde{g}^i$ such that

$$G^i(x) = \int_{b \le x} \widetilde{g}^i(b)\, db + \sum_{b \le x,\, b \in E^i} \mathbb{P}_{G^i}(b^i = b).$$

Note that this regularity condition is mild.[11] In particular, pure strategies (point masses), finite mixtures of point masses, and mixed strategies with smooth densities are all included. With this in mind, we will prove the following result, which generalizes Part (ii) of Theorem 1.

PROPOSITION EC.3. *Suppose that $\boldsymbol{G} = (G^i)_{i=1}^N$ is an ε'-equilibrium in the game $\overline{\Gamma}_{N,\infty}$ such that $\varepsilon' \ge 0$ and $\mathbb{P}_{\boldsymbol{G}}\{b^j = b^i\} = 0$ for every $i \ne j$. In other words, the probability that any two agents have the same inflation intensity is 0. Then, for any $\varepsilon > 0$, there exists large $T_0 > 0$ such that $\boldsymbol{G}$ is an $(\varepsilon' + \varepsilon)$-equilibrium in the game $\overline{\Gamma}_{N,T}$ for all $T > T_0$.*

[11] By the Lebesgue–Radon–Nikodym theorem, any probability measure μ on $\mathbb{R}$ can be decomposed as $\mu = g(x)\,dx + \sum_{k=1}^{\infty} p_k \delta_{x_k} + \mu_{sc}$; see Folland 1999, Thm. 3.8. The three components are: (i) an absolutely continuous part with density g, (ii) a countable atomic part, and (iii) a singular-continuous part (e.g., the Cantor distribution). Our regularity condition imposes only mild restrictions on this decomposition: it requires g to be bounded, the atomic component to be finite, and rules out the singular-continuous part. The latter is mainly a technical exclusion, as such distributions are highly irregular and rarely arise in economic models.

We will start with the following lemma.

LEMMA EC.2. *If $\boldsymbol{G}$ is a regular Nash equilibrium, possibly in mixed strategies, of $\overline{\Gamma}_{N,\infty}$, then for every $\varepsilon > 0$, $\boldsymbol{G}$ is an ε-equilibrium of $\overline{\Gamma}_{N,T}$ for all sufficiently large T. That is, for every agent $i \in [N]$,*

$$v^i(G^i, \boldsymbol{G}^{-i}) \geq v^i(b, \boldsymbol{G}^{-i}) - \varepsilon, \qquad \forall b \geq 0.$$

Proof of Lemma EC.2. We break the proof into two main steps.

Step 1: When $\boldsymbol{G}$ is a regular NE, quantify the proximity between $v_T^i(\cdot, \boldsymbol{G}^{-i})$ and $\overline{v}^i(\cdot, \boldsymbol{G}^{-i})$. More precisely, let $\boldsymbol{G}$ be a regular mixed strategy Nash Equilibrium. Invoking Definition EC.3, we have that $G^i(x) = \int_{b \leq x} \widetilde{g}^i(b)\, db + \sum_{b \leq x, b \in E^i} \mathbb{P}_{G^i}(b^i = b)$, where $E^i = \{b \,|\, \mathbb{P}_{G^i}(b^i = b) > 0\}$ is the singular set that G^i has a positive point measure on. Let $E^{-i} := \bigcup_{j \neq i} E^j$. On a high level, the difficulty of this step comes from the deterioration of convergence properties when x becomes closer to the singular set E^{-i}. In that case, the tie-breaking rule becomes potentially important. To overcome this challenge, we simultaneously work with two tie-breaking rules: one that agent i most prefers and the other one that he least prefers. We then develop two one-sided, but uniform bounds for both tie-breaking rules.

We introduce a few constants to be used in this step. First, let $\|\widetilde{g}\|_\infty = \max_{i \in [N]} \max_{x \geq 0} \widetilde{g}^i(x) < +\infty$ be the uniform upper bound of $\{\widetilde{g}^i\}$. Second, Proposition EC.2 implies that the probability of tied inflation is zero at equilibrium. Therefore, the sets $\{E^i : i \in [N]\}$ are disjoint, i.e., $E^i \cap E^j = \emptyset$ for every $i \neq j$. Therefore, we may define $d := \min\{|x - y| : x \neq y \text{ and } x, y \in \cup_j E^j\} > 0$ as the minimum distance between any two singular points. Both constants $\|\widetilde{g}\|_\infty$ and d only depend on the profile $\boldsymbol{G}$. We will show that for every $\varepsilon > 0$, there exists $\hat{T}_i$ such that when $T > \hat{T}_i$,

$$v_T^i(b^i, \boldsymbol{G}^{-i}) - \overline{v}_{M^i}^i(G^i, \boldsymbol{G}^{-i}) \leq \varepsilon/2 \quad \text{for all } b^i \in \mathbb{R}_+ \tag{EC.12}$$
$$\overline{v}^i(G^i, \boldsymbol{G}^{-i}) - v_T^i(G^i, \boldsymbol{G}^{-i}) \leq \varepsilon/2 \tag{EC.13}$$

Verifying (EC.12). Since $\tau = o(\sqrt{T})$, given $i \in [N]$, there exists $T_i < +\infty$, only dependent on the profile $\boldsymbol{G}$, agent i, and problem primitives, such that as long as $T > T_i$, it holds that $T > \max\left\{\tau \max_i\{|\mu^N - \mu^i|, |\mu^1 - \mu^i|\}, \frac{2}{d}\right\}^2 = \max\left\{\tau \Delta_\mu, \frac{2}{d}\right\}^2$, where we define the constant $\Delta_\mu := \max_i\{|\mu^N - \mu^i|, |\mu^1 - \mu^i|\}$ for shorthand notation.

Fix any $b^i \geq 0$ for this part. Consider the interval $[b^i, b^i + 1/\sqrt{T})$. Since $\sqrt{T} > 2/d$, we have $1/\sqrt{T} < d/2$. Therefore, the interval $[b^i, b^i + 1/\sqrt{T})$ contains at most one singular point (i.e., a point in E^{-i}). In other words, there is at most one agent $k \neq i$ and a singular inflation intensity b_* such that $E^k \cap [b^i, b^i + 1/\sqrt{T}) = \{b_*\}$. With those in mind, consider the following events regarding agent i's opponents' (agents $j \neq i$) inflation intensities:

- $\mathcal{E}_G := \{\boldsymbol{b}^{-i} : \text{ either } b^j < b^i \text{ or } b^j \geq b^i + 1/\sqrt{T}, \text{ for all } j \neq i\}$. This is the "good" event where agent i's opponents' inflation intensities are either (i) strictly less than b^i; or (ii) larger than b^i by at least a margin of $1/\sqrt{T}$.
- $\mathcal{E}_B := \{\boldsymbol{b}^{-i} : \ b^j \in [b^i, b^i + 1/\sqrt{T}), \text{ for some } j \neq i\} \setminus \{\boldsymbol{b}^{-i} : \ b^k = b_*\}$. This is the "bad" event where some opponents' inflation intensities lie in the interval $[b^i, b^i + 1/\sqrt{T})$, but no one uses the singular inflation intensity b_*.
- $\mathcal{E}_S := \{\boldsymbol{b}^{-i} : \ b^j \in [b^i, b^i + 1/\sqrt{T}), \text{ for some } j \neq i\} \cap \{\boldsymbol{b}^{-i} : \ b^k = b_*\}$. This is a "singular" event where some opponent's inflation intensity lies in the interval $[b^i, b^i + 1/\sqrt{T})$, among whom agent k uses the singular intensity b_*. If $[b^i, b^i + 1/\sqrt{T})$ contains no singular points, this set is empty and we ignore this event.

By construction, $\{\mathcal{E}_G, \mathcal{E}_S, \mathcal{E}_B\}$ forms a partition of all possibilities about $\boldsymbol{b}^{-i}$. Suppose without loss that the singular point $b_* \in [b^i, b^i + 1/\sqrt{T})$ exists.[12] It holds that

$$\begin{aligned}
& v_T^i(b^i, \boldsymbol{G}^{-i}) - \overline{v}_{M^i}^i(G^i, \boldsymbol{G}^{-i}) \\
\overset{(a)}{\leq}\ & v_T^i(b^i, \boldsymbol{G}^{-i}) - \overline{v}_{M^i}^i(b_*, \boldsymbol{G}^{-i}) \\
\leq\ & \int \left[v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}_{M^i}^i(b_*, \boldsymbol{b}^{-i})\right] d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i}) \\
=\ & \int_{\mathcal{E}_B} \left[v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}_{M^i}^i(b_*, \boldsymbol{b}^{-i})\right] d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i}) + \int_{\mathcal{E}_G \cup \mathcal{E}_S} \left[v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}_{M^i}^i(b_*, \boldsymbol{b}^{-i})\right] d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i}) \\
\overset{(b)}{\leq}\ & \underbrace{\Delta_F\, \mathbb{P}_{\boldsymbol{G}^{-i}}(\mathcal{E}_B)}_{(*)} + \sup_{\boldsymbol{b}^{-i} \in \mathcal{E}_G} \underbrace{\left[v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}_{M^i}^i(b_*, \boldsymbol{b}^{-i})\right]}_{(**)} + \underbrace{\int_{\mathcal{E}_S} \left[v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}_{M^i}^i(b_*, \boldsymbol{b}^{-i})\right] d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i})}_{(***)},
\end{aligned} \tag{EC.14}$$

where (a) is because $\boldsymbol{G}$ is a Nash equilibrium; (b) is because $v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}_{M^i}^i(b^i, \boldsymbol{b}^{-i}) \leq \Delta_F = F(1,1) - F(0,0)$. We will separately bound the three terms on the right hand side of the last inequality.

First, note that

$$\begin{aligned}
(*) = \Delta_F\, \mathbb{P}_{\boldsymbol{G}^{-i}}(\mathcal{E}_B) \overset{(a)}{\leq}\ & \Delta_F \Big[\sum_{j \in [N] \setminus \{i,k\}} \mathbb{P}_{\boldsymbol{G}^{-i}}(b^i \leq\ b^j < b^i + \tfrac{1}{\sqrt{T}}) + \mathbb{P}_{\boldsymbol{G}^{-i}}\big(b^k \in [b^i, b^i + \tfrac{1}{\sqrt{T}}) \setminus \{b_*\}\big) \Big] \\
\overset{(b)}{=}\ & \Delta_F \sum_{j \neq i} \int_{b^i}^{b^i + 1/\sqrt{T}} \widetilde{g}^j(y)\, dy \leq \tfrac{N-1}{\sqrt{T}} \Delta_F \|\widetilde{g}\|.
\end{aligned} \tag{EC.15}$$

In the derivations above, (a) is due to the definition of $\mathcal{E}_B$; (b) comes from the fact that $\boldsymbol{G}$ is regular (see Definition EC.3) as well as the union bound.

If $\boldsymbol{b}^{-i} \in \mathcal{E}_G$, we have either $b^j < b^i$ or $b^j \geq b^i + 1/\sqrt{T}$ for all $j \neq i$. For simplicity of notation, let $\boldsymbol{b}^i$ be the completed pure-strategy inflation profile with b^i for agent i. For every $j \in \{j : b^j > b^i\}$, we have

$$\tau(\mu^j - \mu^i) + T(b^j - b^i) \overset{(a)}{\geq}\ \sqrt{T} - \tau|\mu^j - \mu^i| \overset{(b)}{\geq}\ \sqrt{T} - \tau \max\{|\mu^N - \mu^i|, |\mu^1 - \mu^i|\} \geq \sqrt{T} - \tau\Delta_\mu \overset{(c)}{>} 0.$$

In the derivation above, (a) is because $b^j - b^i \geq 1/\sqrt{T}$; (b) holds by the monotonicity of $\{\mu^j\}$; (c) holds because $T > T_i$. Therefore, (EC.6) holds. In other words, $T > \overline{T}_i(\boldsymbol{b}^i)$ and (EC.9) is applicable to $\boldsymbol{b}^i$. As a result,

$$\begin{aligned}
(**) \leq v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}_{M^i}^i(b^i, \boldsymbol{b}^{-i}) \ &\leq\ \tfrac{\Delta_F}{T}\Big[\tau + (T-\tau)N \exp\Big(-\tfrac{1}{4\tau\eta^2} \min_{j: b^j > b^i} |(\mu^i - \mu^j)\tau + (b^i - b^j)T|^2\Big)\Big] \\
&\leq\ \tfrac{\Delta_F}{T}\Big[\tau + (T-\tau)N \exp\Big(-\tfrac{1}{4\tau\eta^2}\big(\sqrt{T} - \tau\Delta_\mu\big)^2\Big)\Big],
\end{aligned} \tag{EC.16}$$

which is a term independent of $\boldsymbol{b}^{-i}$.

If $\boldsymbol{b}^{-i} \in \mathcal{E}_S$, there is a unique agent $k \neq i$ and singular inflation intensity $b_* \in E^k$ such that $E^k \cap [b^i, b^i + 1/\sqrt{T}) = \{b_*\}$. Therefore, we could write the complete inflation profile as $(b^i, b_*, \boldsymbol{b}^{-(i,k)})$. Similar to the arguments above, we can partition the values of $\boldsymbol{b}^{-(i,k)}$ into two events. Let $\mathcal{E}'_G := \{\boldsymbol{b}^{-(i,k)} :$ either $b^j < b_*$ or $b^j \geq b_* + 1/\sqrt{T},$ for all $j \in [N] \setminus \{i,k\}\}$ be the "good" event and let $\mathcal{E}'_B :=$

[12] Otherwise, $[b^i, b^i + 1/\sqrt{T})$ contains no singular points. That case is easier and we will comment on it later.

$\{\boldsymbol{b}^{-(i,k)} : b^j \in [b_*, b_* + 1/\sqrt{T}),\ \text{for some } j \in [N] \setminus \{i,k\}\}$. With slight abuse of notation, it holds that for all $\boldsymbol{b}^{-i} \in \mathcal{E}_S$,

$$
\begin{aligned}
& v_T^i(b^i, \boldsymbol{b}^{-i}) - \overline{v}_{M^i}^i(b_*, \boldsymbol{b}^{-i}) \\
\overset{(a)}{\leq}\ & v_T^i(b_*, \boldsymbol{b}^{-i}) + \tfrac{c}{\sqrt{T}} - \overline{v}_{M^i}^i(b_*, \boldsymbol{b}^{-i}) \\
=\ & v_T^i(b_*, b_*, \boldsymbol{b}^{-(i,k)}) - \overline{v}_{M^i}^i(b_*, b_*, \boldsymbol{b}^{-(i,k)}) + \tfrac{c}{\sqrt{T}} \\
\overset{(b)}{\leq}\ & \Delta_F\, \mathbb{P}_{\boldsymbol{G}^{-(i,k)}}(\mathcal{E}'_B) + \sup_{\boldsymbol{b}^{-(i,k)} \in \mathcal{E}'_G} \big[v_T^i(b_*, b_*, \boldsymbol{b}^{-(i,k)}) - \overline{v}_{M^i}^i(b_*, b_*, \boldsymbol{b}^{-(i,k)}) \big] + \tfrac{c}{\sqrt{T}} \\
\overset{(c)}{\leq}\ & \tfrac{N-2}{\sqrt{T}} \Delta_F \|\widetilde{g}\| + \tfrac{\Delta_F}{T} \Big[\tau + (T-\tau) N \exp\Big(-\tfrac{1}{4\tau\eta^2} \min_{j: b^j > b_*} |(\mu^i - \mu^j)\tau + (b_* - b^j)T|^2 \Big) \Big] + \tfrac{c}{\sqrt{T}} \\
<\ & \tfrac{N-1}{\sqrt{T}} \Delta_F \|\widetilde{g}\| + \tfrac{\Delta_F}{T} \Big[\tau + (T-\tau) N \exp\Big(-\tfrac{1}{4\tau\eta^2} \min_{j: b^j > b_*} |(\mu^i - \mu^j)\tau + (b_* - b^j)T|^2 \Big) \Big] + \tfrac{c}{\sqrt{T}} \\
\overset{(d)}{\leq}\ & \tfrac{N-1}{\sqrt{T}} \Delta_F \|\widetilde{g}\| + \tfrac{\Delta_F}{T} \Big[\tau + (T-\tau) N \exp\Big(-\tfrac{1}{4\tau\eta^2} \big(\sqrt{T} - \tau\Delta_\mu\big)^2 \Big) \Big] + \tfrac{c}{\sqrt{T}}.
\end{aligned}
\tag{EC.17}
$$

In derivations above, (a) is because $b_* \in [b^i, b^i + 1/\sqrt{T})$, which implies that $v_T^i(b^i, \boldsymbol{b}^{-i}) \leq v_T^i(b_*, \boldsymbol{b}^{-i}) + c(b_* - b^i) < v_T^i(b_*, \boldsymbol{b}^{-i}) + \frac{c}{\sqrt{T}}$; (b) follows from similar steps in (EC.14); (c) follows from similar steps for (EC.15) and (EC.16), respectively, plus the fact that the interval $[b^i, b^i + 1/\sqrt{T})$ contains a unique singular point; in (d), given that $\boldsymbol{b}^{-(i,k)} \in \mathcal{E}'_G$, we can apply the same arguments in the second inequality of (EC.16) to obtain the same upper bound.

Combining all the pieces together from (EC.14), (EC.15), (EC.16), and (EC.17), we conclude that

$$
v_T^i(b^i, \boldsymbol{G}^{-i}) - \overline{v}_{M^i}^i(G^i, \boldsymbol{G}^{-i}) \leq \tfrac{2(N-1)}{\sqrt{T}} \Delta_F \|\widetilde{g}\| + \tfrac{2\Delta_F}{T} \Big[\tau + (T-\tau) N \exp\Big(-\tfrac{1}{4\tau\eta^2} \big(\sqrt{T} - \tau\Delta_\mu\big)^2 \Big) \Big] + \tfrac{c}{\sqrt{T}},
$$

which is a bound independent of b^i. Moreover, the upper bound diminishes to zero when $\tau = o(\sqrt{T})$ and T grows. That proves (EC.12). Note that for now, we have assumed that the singular point $b_* \in [b^i, b^i + 1/\sqrt{T})$ exists. (See Footnote 12.) Otherwise, it suffices to replace "b_*" with "b^i" in (EC.14) and repeat the same process while ignoring term $(***)$.

Verifying (EC.13). Note that $\overline{v}_{M^i}^i(G^i, \boldsymbol{G}^{-i})$ and $\overline{v}^i(G^i, \boldsymbol{G}^{-i})$ represent agent i's utilities under his most preferable and least preferable tie-breaking rules, respectively. Therefore, we can repeat the same process for (EC.12) while flipping the direction, thus proving (EC.13).

Step 2: Establish the ε-equilibrium. Let $\boldsymbol{G}$ be a regular Nash Equilibrium. Invoking Proposition EC.2, $\boldsymbol{G}$ is also an equilibrium in the M^i-limiting game. Therefore,

$$
\begin{aligned}
\overline{v}_{M^i}^i(G^i, \boldsymbol{G}^{-i}) \ =\ & \int_{\mathbb{R}_+/E^{-i}} \int_{\mathbb{R}_+^{N-1}} \overline{v}_{M^i}^i(b^i, \boldsymbol{b}^{-i})\, d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i})\, dG^i(b^i) + \sum_{b^i \in E^{-i}} \mathbb{P}_{G^i}(\{b^i\}) \int_{\mathbb{R}_+^{N-1}} \overline{v}_{M^i}^i(b^i, \boldsymbol{b}^{-i})\, d\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i}) \\
\overset{(a)}{=}\ & \int_{\mathbb{R}_+/E^{-i}} \int_{\mathbb{R}_+^{N-1}} \overline{v}_{M^i}^i(b^i, \boldsymbol{b}^{-i})\, d\,\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i})\, d\, G^i(b^i) \\
\overset{(b)}{=}\ & \int_{\mathbb{R}_+/E^{-i}} \int_{\mathbb{R}_+^{N-1}} \overline{v}^i(b^i, \boldsymbol{b}^{-i})\, d\,\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i})\, d\, G^i(b^i) \\
\overset{(c)}{=}\ & \int_{\mathbb{R}_+/E^{-i}} \int_{\mathbb{R}_+^{N-1}} \overline{v}^i(b^i, \boldsymbol{b}^{-i})\, d\,\boldsymbol{G}^{-i}(\boldsymbol{b}^{-i})\, d\, G^i(b^i) + \sum_{b^i \in E^{-i}} \mathbb{P}_{G^i}(\{b^i\}) \int_{\mathbb{R}_+^{N-1}} \overline{v}^i(b^i, \boldsymbol{b}^{-i})\, d\, G(\boldsymbol{b}^{-i}) \\
=\ & \overline{v}^i(G^i, \boldsymbol{G}^{-i}).
\end{aligned}
\tag{EC.18}
$$

In (a) and (c) of the derivations above, by Proposition EC.2, we know that $\mathbb{P}_{G^i}(\{x\}) = 0$ for every $x \in E^{-i}$; therefore, the second terms in the LHS of (a) and in the RHS of (c) are both zero. For (b),

it suffices to notice that $\overline{v}^i_{M^i}(b^i, \boldsymbol{b}^{-i}) = \overline{v}^i(b^i, \boldsymbol{b}^{-i})$ if $b^i \neq b^j$ for any $j \neq i$, i.e., there is no tie. Now, pick an arbitrary $\varepsilon > 0$, $T \geq \max\{\hat{T}_i\}$, agent $i \in [N]$ and intensity $b^i \in \mathbb{R}_+$. It holds that

$$\begin{aligned} v^i_T(b^i, \boldsymbol{G}^{-i}) - v^i_T(G^i, \boldsymbol{G}^{-i}) &= v^i_T(b^i, \boldsymbol{G}^{-i}) - \overline{v}^i(G^i, \boldsymbol{G}^{-i}) + \overline{v}^i(G^i, \boldsymbol{G}^{-i}) - v^i_T(G^i, \boldsymbol{G}^{-i}) \\ &\overset{(a)}{=} v^i_T(b^i, \boldsymbol{G}^{-i}) - \overline{v}^i_{M^i}(G^i, \boldsymbol{G}^{-i}) + \overline{v}^i(G^i, \boldsymbol{G}^{-i}) - v^i_T(G^i, \boldsymbol{G}^{-i}) \\ &\overset{(b)}{\leq} \varepsilon/2 + \varepsilon/2 = \varepsilon, \end{aligned}$$

where (a) follows from (EC.18); (b) follows from applying (EC.12) and (EC.13). ■

We are now ready to prove Proposition EC.3.

Proof of Proposition EC.3. Since $\mathbb{P}_{\boldsymbol{G}}(\widetilde{b}^j = \widetilde{b}^i) = 0$ for any $i \neq j$, the utility of inflation intensity profile $\boldsymbol{G}$ in all M-limiting games is the same as they only differ in the tie-breaking rule. Since $\boldsymbol{G}$ is ε'-equilibrium in $\overline{M}$-limiting game, by Corollary EC.1, $\boldsymbol{G}$ is also ε'-equilibrium in M^i-limiting game, i.e., for any inflation intensity $x \in \mathbb{R}_+$, $\overline{v}^i_{M^i}(x, \boldsymbol{G}^{-i}) \leq \varepsilon' + \overline{v}^i_{M^i}(G^i, \boldsymbol{G}^{-i}) = \varepsilon' + \overline{v}^i(G^i, \boldsymbol{G}^{-i})$. Hence, for any $\varepsilon > 0$, there exists T_0 such that for all $T \geq T_0$, for any inflation intensity $x \in \mathbb{R}_+$,

$$v^i(x, \boldsymbol{G}^{-i}) - v^i(G^i, \boldsymbol{G}^{-i}) \leq v^i(x, \boldsymbol{G}^{-i}) - \overline{v}^i_{M^i}(x, \boldsymbol{G}^{-i}) + \varepsilon' + \overline{v}^i(G^i, \boldsymbol{G}^{-i}) - v^i(G^i, \boldsymbol{G}^{-i}) \leq \varepsilon + \varepsilon', \quad \text{(EC.19)}$$

where the two inequalities follow from Proposition EC.2 and Theorem 1, in turn. Therefore, $\boldsymbol{G}$ is an $(\varepsilon + \varepsilon')$-equilibrium for the game $\overline{\Gamma}_{N,T}$ with $T > T_0$. ■

Proof of Theorem 2. Pick any continuous and strictly increasing π and $N \geq 2$ as required. We prove the two claims respectively.

Claim 1: $\left|\overline{v}^i(\cdot, \boldsymbol{\pi}_N^{-i}) - U_\pi(q^i, \cdot)\right|$ is "uniformly small." Pick any $i \in [N]$ and b^i. Let us start by bounding the gap between the ranking positions. Recall the expressions of $H^N(b^i|\boldsymbol{b}^{-i})$ in (5) and $H(b^i|\pi)$ in (7), respectively. Since π is continuous and strictly increasing, the inverse mapping π^{-1} exists on the interval $[\pi(0), \pi(1)]$. Without loss, suppose $b^i \in [\pi(0), \pi(1)]$. Therefore,

$$H(b^i|\pi) = \int_0^1 \mathbb{I}\{\pi(q) \leq b^i\}\, dq = \int_0^1 \mathbb{I}\{q \leq \pi^{-1}(b^i)\}\, dq = \int_0^{\pi^{-1}(b^i)} 1\, dq = \pi^{-1}(b^i).$$

Similarly, recall that $\{q^i\}_i = \{\frac{i-1}{N-1}\}_i$ form a uniform mesh grid on the uniform interval $[0, 1]$. Hence,

$$H^N(b^i \,|\, \boldsymbol{\pi}_N^{-i}) = \tfrac{1}{N-1} \sum_{j \in [N]\setminus\{i\}} \mathbb{I}\{\pi(q^j) < b^i\} = \tfrac{1}{N-1} \sum_{j \in [N]\setminus\{i\}} \mathbb{I}\{q^j < \pi^{-1}(b^i)\} = \tfrac{1}{N-1} \lceil (N-1)\pi^{-1}(b^i) \rceil.$$

As a result,

$$\left|H^N(b^i \,|\, \boldsymbol{\pi}_N^{-i}) - H(b^i|\pi)\right| \;=\; \left|\tfrac{1}{N-1} \lceil (N-1)\pi^{-1}(b^i) \rceil - \pi^{-1}(b^i)\right| \;\leq\; \tfrac{1}{N-1}. \quad \text{(EC.20)}$$

As a consequence,

$$\begin{aligned} \left|\overline{v}^i(b^i, \boldsymbol{\pi}_N^{-i}) - U_\pi(q^i, b^i)\right| &\overset{(a)}{=} \left|\Big(F\big(q^i, H^N(b^i|\boldsymbol{\pi}_N^{-i})\big) - cb^i\Big) - \Big(F(q^i, H(b^i|\pi)) - cb^i\Big)\right| \\ &= \left|F\big(q^i, H^N(b^i|\boldsymbol{\pi}_N^{-i})\big) - F(q^i, H(b^i|\pi))\right| \overset{(b)}{\leq} \tfrac{L_F}{N-1}. \end{aligned}$$

In the derivations above, (a) follows directly from the definition of $\overline{v}^i$ in (6) and U_π in (8); in (b), we use Assumption 2, which states that $F(\cdot)$ is continuously differentiable in $[0, 1]^2$ and hence Lipschitz continuous. Therefore, there exists a constant $L_F > 0$, depending only on F, so that $|F\big(q^i, H^N(b^i|\boldsymbol{\pi}_N^{-i})\big) - F(q^i, H(b^i|\pi))| \leq \frac{L_F}{N-1}$.

Claim 2: $\boldsymbol{\pi}_N$ is an approximate equilibrium in $\overline{\Gamma}_{N,\infty}$. Suppose further that $\{\pi(q): q\in[0,1]\}$ forms an NAE. It holds that for every $i\in[N]$ and $b^i\geq 0$

$$\overline{v}^i(\pi(q^i),\boldsymbol{\pi}_N^{-i}) \overset{(a)}{\geq} U_\pi(q^i,\pi(q^i)) - \frac{L_F}{N} \overset{(b)}{\geq} U_\pi(q^i,b^i) - \frac{L_F}{N} \overset{(c)}{\geq} \overline{v}^i(b^i,\boldsymbol{\pi}_N^{-i}) - \frac{2L_F}{N},$$

where (a) and (c) follow from the observation in Claim 1 of this result; (b) follows from the definition of NAE. The claim is proved by picking $N_0 = 100L_F/\varepsilon$. ■

D. Analysis of $\overline{\Gamma}_{\infty,\infty}$ and the NAE Proof of Theorem 3. First, notice that with the closed-form expression $\overline{\pi}(q) = \int_0^q \frac{F_2(t,t)}{c}dt$, it can be verified that $\overline{\pi}$ is a differentiable and strictly increasing function because Assumption 2 implies that $F_2(t,t)>0$ for all $t\in[0,1]$. Following the same argument in the proof of Theorem 2, the inverse mapping $\overline{\pi}^{-1}$ exists on the interval $[\overline{\pi}(0),\overline{\pi}(1)] = [0,\overline{\pi}(1)]$, and

$$H(b|\overline{\pi}) = \begin{cases} \overline{\pi}^{-1}(b) & \text{if } b\in[0,\overline{\pi}(1)];\\ 1 & \text{if } b\geq\overline{\pi}(1).\end{cases}$$

Therefore,

$$U_{\overline{\pi}}(q,b) = F(q,H(b|\overline{\pi})) - cb = \begin{cases} F(q,\overline{\pi}^{-1}(b)) - cb & \text{if } b\in[0,\overline{\pi}(1)];\\ F(q,1) - cb & \text{if } b>\overline{\pi}(1).\end{cases}$$

To verify that $\{\overline{\pi}(q): q\in[0,1]\}$ forms an NAE, we pick an arbitrary $q\in[0,1]$ and show that the function $U_{\overline{\pi}}(q,\cdot)$ increases in the interval $[0,\overline{\pi}(q)]$ and decreases in the interval $[\overline{\pi}(q),+\infty)$. That implies $\overline{\pi}(q)\in\arg\max_{b\geq 0} U_{\overline{\pi}}(q,b)$ and thus finishes the proof. Since it is straightforward to verify that $U_{\overline{\pi}}(q,b)$ decreases when $b>\overline{\pi}(1)$, let us pick $b\in[0,\overline{\pi}(1))$ and make the following observations:

$$\frac{\partial}{\partial b}U_{\overline{\pi}}(q,b) = \frac{\partial}{\partial b}\Big[F(q,\overline{\pi}^{-1}(b)) - cb\Big] = \frac{F_2(q,\overline{\pi}^{-1}(b))}{\overline{\pi}'(\overline{\pi}^{-1}(b))} - c = c\frac{F_2(q,\overline{\pi}^{-1}(b))}{F_2(\overline{\pi}^{-1}(b),\overline{\pi}^{-1}(b))} - c.$$

When $b=\overline{\pi}(q)$, we have $\overline{\pi}^{-1}(b)=q$, and hence $\frac{F_2(q,\overline{\pi}^{-1}(b))}{F_2(\overline{\pi}^{-1}(b),\overline{\pi}^{-1}(b))} = \frac{F_2(q,q)}{F_2(q,q)} = 1$. In other words, the first-order condition holds. To verify the global optimality of the inflation intensity $\overline{\pi}(q)$, note that when $b<\overline{\pi}(q)$, we have $\overline{\pi}^{-1}(b)<q$ and therefore $\frac{F_2(q,\overline{\pi}^{-1}(b))}{F_2(\overline{\pi}^{-1}(b),\overline{\pi}^{-1}(b))}\geq 1$ because F is supermodular (see Assumption 2), which further implies that $\frac{\partial}{\partial b}U_{\overline{\pi}}(q,b)\geq 0$. Similarly, when $b>\overline{\pi}(q)$, it holds that $\frac{\partial}{\partial b}U_{\overline{\pi}}(q,b)\leq 0$. ■

Next, we prove Theorem 4. Note that it utilizes the analysis of game $\overline{\Gamma}_{N,\infty}$, i.e., Theorem 2.

Proof of Theorem 4. We first note that here we no longer require τ to be a multiple of N. As such, v^i may potentially depend on the initial ranking. However, that is negligible in our analysis. In fact, let us denote agent i's expected reward in the experimentation phase as $\Delta_\tau^i := \mathbb{E}\Big[\sum_{t=1}^\tau F(q^i,\sigma_t(i))\Big]$. We will use the following observation later:

$$\big|\Delta_\tau^i - \Delta_\tau^j\big| \leq \sum_{t=1}^\tau \mathbb{E}\left[\big|F(q^i,\sigma_t(i)) - F(q^j,\sigma_t(j))\big|\right] \leq \tau[F(1,1)-F(0,0)] = \tau\Delta_F, \tag{EC.21}$$

which is a constant independent of agents i,j and the initial ranking.

Now, fix an agent $i\in[N]$ and an inflation intensity $b^i\geq 0$. Suppose agent i uses a total corruption budget of Tb^i and the rest of the agents use $T\overline{\boldsymbol{\pi}}_N^{-i}$. We break the rest of the proof into the following steps.

Step 1: Quantify the proximity between $\overline{\sigma}(i)$ and $H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i})$. Recall that agent i locks in the position $\overline{\sigma}(i)$ during the commitment phase, which is expressed as $\overline{\sigma}(i) = \frac{1}{N-1}\sum_{j\in[N]\setminus\{i\}} \mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\}$. At the same time, recall from (5) that $H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i}) = \frac{1}{N-1}\sum_{j\in[N]\setminus\{i\}} \mathbb{I}\{\overline{\pi}(q^j) < b^i\}$ is the inflation-intensity based ranking. Let $s_\tau := (\tau\Delta_F + \log T)/T$ for shorthand notation. We observe that

$$\begin{aligned}
&\mathbb{E}\Big[\big|\overline{\sigma}(i) - H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i})\big|\Big] \\
\le\ & \mathbb{E}\Big[\Big|\big(\tfrac{1}{N-1}\sum_{j\in[N]\setminus\{i\}} \mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\}\big) - \big(\tfrac{1}{N-1}\sum_{j\in[N]\setminus\{i\}} \mathbb{I}\{\overline{\pi}(q^j) < b^i\}\big)\Big|\Big] \\
\le\ & \tfrac{1}{N-1}\sum_{j\in[N]\setminus\{i\}} \mathbb{E}\Big[\big|\mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\} - \mathbb{I}\{\overline{\pi}(q^j) < b^i\}\big|\Big] \\
\le\ & \tfrac{1}{N-1}\left(\underbrace{\sum_{j:|\overline{\pi}(q^j)-b^i|<s_\tau} 1}_{(*)} + \underbrace{\sum_{j:|\overline{\pi}(q^j)-b^i|\ge s_\tau} \mathbb{E}\Big[\big|\mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\} - \mathbb{I}\{\overline{\pi}(q^j) < b^i\}\big|\Big]}_{(**)}\right)
\end{aligned}$$

We have thus decomposed the expected discrepancy between $\overline{\sigma}$ and H^N into two parts. The first term $(*)$ corresponds to the agents whose inflation intensities are "close" to b^i; the second term $(**)$ corresponds to agents whose inflation intensities are "far" from b^i. In the derivations below, we analyze terms $(*)$ and $(**)$, respectively. Intuitively, term $(*)$ is small because under the equilibrium inflation intensity function $\overline{\pi}(q) = \int_0^q \frac{F_2(t,t)}{c}dt$ in Theorem 3, there is no pathological "clustering" of the agents' inflation intensities. Term $(**)$ is small because when the inflation intensity difference is large, the learner's ranking is almost dominated by inflation.

To analyze term $(*)$, recall from Assumption 2 that F_2 is positively valued and continuous on $[0,1]^2$. Therefore, there exists a constant $\delta_F^* := \min_{t\in[0,1]} F_2(t,t) > 0$. As a result, pick arbitrary q, q' such that $q' - q \ge cs_\tau/\delta_F^*$. Then,

$$|\overline{\pi}(q') - \overline{\pi}(q)| \ =\ |\textstyle\int_q^{q'} \frac{F_2(t,t)}{c}dt| \overset{(a)}{\ge} |\int_q^{q'} \frac{\delta_F^*}{c}dt| = \frac{\delta_F^*}{c}|q'-q| \ \ge\ \frac{\delta_F^*}{c}\frac{cs_\tau}{\delta_F^*} \ =\ s_\tau. \tag{EC.22}$$

In the derivation above, (a) comes from the definition of δ_F^*. Invoking (EC.22), $|\overline{\pi}(q^j) - b^i| < s_\tau$ implies that $|\overline{\pi}^{-1}(b^i) - q^j| < \frac{cs_\tau}{\delta_F^*}$. Therefore,

$$(*) = \sum_{j:|\overline{\pi}(q^j)-b^i|<s_\tau} 1 \ \le\ \big|\{j : |q^j - \overline{\pi}^{-1}(b^i)| < \tfrac{cs_\tau}{\delta_F^*}\}\big| \overset{(a)}{\le} \frac{2cs_\tau(N-1)}{\delta_F^*} \overset{(b)}{=} \frac{2(N-1)c(\tau\Delta_F + \log T)}{T\delta_F^*},$$

where (a) is because $\{q^j\}_j = \{(j-1)/(N-1)\}_j$ forms a uniform mesh grid on the unit interval $[0,1]$; (b) follows from the fact that $s_\tau = (\tau\Delta_F + \log T)/T$.

To analyze term $(**)$, note that for every j such that $b^i \ge \overline{\pi}(q^j) + s_\tau$, it holds that

$$\begin{aligned}
&\mathbb{P}\Big[\mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\} \ne \mathbb{I}\{\overline{\pi}(q^j) < b^i\}\Big] \\
=\ & \mathbb{P}\Big[\mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\} = 0\Big] \\
=\ & \mathbb{P}\Big[\Delta_\tau^i + \sum_{t=1}^\tau \epsilon_t^i + Tb^i \le \Delta_\tau^j + \sum_{t=1}^\tau \epsilon_t^j + T\overline{\pi}(q^j)\Big] \\
=\ & \mathbb{P}\Big[\sum_{t=1}^\tau (\epsilon_t^i - \epsilon_t^j) \le \Delta_\tau^j - \Delta_\tau^i + T(\overline{\pi}(q^j) - b^i)\Big]
\end{aligned}$$

$$\overset{(a)}{\leq} \mathbb{P}\Big[\sum_{t=1}^{\tau}(\epsilon_t^i - \epsilon_t^j) \leq -\log T\Big]$$
$$\overset{(b)}{\leq} \exp\Big(-\frac{\log^2 T}{4\tau\eta^2}\Big).$$

In the derivation above, (a) is because $\Delta_\tau^j - \Delta_\tau^i + T(\overline{\pi}(q^j) - b^i) \leq \Delta_\tau^j - \Delta_\tau^i - Ts_\tau = \Delta_\tau^j - \Delta_\tau^i - \tau\Delta_F - \log T \leq -\log T$ by the construction of s_τ and (EC.21); (b) is due to the Hoeffding's inequality. Similarly, if $b^i \leq \overline{\pi}(q^j) - s_\tau$, it holds that $\Delta_\tau^j - \Delta_\tau^i + T(\overline{\pi}(q^j) - b^i) \geq \Delta_\tau^j - \Delta_\tau^i + Ts_\tau = \Delta_\tau^j - \Delta_\tau^i + \tau\Delta_F + \log T \geq \log T$. Therefore,

$$\begin{aligned}
&\mathbb{P}\Big[\mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\} \neq \mathbb{I}\{\overline{\pi}(q^j) < b^i\}\Big] \\
&= \mathbb{P}\Big[\mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\} = 1\Big] \\
&= \mathbb{P}\Big[\Delta_\tau^i + \sum_{t=1}^{\tau}\epsilon_t^i + Tb^i > \Delta_\tau^j + \sum_{t=1}^{\tau}\epsilon_t^j + T\overline{\pi}(q^j)\Big] \\
&\leq \mathbb{P}\Big[\sum_{t=1}^{\tau}(\epsilon_t^i - \epsilon_t^j) > \log T\Big] \\
&\leq \exp\Big(-\frac{\log^2 T}{4\tau\eta^2}\Big).
\end{aligned}$$

As a result,

$$(**) = \sum_{j:|\overline{\pi}(q^j)-b^i|\geq s_\tau} \mathbb{P}\Big[\mathbb{I}\{\widetilde{D}^j < \widetilde{D}^i\} \neq \mathbb{I}\{\overline{\pi}(q^j) < b^i\}\Big] \leq (N-1)\exp\Big(-\tfrac{\log^2 T}{4\tau\eta^2}\Big).$$

To summarize,

$$\begin{aligned}
\mathbb{E}\Big[\big|\overline{\sigma}(i) - H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i})\big|\Big] &\leq \tfrac{1}{N-1}\Big(\tfrac{2(N-1)c(\tau\Delta_F+\log T)}{T\delta_F^*} + (N-1)\exp\Big(-\tfrac{\log^2 T}{4\tau\eta^2}\Big)\Big) \\
&= \tfrac{2c(\tau\Delta_F+\log T)}{T\delta_F^*} + \exp\Big(-\tfrac{\log^2 T}{4\tau\eta^2}\Big) \qquad \text{(EC.23)}
\end{aligned}$$

Step 2: Quantify the proximity between $\overline{\sigma}(i)$ and $H(b^i \,|\, \overline{\boldsymbol{\pi}})$. Theorem 3 implies that $\overline{\pi}(\cdot)$ is strictly increasing. Therefore, following the same steps for (EC.20) in the proof of Theorem 2, it holds that

$$\Big|H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i}) - H(b^i|\overline{\pi})\Big| \leq \tfrac{1}{N-1}.$$

With both (EC.23) and the result above in mind, if agent i uses budget Tb^i and all other agents use budgets $T\overline{\boldsymbol{\pi}}_N^{-i}$,

$$\begin{aligned}
\mathbb{E}\Big[\big|\overline{\sigma}(i) - H(b^i|\overline{\pi})\big|\Big] &\leq \mathbb{E}\Big[\big|\overline{\sigma}(i) - H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i})\big|\Big] + \mathbb{E}[\Big|H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i}) - H(b^i|\overline{\pi})\Big|] \\
&\leq \tfrac{2c(\tau\Delta_F+\log T)}{T\delta_F^*} + \exp(-\tfrac{\log^2 T}{4\tau\eta^2}) + \tfrac{1}{N-1} \\
&= \tfrac{2c\Delta_F}{\delta_F^*}\tfrac{\tau}{T} + \tfrac{2c}{\delta_F^*}\tfrac{\log T}{T} + \exp(-\tfrac{1}{4\eta^2}\tfrac{\log^2 T}{\tau}) + \tfrac{1}{N-1}.
\end{aligned}$$

Step 3: Quantify the proximity between the pre-limit utility value $v^i(b^i, \overline{\boldsymbol{\pi}}_N^{-i})$ and $U_{\overline{\pi}}(q^i, b^i)$. Recall from (8) that $U_\pi(q,b) = F(q, H(b|\pi)) - cb$ for $b \geq 0$. Therefore,

$$T\Big|v^i(b^i, \overline{\boldsymbol{\pi}}_N^{-i}) - U_{\overline{\pi}}(q^i, b^i)\Big|$$

$$
\begin{aligned}
&= \Big|\mathbb{E}\Big[\sum\nolimits_{t=1}^{T} F(q^i,\sigma_t(i))\Big] - cTb^i - TF(q^i,H(b^i|\overline{\pi})) + cTb^i\Big| \\
&= \Big|\mathbb{E}\Big[\sum_{t=1}^{\tau} F(q^i,\sigma_t(i)) + (T-\tau)F(q^i,\overline{\sigma}(i))\Big] - TF(q^i,H(b^i|\overline{\pi}))\Big| \\
&\overset{(a)}{\leq} \Big|\sum_{t=1}^{\tau}(F(q^i,\sigma_t(i)) - \tau F(q^i,H(b^i|\overline{\pi}))\Big| + (T-\tau)\Big|\mathbb{E}\big[F(q^i,\overline{\sigma}(i))\big] - F(q^i,H(b^i|\overline{\pi}))\Big| \\
&\leq \tau\Delta_F + (T-\tau)\mathbb{E}\Big[\big|F(q^i,\overline{\sigma}(i)) - F(q^i,H(b^i|\overline{\pi}))\big|\Big] \\
&\leq \tau\Delta_F + (T-\tau)\|F_2\|_\infty\,\mathbb{E}\big[|\overline{\sigma}(i) - H(b^i|\overline{\pi})|\big] \\
&\overset{(b)}{\leq} \tau\Delta_F + (T-\tau)\|F_2\|_\infty\left(\tfrac{2c\tau\Delta_F}{T\delta_F^*} + \tfrac{2c\log T}{T\delta_F^*} + \exp(-\tfrac{\log^2 T}{4\tau\eta^2}) + \tfrac{1}{N-1}\right) \\
&\overset{(c)}{\leq} \left(\Delta_F + \tfrac{2c\|F_2\|_\infty\Delta_F}{\delta_F^*}\right)\tau + \tfrac{2c\|F_2\|_\infty}{\delta_F^*}\log T + \|F_2\|_\infty\exp(-\tfrac{\log^2 T}{4\tau\eta^2} + \log T) + \|F_2\|_\infty\tfrac{T}{N-1},
\end{aligned}
$$

where the inequality in (a) follows from the definition of ETC; in (b), we apply the results in the previous step; (c) follows from the fact that $\frac{T-\tau}{T} \leq 1$; the fourth component follows from the observation that $T - \tau \leq T = \exp(\log T)$. Notably, the bound above depends only on the problem primitives such as F, c, τ, T and N, but not on the agent i and inflation intensity b. Rearranging terms, we have

$$
\Big|v^i(b^i,\overline{\boldsymbol{\pi}}_N^{-i}) - U_{\overline{\pi}}(q^i,b^i)\Big| \;\leq\; \left(\Delta_F + \tfrac{2c\|F_2\|_\infty\Delta_F}{\delta_F^*}\right)\tfrac{\tau}{T} + \tfrac{2c\|F_2\|_\infty}{\delta_F^*}\tfrac{\log T}{T} + \|F_2\|_\infty\exp(-\tfrac{\log^2 T}{4\tau\eta^2}) + \|F_2\|_\infty\tfrac{1}{N-1}. \tag{EC.24}
$$

We have thus proved the first statement of the theorem by picking $\delta_1 = 2\left(\Delta_F + \frac{2c\|F_2\|_\infty\Delta_F}{\delta_F^*}\right)$, $\delta_2 = \frac{4c\|F_2\|_\infty}{\delta_F^*}$, $\delta_3 = 2\|F_2\|_\infty$, and $\delta_4 = 2\|F_2\|_\infty$.

Step 4: Quantify the benefit of deviation. It suffices to check deviations in the form of LSI. Building on the result in the previous step, we can bound the benefit of deviation from the NAE-implied intensity $\overline{\pi}(q^i)$:

$$
\begin{aligned}
& v^i(b^i,\overline{\boldsymbol{\pi}}_N^{-i}) - v^i(\overline{\pi}(q^i),\overline{\boldsymbol{\pi}}_N^{-i}) \\
=\;& v^i(b^i,\overline{\boldsymbol{\pi}}_N^{-i}) - U_{\overline{\pi}}(q^i,b^i) + U_{\overline{\pi}}(q^i,b^i) - U_{\overline{\pi}}(q^i,\overline{\pi}(q^i)) + U_{\overline{\pi}}(q^i,\overline{\pi}(q^i)) - v^i(\overline{\pi}(q^i),\overline{\boldsymbol{\pi}}_N^{-i}) \\
\overset{(a)}{\leq}\;& \Big(v^i(b^i,\overline{\boldsymbol{\pi}}_N^{-i}) - U_{\overline{\pi}}(q^i,b^i)\Big) + \Big(U_{\overline{\pi}}(q^i,\overline{\pi}(q^i)) - v^i(\overline{\pi}(q^i),\overline{\boldsymbol{\pi}}_N^{-i})\Big) \\
\overset{(b)}{\leq}\;& \delta_1\tfrac{\tau}{T} + \delta_2\tfrac{\log T}{T} + \delta_3\exp(-\tfrac{1}{4\eta^2}\tfrac{\log^2 T}{\tau}) + \delta_4\tfrac{1}{N-1},
\end{aligned} \tag{EC.25}
$$

where (a) holds because $U_{\overline{\pi}}(q^i,b^i) \leq U_{\overline{\pi}}(q^i,\overline{\pi}(q^i))$ under the NAE; in (b), we apply the uniform bound (EC.24) in the previous step twice for both components. ■

Proof of Theorem 5. Let the number of agents N, the experimentation length τ, and time horizon T be given. We first note that

$$
\begin{aligned}
\mathcal{R}_T \;&= \mathbb{E}\Big[\sum_{t=1}^{\tau}\Big(\sum_{i=1}^{N}F(q^i,z^i) - \sum_{i=1}^{N}F(q^i,\sigma_t(i))\Big) + \sum_{t=\tau+1}^{T}\Big(\sum_{i=1}^{N}F(q^i,z^i) - \sum_{i=1}^{N}F(q^i,\sigma_t(i))\Big)\Big] \\
&\leq \tau N\Delta_F + \sum_{t=\tau+1}^{T}\mathbb{E}\Big[\sum_{i=1}^{N}F(q^i,z^i) - \sum_{i=1}^{N}F(q^i,\sigma_t(i))\Big] \\
&= \tau N\Delta_F + (T-\tau)\underbrace{\sum_{i=1}^{N}\mathbb{E}\Big[F(q^i,H^N(\overline{\boldsymbol{\pi}}_N^i\,|\,\overline{\boldsymbol{\pi}}_N^{-i})) - F(q^i,\overline{\sigma}(i))\Big]}_{(*)},
\end{aligned}
$$

where recall that $H^N(\overline{\boldsymbol{\pi}}_N^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i})$ is agent i's inflation-intensity based ranking position under the inflation profile $\overline{\boldsymbol{\pi}}_N$. Since $\overline{\pi}$ is strictly increasing in quality, agent i obtains the *quality-based* position under this inflation profile, i.e., $H^N(\overline{\boldsymbol{\pi}}_N^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i}) = z^i$. At the same time $\overline{\sigma}$ is the ranking that ETC commits to in the second phase.

We will next provide two separate upper bounds on the regret based on two different methods of controlling the term $(*)$, one tighter when $N \gg T$, and the other tighter when $T \gg N$.

The first upper bound. Recall from (EC.23) in the proof of Theorem 4, we have

$$\mathbb{E}\Big[\big|\overline{\sigma}(i) - H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i})\big|\Big] \;\le\; \tfrac{2c(\tau\Delta_F + \log T)}{T\delta_F^*} + \exp\Big(-\tfrac{\log^2 T}{4\tau\eta^2}\Big),$$

which is a bound uniform in $i \in [N]$. Therefore, after rearranging the terms, we have

$$\begin{aligned}
\mathcal{R}_T \;&\le \tau N\Delta_F + (T-\tau)N\|F_2\|_\infty \mathbb{E}\Big[\big|\overline{\sigma}(i) - H^N(b^i \,|\, \overline{\boldsymbol{\pi}}_N^{-i})\big|\Big] \\
&\le \tau N\Delta_F + TN\|F_2\|_\infty \Big(\tfrac{2c(\tau\Delta_F+\log T)}{T\delta_F^*} + \exp(-\tfrac{\log^2 T}{4\tau\eta^2})\Big) \\
&= \Delta_F\tau N + \tfrac{2\Delta_F\tau\|F_2\|_\infty}{\delta_F^*}cN + \tfrac{2\|F_2\|_\infty}{\delta_F^*}cN\log T + \|F_2\|_\infty TN\exp(-\tfrac{\log^2 T}{4\tau\eta^2}) \\
&\le \Delta_F\tau N + \tfrac{2\Delta_F\tau\|F_2\|_\infty}{\delta_F^*}cN + \tfrac{2\|F_2\|_\infty}{\delta_F^*}cN\log T + \|F_2\|_\infty \exp(\tau\eta^2)N \\
&= \Big[\Delta_F\tau + \|F_2\|_\infty\exp(\tau\eta^2) + \tfrac{2\Delta_F\tau\|F_2\|_\infty}{\delta_F^*}c\Big]N + \tfrac{2\|F_2\|_\infty}{\delta_F^*}cN\log T \\
&= \phi_1 N + \phi_2 cN\log T,
\end{aligned}$$

where $\phi_1 := \Delta_F\tau + \|F_2\|_\infty\exp(\tau\eta^2) + \frac{2\Delta_F\tau\|F_2\|_\infty}{\delta_F^*}c$ and $\phi_2 := \frac{2\|F_2\|_\infty}{\delta_F^*}$; the last inequality is because the function $f(T) := TN\exp(-\frac{\log^2 T}{4\tau\eta^2})$ attains maximum of $\exp(\tau\eta^2)N$ when $\log T = 2\tau\eta^2$.

Second upper bound. Recall that $\Delta_F = F(1,1) - F(0,0)$ and $\delta_F^* = \min_{t\in[0,1]} F_2(t,t)$ are strictly positive constants that only depend on the reward function F. Let $\overline{\delta} := \frac{\delta_F^* T}{4cN}$ and $T_0 = \frac{4\Delta_F}{\delta_F^*}cN\tau$. In the event of $\{|\sum_{t=1}^\tau \epsilon_t^i| \le \overline{\delta},\, \forall\, i\in[N]\}$, we can verify that for all $T \ge \max\{T_0, \tau\}$ and $i > j$,

$$\begin{aligned}
&\widetilde{D}^i - \widetilde{D}^j \\
&= \sum_{t=1}^\tau \Big(F(q^i,\sigma_t(i)) + \epsilon_t^i - F(q^j,\sigma_t(j)) - \epsilon_t^j\Big) + T\overline{\pi}(q^i) - T\overline{\pi}(q^j) \\
&= \sum_{t=1}^\tau (\epsilon_t^i - \epsilon_t^j) + \sum_{t=1}^\tau \big[F(q^i,\sigma_t(i)) - F(q^j,\sigma_t(j))\big] + T\frac{1}{c}\int_{q^j}^{q^i} F_2(t,t)dt \\
&\ge -2\overline{\delta} - \tau\Delta_F + \tfrac{\delta_F^*}{cN}T \\
&= -\tfrac{\delta_F^*}{2cN}T - \tau\Delta_F + \tfrac{\delta_F^*}{cN}T \\
&= \tfrac{\delta_F^*}{2cN}T - \tau\Delta_F \\
&\ge \tfrac{\delta_F^*}{2cN}\cdot\tfrac{4\tau cN\Delta_F}{\delta_F^*} - \tau\Delta_F = \tau\Delta_F > 0,
\end{aligned}$$

where the second-to-last inequality holds because $T \ge T_0 = \frac{4\Delta_F}{\delta_F^*}cN\tau$.

Therefore, under ETC (with any initial ranking), and when $T \ge \max\{T_0, \tau\}$,

$$\mathbb{P}(\overline{\sigma} \ne \boldsymbol{e}) \le \mathbb{P}\Big\{|\sum_{t=1}^\tau \epsilon_t^i| \ge \overline{\delta},\ \text{for some } i\in[N]\Big\} \overset{(a)}{\le} 2N\exp\Big(-\tfrac{\overline{\delta}^2}{2\tau\eta^2}\Big) = 2N\exp\Big(-\tfrac{(\delta_F^*)^2}{32\eta^2}\cdot\tfrac{T^2}{c^2N^2\tau}\Big),$$

where part (a) follows from the generalized Hoeffding's inequality for sub-Gaussian variables. As a further consequence,

$$\mathcal{R}_T \;\le \tau N\Delta_F + (T-\tau)N\Delta_F\mathbb{P}(\overline{\sigma}\ne\boldsymbol{e}) \;\le\; \tau N\Delta_F + 2TN^2\Delta_F\exp\Big(-\tfrac{(\delta_F^*)^2}{32\eta^2}\cdot\tfrac{T^2}{c^2N^2\tau}\Big).$$

When $T \le \max\{T_0, \tau\}$, the regret can be simply bounded by

$$\mathcal{R}_T \;\le\; \max\{T_0,\tau\}N\Delta_F \;=\; \max\{\tfrac{4\Delta_F^2}{\delta_F^*}cN^2\tau, \tau N\Delta_F\}.$$

Combining two cases, it holds that for all $T \ge 1$,

$$\begin{aligned}
\mathcal{R}_T \;&\le\; \max\Big\{\tau N\Delta_F + 2TN^2\Delta_F \exp\Big(-\tfrac{(\delta_F^*)^2}{32\eta^2}\cdot\tfrac{T^2}{c^2N^2\tau}\Big),\; \tfrac{4\Delta_F^2}{\delta_F^*}cN^2\tau,\; \tau N\Delta_F\Big\} \\
&=\; \max\Big\{\tau N\Delta_F + 2TN^2\Delta_F \exp\Big(-\tfrac{(\delta_F^*)^2}{32\eta^2}\cdot\tfrac{T^2}{c^2N^2\tau}\Big),\; \tfrac{4\Delta_F^2}{\delta_F^*}cN^2\tau\Big\} \\
&\le \tau N\Delta_F + 2TN^2\Delta_F \exp\Big(-\tfrac{(\delta_F^*)^2}{32\eta^2}\cdot\tfrac{T^2}{c^2N^2\tau}\Big) + \tfrac{4\Delta_F^2}{\delta_F^*}cN^2\tau \\
&= \tau\Delta_F N + \tfrac{4\Delta_F^2}{\delta_F^*}c\tau N^2 + 2TN^2\Delta_F \exp\Big(-\tfrac{(\delta_F^*)^2}{32\eta^2}\cdot\tfrac{T^2}{c^2N^2\tau}\Big) \\
&\overset{(a)}{\le} \tau\Delta_F N + \tfrac{4\Delta_F^2}{\delta_F^*}c\tau N^2 + \tfrac{8\eta\Delta_F\sqrt{\tau}}{\sqrt{e}\delta_F^*}cN^3 \\
&\le \Big[\tau\Delta_F + \big(\tfrac{4\Delta_F^2}{\delta_F^*}\tau + \tfrac{8\eta\Delta_F\sqrt{\tau}}{\sqrt{e}\delta_F^*}\big)c\Big]N^3 \\
&= \phi_3 N^3
\end{aligned}$$

where $\phi_3 := \tau\Delta_F + \big(\frac{4\Delta_F^2}{\delta_F^*}\tau + \frac{8\eta\Delta_F\sqrt{\tau}}{\sqrt{e}\delta_F^*}\big)c$, and part (a) above is because the function $T \mapsto 2TN^2\Delta_F \exp\Big(-\frac{(\delta_F^*)^2}{32\eta^2}\cdot\frac{T^2}{c^2N^2\tau}\Big)$ attains maximum at $\frac{8\eta\Delta_F\sqrt{\tau}}{\sqrt{e}\delta_F^*}cN^3$ when $T = \frac{4\,\eta\, c\, N\sqrt{\tau}}{\delta_F^*}$. Combining both upper bounds, we have

$$\mathcal{R}_T \le \min\{\phi_1 N + \phi_2 cN\log T, \phi_3 N^3\}.$$

■

Corollary EC.2. *Consider the ETC policy with $\tau = o(\log T)$. Suppose further that the function $F(q,z)$ is strictly supermodular. Then for every agent i, $\overline{\boldsymbol{\pi}}_N^i$ is close to his best response to $\overline{\boldsymbol{\pi}}_N^{-i}$. That is,*

$$\sup_{i,\; b_*^i \in \arg\max_{b\ge 0} v^i(b, \overline{\boldsymbol{\pi}}_N^{-i})} |b_*^i - \overline{\boldsymbol{\pi}}_N^i| \to 0 \qquad \text{as } T, N \to \infty.$$

Proof of Corollary EC.2. Suppose towards a contradiction there exists $\varepsilon > 0$ and a sequence $\{N_k, T_k, i_k\}_k$ such that (i) $N_k, T_k \to \infty$ as $k \to \infty$; and (ii) for every k, there exists $b_*^{N_k,T_k,i_k} \in \arg\max_{b\ge 0} v^{i_k}(b, \overline{\boldsymbol{\pi}}_N^{-i_k})$ as a best response, but $|b_*^{N_k,T_k,i_k} - \overline{\pi}(q^{i_k})| \ge \varepsilon$. We show that $\{b_*^{N_k,T_k,i_k}\}$ is a bounded sequence. In fact, let $\overline{b} := \max_{(q,z)\in[0,1]^2} F(q,z)/c$, which is finite because F is continuous in $[0,1]^2$. For every N, T, $i \in [N]$, $b^i > \overline{b}$ and $\boldsymbol{b}^{-i} \in \mathbb{R}_+^{N-1}$, agent i's utility is such that $v^i(b^i, \boldsymbol{b}^{-i}) = \mathbb{E}_{\boldsymbol{\Pi}_T \boldsymbol{b}}\big[\sum_{t=1}^T F(q^i, \sigma_t(i))\big]/T - cb^i \le \max_{(q,z)\in[0,1]^2} F(q,z) - cb^i < 0$, which implies that it is not optimal for agent i to choose the intensity $b^i > \overline{b}$. Therefore, $\{b_*^{N_k,T_k,i_k}\}$ is bounded in $[0,\overline{b}]$. In addition, $\{q^{i_k}\}$ is bounded in the unit interval $[0,1]$. Therefore, both $\{b_*^{N_k,T_k,i_k}\}$ and $\{q^{i_k}\}$ have convergent subsequences. After proper re-indexing, suppose that $b_*^{N_k,T_k,i_k} \to b'$ and $q^{i_k} \to q'$ as $k \to \infty$ without loss of generality. That implies that $|b_*^{N_k,T_k,i_k} - \overline{\pi}(q^{i_k})| \to |b' - \overline{\pi}(q')|$, and therefore, $|b' - \overline{\pi}(q')| \ge \varepsilon$.

At the same time, Theorem 4 implies that $\sup_b |v^{i_k}(b, \overline{\boldsymbol{\pi}}_{N_k}^{-i_k}) - U_{\overline{\pi}}(q^{i_k}, b)| \to 0$. Moreover, the function $U_{\overline{\pi}}(q,b) = F(q, \int_0^q \frac{F_2(t,t)}{c}dt) - cb$ is uniformly continuous due to the smoothness of F; see Assumption 2. Therefore, $\sup_b |U_{\overline{\pi}}(q^{i_k}, b) - U_{\overline{\pi}}(q', b)| \to 0$. As a result, $\sup_b |v^{i_k}(b, \overline{\boldsymbol{\pi}}_{N_k}^{-i_k}) - U_{\overline{\pi}}(q', b)| \to 0$. In other words, the function $v^{i_k}(\cdot, \overline{\boldsymbol{\pi}}_{N_k}^{-i_k})$ uniformly converges to the function $U_{\overline{\pi}}(q', \cdot)$ as $k \to \infty$. Along with the facts that $b_*^{N_k,T_k,i_k} \to b'$ and $b_*^{N_k,T_k,i_k} \in \arg\max_{b\ge 0} v^{i_k}(b, \overline{\boldsymbol{\pi}}_N^{-i_k})$, we conclude that b' is also maximizer of the function $U_{\overline{\pi}}(q', \cdot)$. Since $F(q,z)$ is strictly supermodular, and given that $\frac{\partial}{\partial b}U_{\overline{\pi}}(q', b) = c\frac{F_2(q', \overline{\pi}^{-1}(b))}{F_2(\overline{\pi}^{-1}(b), \overline{\pi}^{-1}(b))} - c$ from the proof of Theorem 3, it holds that $\overline{\pi}(q')$ is the unique maximizer of the function $U_{\overline{\pi}}(q', \cdot)$. Consequently, $b' = \overline{\pi}(q')$, which contradicts the fact that $|b' - \overline{\pi}(q')| \ge \varepsilon$. ■

E. Equilibrium Analysis for Finite N (Section 6.1) Proof of Proposition 5. Fix T and an even integer τ as the proposition statement requires. Let $w := F(1,1)+F(1,0)-F(0,1)-F(0,0) > 0$ by the monotonicity of F. When the action profile of the agents is (b^1, b^2), note that the observed signal in the experimentation stage satisfies $\widetilde{D}^2 - \widetilde{D}^1 = T(b^2-b^1) + \frac{1}{2}w\tau + \sum_{\ell=1}^{\tau}(\epsilon_\ell^2 - \epsilon_\ell^1)$.[13] Note in addition that $\sum_{\ell=1}^{\tau}(\epsilon_\ell^2 - \epsilon_\ell^1)$ follows a zero-mean normal distribution with variance $2\tau\eta^2$. Hence we denote the probability density function for $\sum_{\ell=1}^{\tau}(\epsilon_\ell^2 - \epsilon_\ell^1)$ as φ and the corresponding distribution function as Φ. Moreover, for shorthand notation, given $i \in \{1,2\}$, we let

$$v_C^i(x) := \tfrac{1}{T}\mathbb{E}\big[\sum_{\ell=\tau+1}^{T} F(q^i, \overline{\sigma}(i)) \,|\, \widetilde{D}^2 - \widetilde{D}^1 = x\big]$$

be agent i's expected continuation value in the commitment phase (scaled by T) given the observed signals in the experimentation phase. Under ETC, these continuation values admit the following closed forms:

$$v_C^1(x) = \begin{cases} \frac{T-\tau}{T}F(0,0) & \text{if } x>0 \\ \frac{T-\tau}{T}F(0,1) & \text{if } x<0 \end{cases} \quad \text{and} \quad v_C^2(x) = \begin{cases} \frac{T-\tau}{T}F(1,1) & \text{if } x>0 \\ \frac{T-\tau}{T}F(1,0) & \text{if } x<0. \end{cases}$$

As a result, for any $b^i \geq 0$ with $i \in \{1,2\}$, the expected utility for agent i can be expressed as

$$\begin{aligned} v^i(b^i, b^{-i}) &= \tfrac{\tau}{2T}[F(q^i,1)+F(q^i,0)] + \int_{-\infty}^{\infty} v_C^i\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau + x\Big)\varphi(x)\,dx - cb^i \\ &\overset{(a)}{=} \tfrac{\tau}{2T}[F(q^i,1)+F(q^i,0)] + \int_{-\infty}^{\infty} v_C^i\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau + x\Big)\varphi(-x)\,dx - cb^i \\ &= \tfrac{\tau}{2T}[F(q^i,1)+F(q^i,0)] + \int_{-\infty}^{\infty} v_C^i\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau - x\Big)\varphi(x)\,dx - cb^i \\ &\overset{(b)}{=} \tfrac{\tau}{2T}[F(q^i,1)+F(q^i,0)] + \int_{-\infty}^{\infty} v_C^i(x)\varphi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau - x\Big)\,dx - cb^i, \end{aligned} \tag{EC.26}$$

where step (a) follows from the fact that $\varphi(x) = \varphi(-x)$. In Step (b), the second component holds by the commutativity property of convolution, i.e., $\int_{-\infty}^{\infty} v_C^i(T(b^2-b^1)+\frac{1}{2}w\tau - x)\varphi(x)dx = \int_{-\infty}^{\infty} v_C^i(x)\varphi(T(b^2-b^1)+\frac{1}{2}w\tau - x)dx$. In addition, with the expressions of v_C^i, we may further derive that

$$\begin{aligned} &v^1(b^1,b^2) \\ &= \tfrac{\tau}{2T}[F(0,1)+F(0,0)] \\ &\quad + \int_{-\infty}^{0} \tfrac{T-\tau}{T}F(0,1)\varphi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau - x\Big)\,dx + \int_{0}^{\infty} \tfrac{T-\tau}{T}F(0,0)\varphi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau - x\Big)\,dx - cb^1 \\ &\overset{(a)}{=} \tfrac{\tau}{2T}[F(0,1)+F(0,0)] + \tfrac{T-\tau}{T}F(0,1)\Big(1-\Phi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau\Big)\Big) + \tfrac{T-\tau}{T}F(0,0)\Phi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau\Big) - cb^1 \\ &= \tfrac{\tau[F(0,1)+F(0,0)]/2+(T-\tau)F(0,1)}{T} - \tfrac{T-\tau}{T}[F(0,1)-F(0,0)]\Phi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau\Big) - cb^1. \end{aligned}$$

Similarly,

$$v^2(b^2,b^1) = \tfrac{\tau[F(1,1)+F(1,0)]/2+(T-\tau)F(1,0)}{T} + \tfrac{T-\tau}{T}[F(1,1)-F(1,0)]\Phi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau\Big) - cb^2,$$

With both agents' utility functions in closed form, let us prove all the claims below.

[13] We would also like to remind the reader that the superscripts throughout the proof, such as q^i, b^i and $\widetilde{D}^i$ all refer to the agent index rather than the power function.

Claim (i). It holds that

$$\begin{aligned}\tfrac{\partial v^1(b^1,b^2)}{\partial b^1} &= (T-\tau)[F(0,1)-F(0,0)]\varphi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau\Big)-c;\\ \tfrac{\partial v^2(b^2,b^1)}{\partial b^2} &= (T-\tau)[F(1,1)-F(1,0)]\varphi\Big(T(b^2-b^1)+\tfrac{1}{2}w\tau\Big)-c.\end{aligned}$$

Since F is strictly supermodular, we have $\frac{\partial v^1(b^1,b^2)}{\partial b^1} < \frac{\partial v^2(b^2,b^1)}{\partial b^2}$ for all $b^1,b^2\in\mathbb{R}_+$. Intuitively, that means the high-type agent's marginal benefit of inflation is always strictly larger than that of the low-type.

This observation rules out two types of Nash equilibrium. First, the first-order condition for any Nash equilibrium in the interior (i.e., $b^1>0$ and $b^2>0$) requires that $\frac{\partial v^1(b^1,b^2)}{\partial b^1}=\frac{\partial v^2(b^2,b^1)}{\partial b^2}=0$, which is impossible. Similarly, the first-order condition for any pure-strategy Nash equilibrium with $b^1>0$ and $b^2=0$ requires $\frac{\partial v^1(b^1,b^2)}{\partial b^1}=0$ yet $\frac{\partial v^2(b^2,b^1)}{\partial b^2}\le 0$, which is also impossible.

Claim (ii). Fix $b^1=0$. Pick $(T-\tau)F(1,1)\varphi(0)=(T-\tau)F(1,1)\frac{1}{\eta\sqrt{2\tau}}<c$. This happens when T is sufficiently small, η is sufficiently large, or c is sufficiently large. Then

$$\tfrac{\partial v^2(b^2,0)}{\partial b^2} \;=\; (T-\tau)[F(1,1)-F(1,0)]\varphi\Big(Tb^2+\tfrac{1}{2}w\tau\Big)-c \;\le\; (T-\tau)F(1,1)\varphi(0)-c<0.$$

Thus, the marginal benefit of inflation is negative for any $b^2\ge 0$, and we can conclude that the high-quality agent would not inflate when $b^1=0$. Similarly,

$$\tfrac{\partial v^1(b^1,0)}{\partial b^1} = (T-\tau)[F(0,1)-F(0,0)]\varphi\Big(-Tb^1+\tfrac{1}{2}w\tau\Big)-c \;\le\; (T-\tau)F(1,1)\varphi(0)-c<0.$$

Hence the manipulation-free profile is a Nash equilibrium.

Claim (iii). In light of Claim (i), it is left to be seen that there is no pure-strategy Nash equilibrium (b^1,b^2) where $b^1=0$. Now, fix $b^1=0$. It follows that

$$\tfrac{\partial v^2(b^2,0)}{\partial b^2} = (T-\tau)[F(1,1)-F(1,0)]\varphi\Big(Tb^2+\tfrac{1}{2}w\tau\Big)-c$$

is a decreasing function in b^2. In addition, $\frac{\partial v^2(b^2,0)}{\partial b^2}\Big|_{b^2=0} = (T-\tau)[F(1,1)-F(1,0)]\varphi\Big(\frac{1}{2}w\tau\Big)-c$. We prove the first two claims in Case 1 and the last claim in Case 2.

Case 1: c is small or T is large (for fixed η). Fix η and suppose T is sufficiently large or c is sufficiently small such that $(T-\tau)[F(1,1)-F(1,0)]\varphi\Big(\frac{1}{2}w\tau\Big)>c$. As a result, the function $v^2(\cdot,0)$ first increases and then decreases. Hence, agent 2's best response to $b^1=0$, denoted by b^2_*, satisfies

$$(T-\tau)[F(1,1)-F(1,0)]\varphi\Big(Tb^2_*+\tfrac{1}{2}w\tau\Big)-c=0.$$

Recall that φ is the pdf for a zero-mean normal distribution with variance $2\tau\eta^2$, i.e., $\varphi(x)=\frac{1}{\eta\sqrt{4\pi\tau}}\exp\Big(-\frac{x^2}{4\tau\eta^2}\Big)$. Solving this equation gives

$$\varphi\Big(Tb^2_*+\tfrac{1}{2}w\tau\Big)=\tfrac{c}{(T-\tau)[F(1,1)-F(1,0)]} \implies Tb^2_*+\tfrac{1}{2}w\tau=2\,\eta\sqrt{\tau}\sqrt{\log\Big(\tfrac{(T-\tau)[F(1,1)-F(1,0)]}{c\,\eta\sqrt{4\pi\tau}}\Big)}=\Theta(\sqrt{\log\tfrac{T}{c}}).$$

However, we next show that agent 1's best response to b^2_* is not $b^1=0$. To see why, note that

$$\begin{aligned}&v^1(\varepsilon,b^2_*)-v^1(0,b^2_*)\\ &=\tfrac{T-\tau}{T}[F(0,1)-F(0,0)]\Big[\Phi\big(Tb^2_*+\tfrac{1}{2}w\tau\big)-\Phi\big(T(b^2_*-\varepsilon)+\tfrac{1}{2}w\tau\big)\Big]-c\varepsilon\\ &=\tfrac{T-\tau}{T}[F(0,1)-F(0,0)]\Big[\Phi_*\Big(\sqrt{\log\Big(\tfrac{(T-\tau)[F(1,1)-F(1,0)]}{c\,\eta\sqrt{4\pi\tau}}\Big)}\Big)-\Phi_*\Big(\sqrt{\log\Big(\tfrac{(T-\tau)[F(1,1)-F(1,0)]}{c\,\eta\sqrt{4\pi\tau}}\Big)}-\tfrac{T\varepsilon}{2\eta\sqrt{\tau}}\Big)\Big]-c\varepsilon,\end{aligned}$$

where Φ_* is the cdf for standard normal. Now, it can be clearly seen that as $T\to\infty$ or $c\to 0$, $\Phi_*\Big(\sqrt{\log\Big(\frac{(T-\tau)[F(1,1)-F(1,0)]}{c\eta\sqrt{4\pi\tau}}\Big)}\Big)\to 1$ and $\Phi_*\Big(\sqrt{\log\Big(\frac{(T-\tau)[F(1,1)-F(1,0)]}{c\eta\sqrt{4\pi\tau}}\Big)}-\frac{T\varepsilon}{2\eta\sqrt{\tau}}\Big)\to 0$. Hence $v^1(\varepsilon,b^2_*)-v^1(0,b^2_*)>0$ for sufficiently large T or sufficiently small c as long as ε is picked to be a small quantity independent of T, e.g., $\varepsilon=0.001[F(0,1)-F(0,0)]/c$.

Case 2: η is small (for fixed large T or small c). First, suppose T and c are properly fixed so that $(T-\tau)[F(0,1)-F(0,0)]-cw\tau>0$, and let η be sufficiently small so that $(T-\tau)[F(1,1)-F(1,0)]\varphi\Big(\frac{1}{2}w\tau\Big)-c<0$. As a result, the function $v^2(\cdot,0)$ decreases. Hence, agent 2's best response to $b^1=0$, denoted by b^2_*, satisfies $b^2_*=0$. However, we next show that agent 1's best response to $b^2_*=0$ is not $b^1=0$. To see why, note that

$$\begin{aligned}
&v^1(\tfrac{w\tau}{T},b^2_*)-v^1(0,b^2_*)\\
&=\tfrac{T-\tau}{T}[F(0,1)-F(0,0)]\Big[\Phi\big(\tfrac{1}{2}w\tau\big)-\Phi\big(-T\cdot\tfrac{w\tau}{T}+\tfrac{1}{2}w\tau\big)\Big]-c\tfrac{w\tau}{T}\\
&=\tfrac{T-\tau}{T}[F(0,1)-F(0,0)]\Big[\Phi_*\big(\tfrac{1}{2\sqrt{2\tau\eta}}w\tau\big)-\Phi_*\big(-\tfrac{1}{2\sqrt{2\tau\eta}}w\tau\big)\Big]-c\tfrac{w\tau}{T},
\end{aligned}$$

where Φ_* is the cdf for standard normal. Now, it can be seen that as $\eta\to 0$, $v^1(\frac{w\tau}{T},b^2_*)-v^1(0,b^2_*)\to\frac{T-\tau}{T}[F(0,1)-F(0,0)]-c\frac{w\tau}{T}>0$, which is strictly positive based on the initial choice of T and c. ■

Proof of Proposition 6. To prove the claim, we suppose towards a contradiction that there exists a pure-strategy Nash equilibrium in $\overline{\Gamma}_{N,\infty}$, which we denote by $\boldsymbol{b}=(b^1,...,b^N)$. We separately discuss two cases.

Case 1: The inflation profile $\boldsymbol{b}$ has ties. Fixing $i\in[N]$, an inflation tie means that there exists $j\neq i$ such that $b^i=b^j$. Recall from (5) and (6) that the agent i's utility satisfies $\overline{v}^i(b^i,\boldsymbol{b}^{-i})=F\big(q^i,H^N(b^i\,|\,\boldsymbol{b}^{-i})\big)-cb^i$, where $H^N(b^i\,|\,\boldsymbol{b}^{-i})=\frac{1}{N-1}\sum_{k\in[N]\setminus\{i\}}\mathbb{I}\{b^k<b^i\}$. Note that there is a strictly positive jump of $H^N(b^i\,|\,\boldsymbol{b}^{-i})$ at the value of b^i. In other words,

$$\lim_{\rho\to 0+}H^N(b^i+\rho\,|\,\boldsymbol{b}^{-i})-H^N(b^i\,|\,\boldsymbol{b}^{-i})\geq\tfrac{1}{N-1}>0.$$

This is because agent i is "boosted" by one position by out-inflating agent j by ρ given that $b^i=b^j$. Therefore, due to Assumption (i), we have

$$\lim_{\rho\to 0+}\overline{v}^i(b^i+\rho,\boldsymbol{b}^{-i})-\overline{v}^i(b^i,\boldsymbol{b}^{-i})>0.$$

Therefore, we can construct a strictly profitable deviation for agent i. There exists a small number $\rho>0$ so that the utility of agent i with action $b^i+\rho$ satisfies

$$\overline{v}^i(b^i+\rho,\boldsymbol{b}^{-i})\;>\;\overline{v}^i(b^i,\boldsymbol{b}^{-i})-c\rho,$$

In other words, agent i finds it strictly beneficial to deviate from action b^i to $b^i+\rho$. But that contradicts the assumption that $\boldsymbol{b}$ is a pure-strategy equilibrium.

Case 2: The inflation profile $\boldsymbol{b}$ does not have ties. Suppose $\boldsymbol{b}$ does not have ties. Let $i,j\in[N]$ be two agents with adjacent ranking positions, i.e. $H^N(b^j|\boldsymbol{b}^{-j})=H^N(b^i|\boldsymbol{b}^{-i})+\frac{1}{N-1}$. Because of the monotonicity of H^N and the fact that $\boldsymbol{b}$ does not have ties, it holds that $b^i<b^j$. Therefore, for sufficiently small $\varepsilon<(b^j-b^i)/100$, we have $H^N(b^j|\boldsymbol{b}^{-j})=H^N(b^j-\varepsilon|\boldsymbol{b}^{-j})$. In other words, agent j keeps the ranking position by slightly inflating less. Therefore, the utility of agent j with action $b^j-\varepsilon$ satisfies

$$\overline{v}^j(b^j-\varepsilon,\boldsymbol{b}^{-j})=F\big(q^j,H^N(b^j-\varepsilon\,|\,\boldsymbol{b}^{-j})\big)-c(b^j-\varepsilon)=\overline{v}^j(b^j,\boldsymbol{b}^{-j})+c\varepsilon>\overline{v}^j(b^j,\boldsymbol{b}^{-j}).$$

In other words, agent j finds it strictly beneficial to deviate from action b^j to $b^j-\varepsilon$. However, this contradicts the assumption that $\boldsymbol{b}$ is a pure-strategy equilibrium. ■

F. Generalization to the Cascade Model (Section 6.2).

Proof of Proposition 7. Fix an arbitrary agent $i \in [N]$, time period $t \in [T]$, history $\boldsymbol{h}_{t-1} = \{\boldsymbol{B}, \sigma_1, \boldsymbol{D}_1, \boldsymbol{b}_1, \ldots, \sigma_{t-1}, \boldsymbol{D}_{t-1}, \boldsymbol{b}_{t-1}\}$, and continuation strategy profiles $\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})$ for all agents. The continuation value of agent i from period t onward is

$$\mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[\sum_{\ell=t}^{T} R_\ell(q^i, \sigma_\ell) \mid \boldsymbol{h}_{t-1}\right] = \begin{cases} \sum_{\ell=t}^{\tau} F(q^i, \sigma_\ell) + (T-\tau)\,\mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[F(q^i, \overline{\sigma}) \mid \boldsymbol{h}_{t-1}\right] & \text{if } t \le \tau, \\ (T-t+1)\,F(q^i, \overline{\sigma}) & \text{if } t \ge \tau+1. \end{cases} \tag{EC.27}$$

For $t \ge \tau+1$, note that the committed ranking $\overline{\sigma}$ is fully determined by history $\boldsymbol{h}_\tau$. Hence, agent i's continuation value is constant and independent of the continuation inflation policy.

For $t \le \tau$, the first term in (EC.27), $\sum_{\ell=t}^{\tau} F(q^i, \sigma_\ell)$, is independent of the agent's strategy $\Pi^i(h^i_{t-1})$, since the rankings $\{\sigma_t : t = 1, \ldots, \tau\}$ are determined by a fixed schedule. Hence, it suffices to focus on the second term, $\mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[F(q^i, \overline{\sigma}) \mid \boldsymbol{h}_{t-1}\right]$. Under the cascade model, we have:

$$F(q^i, \overline{\sigma}) = \lambda N r(q^i) \prod_{j:\overline{\sigma}(j) > \overline{\sigma}(i)} \left[(1 - r(q^j))(1-\zeta)\right],$$

where $\overline{\sigma}(i)$ is given by

$$\overline{\sigma}(i) = \tfrac{1}{N-1} \sum_{j \ne i} \mathbb{I}\left\{\widetilde{D}^i > \widetilde{D}^j\right\} = \tfrac{1}{N-1} \sum_{j \ne i} \mathbb{I}\left\{ \underbrace{D^i + \sum_{\ell=1}^{t-1} b^i_\ell}_{A} + \sum_{\ell=t}^{\tau} b^i_\ell > \underbrace{D^j + \sum_{\ell=1}^{\tau} b^j_\ell}_{B} \right\}.$$

Following the same reasoning as in the proof of Proposition 4, both components A and B are independent of agent i's continuation strategy. Therefore, increasing the continuation inflation amount $\sum_{\ell=t}^{\tau} b^i_\ell$ increases $\overline{\sigma}(i)$, decreases $\overline{\sigma}(j)$ for every $j \ne i$, and thereby increases $F(q^i, \overline{\sigma})$. Consequently, the conditional expectation $\mathbb{E}_{\boldsymbol{\Pi}(\boldsymbol{h}_{t-1})}\left[F(q^i, \overline{\sigma}) \mid \boldsymbol{h}_{t-1}\right]$ is maximized when the entire remaining budget is used up during the experimentation phase, which is precisely the behavior under the LSI strategy. ■

Proof of Theorem 6. For notational convenience, let us define the function:

$$\tilde{F}(q, z) := \lambda q \exp\left(\frac{z^2}{2} + z - \frac{3}{2}\right),$$

so that the nonatomic utility function under the cascade model, as given in (10), can be written as

$$U_\pi(q, b) = \tilde{F}(q, H(b \mid \pi)) - cb.$$

Intuitively, $\tilde{F}(q, z)$ represents the reward for a type-q agent when he attains a ranking position z by *unilaterally* adjusting his inflation intensity, while all other agents follow the inflation profile π. Note that $\tilde{F}(q, z)$ depends only on the agent's own quality q and the resulting position z even though we are working under the cascade model. That is because when a single agent deviates unilaterally, the ranking positions of all other agents are determined by the fixed profile π. This makes the mapping from the agent's position to his reward well-defined in the nonatomic game.

This observation allows us to reduce the verification of NAE to establishing three properties: (1) the equality $\tilde{\pi}(q) = \int_0^q \frac{\tilde{F}_2(t,t)}{c}\, dt$; (2) the strict supermodularity and monotonicity of $\tilde{F}$; and (3) the strict monotonicity of $\tilde{\pi}$. Once these are established, we can apply the same reasoning as in the proof of Theorem 3 to conclude that $\tilde{\pi}$ constitutes an NAE under the cascade model.

To verify properties (1) and (2), note that $\tilde{F}_1(q,z) = \lambda \cdot \exp\left(\frac{z^2}{2} + z - \frac{3}{2}\right) > 0$, $\tilde{F}_2(q,z) = \lambda q(z+1) \cdot \exp\left(\frac{z^2}{2} + z - \frac{3}{2}\right) \geq 0$, and $\frac{\partial^2 \tilde{F}}{\partial q \partial z}(q,z) = \lambda(z+1) \cdot \exp\left(\frac{z^2}{2} + z - \frac{3}{2}\right) > 0$ for $q, z \in [0,1]$.

Given those expressions, the proposed strategy in the theorem statement satisfies $\tilde{\pi}(q) = \int_0^q \frac{\tilde{F}_2(t,t)}{c}\, dt$. Moreover, note that $\tilde{F}_2(q,z) = 0$ if and only if $q = 0$, and the cross derivative $\frac{\partial^2 \tilde{F}}{\partial q \partial z}(q,z)$ is strictly positive in $[0,1]^2$. Therefore, the function $\tilde{F}(q,z)$ is supermodular and strictly increasing over $[0,1]^2$. Lastly, to establish item (3), simply observe that $\tilde{\pi}'(q) = \frac{\lambda}{c} q(q+1) \cdot \exp\left(\frac{q^2}{2} + q - \frac{3}{2}\right) > 0$ for all $q \in (0,1]$. Thus, $\tilde{\pi}$ is strictly increasing and continuously differentiable on $(0,1]$. ■

Proof of Theorem 7. Let us fix T and N as required in the theorem statement. We also introduce a few notations specific to the cascade model. First, similar to the base model, we define the interim long-run limiting utility function under the cascade model:

$$\overline{v}^i(b, \boldsymbol{b}^{-i}) \;:=\; F\big(q^i, H^N(b, \boldsymbol{b}^{-i})\big) - cb \;=\; \lambda q^i \prod_{H_j^N(b,\boldsymbol{b}^{-i}) > H_i^N(b,\boldsymbol{b}^{-i})} \big((1 - r(q^j))\,(1 - \tfrac{1}{N})\big) - cb,$$

where $H^N(b^i, \boldsymbol{b}^{-i})$ is the inflation-intensity based ranking for N agents as defined in (5). Intuitively, $\overline{v}^i(b, \boldsymbol{b}^{-i})$ represents agent i's long-term average utility from inflation intensity b while others play $\boldsymbol{b}^{-i}$. In the proof, it will be used as a bridge between $v^i(b^i, \tilde{\boldsymbol{\pi}}_N^{-i})$ and $U_{\tilde{\pi}}(q^i, b^i)$. We also denote by $C_{\text{us}} \sim \text{Poisson}(\lambda\tau N)$, the total number of customers arriving during the experimentation phase (who may potentially quit without purchasing).

Now, fix $i \in [N]$. Let $\{x_\ell\}_{\ell=1}^\infty$ be a sequence of i.i.d. Bernoulli random variables with mean $r(q^i) = q^i/N$, where x_ℓ represents the ℓth customer's purchase decision who has viewed product i. We also denote agent i's total reward during the experimentation phase as $R^i := \sum_{t=1}^\tau R_t^i$. We derive below an upper bound on the probability that R^i exceeds $2\lambda\tau$:

$$\begin{aligned}
\mathbb{P}\Big\{R^i \geq 2\lambda\tau\Big\} \;&\leq\; \mathbb{P}\Big\{R^i \geq \lambda\tau(1+q^i)\Big\} \\
&\overset{(a)}{\leq}\; \mathbb{P}\left\{\sum_{\ell=1}^{C_{\text{us}}} x_\ell \geq \lambda\tau(1+q^i)\right\} \\
&\overset{(b)}{\leq}\; \exp\left(-\frac{(\lambda\tau)^2}{2\lambda\tau(q^i + 1/3)}\right) \\
&\overset{(c)}{\leq}\; \exp\left(-\frac{3}{8}\lambda\tau\right), \qquad \text{(EC.28)}
\end{aligned}$$

where in step (a), we employ a (path-wise) upper bound of $R^i \leq \sum_{\ell=1}^{C_{\text{us}}} x_\ell$. This is due to the setup of the cascade model: the number of customers who view product i during the experimentation phase is at most C_{us}, and conditional on viewing, each of them making independent purchase decisions represented by $\{x_\ell\}$. In step (b), the sum $\sum_{\ell=1}^{C_{\text{us}}} x_\ell$ is compound Poisson, where the Poisson variable C_{us} and the Bernoulli sequence $\{x_\ell\}$ are independent of each other. Therefore, it follows Poisson distribution with mean $\mathbb{E}[C_{\text{us}}] \cdot q^i/N = \lambda\tau N \cdot q^i/N = \lambda\tau q^i$. We then apply the Bernstein-type inequality for Poisson random variables; see Section 2.2 in Boucheron et al. (2013). Step (c) is because $q^i \in [0,1]$.

Now, further fix $b^i \geq 0$. Also pick $\tau = 3\log(N^2T)/\lambda$. We break the rest of the proofs into the following steps.

Step 1: Quantify the proximity between $v^i(b^i, \tilde{\boldsymbol{\pi}}_N^{-i})$ and $\overline{v}^i(b^i, \tilde{\boldsymbol{\pi}}_N^{-i})$. As an overview of the idea, recall that the learner locks in the ranking $\overline{\sigma}$ during the commitment phase based all agents' observed signals $\{\tilde{D}_i\}$. At the same time, the inflation-intensity based ranking is given by $H^N(b^i, \tilde{\boldsymbol{\pi}}_N^{-i})$. As in the base model, the main idea is to bound the discrepancy between $\overline{\sigma}$ and $H^N(b^i, \tilde{\boldsymbol{\pi}}_N^{-i})$, and that of the resulting reward function value.

We further divide the proof arguments of this step into substeps, in which Step 1-1 and Step 1-2 provide auxiliary results to support the proof of the claim in Step 1-3.

Step 1-1. We show that the relative order of agents i and j in terms of their inflation intensities (i.e., b^i vs. $\tilde{\pi}(q^j)$) is consistent with that in terms of the observed signal (i.e., $\widetilde{D}^i$ vs. $\widetilde{D}^j$) with high probability, as long as agent j is "far away" from agent i. To be precise, let $s_\tau := (2\lambda\tau)/T$ for shorthand notation. Consider any agent j such that $|b^i - \tilde{\pi}(q^j)| > s_\tau$. Suppose $b^i - \tilde{\pi}(q^j) > s_\tau$, it holds that

$$\begin{aligned}\mathbb{P}(\widetilde{D}^j \geq \widetilde{D}^i) &= \mathbb{P}\Big\{R^j + T\tilde{\pi}(q^j) \geq R^i + Tb^i\Big\}\\ &\leq \mathbb{P}\Big\{R^j \geq T(b^i - \tilde{\pi}(q^j))\Big\}\\ &\overset{(a)}{\leq} \mathbb{P}\Big\{R^j \geq 2\lambda\tau\Big\}\\ &\overset{(b)}{\leq} \exp\left(-\tfrac{3}{8}\lambda\tau\right),\end{aligned} \tag{EC.29}$$

where part (a) uses the definition of s_τ and part (b) comes from (EC.28). Similarly, if $\tilde{\pi}(q^j) - b^i > s_\tau$, using the upper bound in (EC.28) again, it holds that

$$\mathbb{P}(\widetilde{D}^j < \widetilde{D}^i) = \mathbb{P}\Big\{R^j + T\tilde{\pi}(q^j) < R^i + Tb^i\Big\} \leq \mathbb{P}\Big\{R^i > T(\tilde{\pi}(q^j) - b^i)\Big\} \leq \mathbb{P}\Big\{R^i > 2\lambda\tau\Big\} \leq \exp\left(-\tfrac{3}{8}\lambda\tau\right). \tag{EC.30}$$

Either way, it can be observed that $\mathbb{I}\{b^i > \tilde{\pi}(q^j)\} = \mathbb{I}\{\widetilde{D}^i > \widetilde{D}^j\}$ with high probability. To this end, let us define the following event

$$\Xi_i := \Big\{\forall j \text{ such that } |b^i - \tilde{\pi}(q^j)| > s_\tau : \ \ \mathbb{I}\{b^i > \tilde{\pi}(q^j)\} = \mathbb{I}\{\widetilde{D}^i > \widetilde{D}^j\}\Big\}. \tag{EC.31}$$

Invoking the union bound and the observations in (EC.29)-(EC.30), it holds that

$$\mathbb{P}(\Xi_i^C) \leq \textstyle\sum_{j:b^i - \tilde{\pi}(q^j) > s_\tau} \mathbb{P}(\widetilde{D}^i < \widetilde{D}^j) + \sum_{j:\tilde{\pi}(q^j) - b^i > s_\tau} \mathbb{P}(\widetilde{D}^i > \widetilde{D}^j) \leq N\exp\left(-\tfrac{3}{8}\lambda\tau\right). \tag{EC.32}$$

As a result, this "good" event happens with high probability, too.

Step 1-2. For any ranking σ, we define the set of agents ranked above agent i under a ranking σ as

$$S^i_\sigma := \Big\{j \neq i : \sigma(j) > \sigma(i)\Big\}.$$

Conditional on the event Ξ_i, the discrepancy between $S^i_{\overline{\sigma}}$ and $S^i_{H^N(b^i, \tilde{\boldsymbol{\pi}}_N^{-i})}$ can only come from agents whose inflation intensities lie in the interval $[b^i - s_\tau, b^i + s_\tau]$. However, the number of such agents can be controlled. More precisely, since $\widetilde{\pi}'(q) = \frac{\lambda}{c} q(q+1)\exp(q^2/2 + q - 3/2)$, for all $q \in [0,1]$, we have $\widetilde{\pi}'(q) \geq \frac{\lambda}{c} e^{-3/2} q$. Hence, for any $0 \leq q < q' \leq 1$,

$$\widetilde{\pi}(q') - \widetilde{\pi}(q) \geq \frac{\lambda}{2c} e^{-3/2}((q')^2 - q^2) \geq \frac{\lambda}{2c} e^{-3/2}(q' - q)^2.$$

Therefore, $|\tilde{\pi}(q^j) - b^i| \leq s_\tau$ implies that $|\tilde{\pi}^{-1}(b^i) - q^j| \leq \sqrt{\frac{2ce^{3/2}}{\lambda} s_\tau}$. Since $\{q^j\} = \{(j-1)/(N-1)\}_j$ forms a uniform mesh grid on the unit interval $[0,1]$, the number of agents whose inflation is between $[b^i - s_\tau, b^i + s_\tau]$ is upper bounded by $2(N-1)\sqrt{\frac{2ce^{3/2}}{\lambda} s_\tau}$.

As a result, we may control $\left|F(q^i, \overline{\sigma}) - F\left(q^i, H^N(b, \widetilde{\boldsymbol{\pi}}_N^{-i})\right)\right|$, too. Recall from (9) that with $r(q^i) = q^i/N$ and $\zeta = 1/N$, the reward function under the cascade model is given by $F(q^i, \sigma) = \lambda q^i \prod_{j \in S^i_\sigma}[\left(1 - \frac{q^j}{N}\right)(1 - \frac{1}{N})]$. The following technical lemma quantifies the sensitivity of $F(q^i, \sigma)$ with respect to σ.

Lemma EC.3. *$|F(q^i, \hat{\sigma}) - F(q^i, \dot{\sigma})| \leq \lambda q^i \left(2 - \exp\left(-\frac{3|S^i_{\hat{\sigma}} \setminus S^i_{\dot{\sigma}}|}{N}\right) - \exp\left(-\frac{3|S^i_{\dot{\sigma}} \setminus S^i_{\hat{\sigma}}|}{N}\right)\right)$ for any two rankings $\hat{\sigma}$ and $\dot{\sigma}$.*

Since $|S^i_{\overline{\sigma}} \setminus S^i_{H^N(b,\boldsymbol{b}^{-i})}|$ and $|S^i_{H^N(b,\boldsymbol{b}^{-i})} \setminus S^i_{\overline{\sigma}}|$ are both no more than $2(N-1)\sqrt{\frac{2ce^{3/2}}{\lambda}s_\tau}$, we find that

$$\begin{aligned}
&\mathbb{E}\Big[\,\big|F(q^i,\overline{\sigma}) - F\left(q^i, H^N(b^i,\widetilde{\boldsymbol{\pi}}_N^{-i})\right)\big|\,|\Xi_i\Big] \\
\leq &\lambda q^i\left(2-\exp\left(-\tfrac{3}{N}|S^i_{\overline{\sigma}} \setminus S^i_{H^N(b,\boldsymbol{b}^{-i})}|\right) - \exp\left(-\tfrac{3}{N}|S^i_{H^N(b,\boldsymbol{b}^{-i})} \setminus S^i_{\overline{\sigma}}|\right)\right), \\
\leq &2\lambda q^i\left(1-\exp\Big(-6\sqrt{\frac{2ce^{3/2}}{\lambda}s_\tau}\Big)\right)
\end{aligned} \tag{EC.33}$$

Step 1-3. To proceed, the absolute difference between the utility in the pre-limit and the limiting games can be bounded from above:

$$\begin{aligned}
&\big|v^i(b^i,\tilde{\boldsymbol{\pi}}_N^{-i}) - \overline{v}^i(b^i,\tilde{\boldsymbol{\pi}}_N^{-i})\big| \\
=\ &|\mathbb{E}[\tfrac{1}{T}\textstyle\sum_{t=1}^T F(q^i,\sigma_t)] - F\big(q^i,H^N(b^i,\tilde{\boldsymbol{\pi}}_N^{-i})\big| \\
\leq\ &\mathbb{E}[|\tfrac{1}{T}\textstyle\sum_{t=1}^T F(q^i,\sigma_t) - F\big(q^i,H^N(b^i,\tilde{\boldsymbol{\pi}}_N^{-i})\big)|] \\
\leq\ &\mathbb{E}[|\tfrac{1}{T}\textstyle\sum_{t=1}^T F(q^i,\sigma_t) - F\big(q^i,H^N(b^i,\tilde{\boldsymbol{\pi}}_N^{-i})\big)|\,|\,\Xi_i^C]\mathbb{P}(\Xi_i^C) + \mathbb{E}[|\tfrac{1}{T}\textstyle\sum_{t=1}^T F(q^i,\sigma_t) - F\big(q^i,H^N(b^i,\tilde{\boldsymbol{\pi}}_N^{-i})\big)|\,|\,\Xi_i] \\
\overset{(a)}{\leq}\ &\lambda q^i\cdot N\exp\left(-\tfrac{3}{8}\lambda\tau\right) + \mathbb{E}[|\tfrac{1}{T}\textstyle\sum_{t=1}^T F(q^i,\sigma_t) - F\big(q^i,H^N(b^i,\tilde{\boldsymbol{\pi}}_N^{-i})\big)|\,|\,\Xi_i] \\
\overset{(b)}{\leq}\ &\lambda q^i\cdot N\exp\left(-\tfrac{3}{8}\lambda\tau\right) + \left[\tfrac{\tau}{T}\lambda q^i + \tfrac{T-\tau}{T}\mathbb{E}[\big|F(q^i,\overline{\sigma}) - F\left(q^i,H^N(b^i,\widetilde{\boldsymbol{\pi}}_N^{-i})\right)\big|\,|\Xi_i]\right] \\
\overset{(c)}{\leq}\ &\lambda q^i\cdot N\exp\left(-\tfrac{3}{8}\lambda\tau\right) + \tfrac{\tau}{T}\lambda q^i + 2\lambda q^i\left(1-\exp\Big(-6\sqrt{\frac{2ce^{3/2}}{\lambda}s_\tau}\Big)\right) \\
\overset{(d)}{\leq}\ &\lambda N\exp\left(-\tfrac{3}{8}\lambda\tau\right) + \tfrac{\tau}{T}\lambda + 2\lambda\left(1-\exp\Big(-6\sqrt{\frac{4ce^{3/2}\tau}{T}}\Big)\right),
\end{aligned}$$

where in step (a), we use the inequality in (EC.32), as well as the universal upper bound of reward function $F(q^i,\sigma) \leq \lambda N q^i/N = \lambda q^i$ under the cascade model. In step (b), we have once again the gap in reward function is upper bounded by λq^i for the first τ periods, and the bound for the last $T-\tau$ periods follows from the fact that $\sigma_t \equiv \overline{\sigma}$ in the commitment phase. Step (c) follows from the observation in (EC.33). Finally, step (d) uses the definition of s_τ and the fact that $q^i \leq 1$.

Step 2: Quantify the proximity between $\overline{v}^i(b^i,\tilde{\boldsymbol{\pi}}_N^{-i})$ and $U_{\tilde{\pi}}(q^i,b^i)$. Since $\tilde{\pi}$ is continuous and strictly increasing, its inverse $\tilde{\pi}^{-1}$ exists on the interval $[\tilde{\pi}(0),\tilde{\pi}(1)]$. Without loss of generality, suppose $b^i \in [\tilde{\pi}(0),\tilde{\pi}(1)]$. Then, by the expression for $H(b^i \mid \tilde{\pi})$ in Equation (7), we have $H(b^i \mid \tilde{\pi}) = \tilde{\pi}^{-1}(b^i)$. Recall that the qualities $\{q^i\}_{i=1}^N = \left\{\frac{i-1}{N-1}\right\}$ form a uniform mesh grid over the interval $[0,1]$. We now bound the gap between $\overline{v}^i(b^i,\tilde{\boldsymbol{\pi}}_N^{-i})$ and $U_{\tilde{\pi}}(q^i,b^i)$.

$$\begin{aligned}
&\big|\overline{v}^i(b^i,\tilde{\boldsymbol{\pi}}_N^{-i}) - U_{\tilde{\pi}}(q^i,b^i)\big| \\
=\ &\left|\lambda q^i\prod_{j:\tilde{\pi}(q^j)>b^i}\left(1-\tfrac{q^j}{N}\right)(1-\tfrac{1}{N}) - cb^i - \left(\lambda q^i\exp\left(\tfrac{H^2(b^i|\tilde{\pi})}{2} + H(b^i|\tilde{\pi}) - \tfrac{3}{2}\right) - cb^i\right)\right| \\
=\ &\lambda q^i\left|\prod_{j:\tilde{\pi}(q^j)>b^i}\left(1-\tfrac{N+j-1}{N^2}\right) - \exp\left(\tfrac{H^2(b^i|\tilde{\pi})}{2} + H(b^i|\tilde{\pi}) - \tfrac{3}{2}\right)\right| \\
=\ &\lambda q^i\left|\exp\left(\sum_{j:\tilde{\pi}(q^j)>b^i}\log\left(1-\tfrac{N+j-1}{N^2}\right)\right) - \exp\left(\tfrac{H^2(b^i|\tilde{\pi})}{2} + H(b^i|\tilde{\pi}) - \tfrac{3}{2}\right)\right|
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \lambda q^i \left| \sum_{j:\tilde{\pi}(q^j)>b^i} \log\left(1-\tfrac{N+j-1}{N^2}\right) - \left(\tfrac{H^2(b^i|\tilde{\pi})}{2} + H(b^i|\tilde{\pi}) - \tfrac{3}{2}\right)\right| \\
&\overset{(b)}{=} \lambda q^i \left| \sum_{j:\tilde{\pi}(q^j)>b^i} \left(-\sum_{\ell=1}^{\infty} \tfrac{1}{\ell}\left(\tfrac{N+j-1}{N^2}\right)^{\ell}\right) - \left(\tfrac{H^2(b^i|\tilde{\pi})}{2} + H(b^i|\tilde{\pi}) - \tfrac{3}{2}\right)\right| \\
&\leq \lambda q^i \Bigg\{ \underbrace{\left| \sum_{j:\tilde{\pi}(q^j)>b^i} \left(-\tfrac{N+j-1}{N^2}\right) - \left(\tfrac{H^2(b^i|\tilde{\pi})}{2} + H(b^i|\tilde{\pi}) - \tfrac{3}{2}\right)\right|}_{(*)} + \underbrace{\left| \sum_{j:\tilde{\pi}(q^j)>b^i} \sum_{\ell=2}^{\infty} \tfrac{1}{\ell}\left(\tfrac{N+j-1}{N^2}\right)^{\ell}\right|}_{(**)} \Bigg\} \\
&\overset{(c)}{\leq} \lambda q^i \left(\tfrac{3}{N} + \tfrac{4}{N-2}\right) \\
&\overset{(d)}{\leq} \lambda \cdot \tfrac{14}{N},
\end{aligned}
$$

where step (a) follows from the inequality $|\exp(x)-\exp(y)| \leq |x-y|$ for $x, y \leq 0$, and noting that both $\sum_{j:\tilde{\pi}(q^j)>b^i} \log\left(1-\frac{N+j-1}{N^2}\right)$ and $\frac{H^2(b^i|\tilde{\pi})}{2} + H(b^i|\tilde{\pi}) - \frac{3}{2}$ are negative. In step (b), we use the Taylor expansion of the logarithm in the first term. In part (c), terms $(*)$ and $(**)$ are handled separately. Term $(*)$ is a bit more tedious, and we summarize the result using the technical lemma below.

LEMMA EC.4. $\left|\sum_{j:\tilde{\pi}(q^j)>b^i} \left(-\frac{N+j-1}{N^2}\right) - \left(\frac{H^2(b^i|\tilde{\pi})}{2} + H(b^i|\tilde{\pi}) - \frac{3}{2}\right)\right| \leq \frac{3}{N}$.

Term $(**)$, which comes from the remainder of the Taylor series, can be bounded using the derivation below:

$$
\left| \sum_{j:\tilde{\pi}(q^j)>b^i} \sum_{\ell=2}^{\infty} \tfrac{1}{\ell}\left(\tfrac{N+j-1}{N^2}\right)^{\ell}\right| \leq \sum_{j=1}^{N} \sum_{\ell=2}^{\infty} \tfrac{1}{\ell}\left(\tfrac{2}{N}\right)^{\ell} \leq \sum_{j=1}^{N} \tfrac{4}{N(N-2)} = \tfrac{4}{N-2}.
$$

Part (d) holds when we assume $N \geq 3$.

Step 3: Quantify the proximity between the pre-limit utility value $v^i(b^i, \widetilde{\boldsymbol{\pi}}_N^{-i})$ and $U_{\tilde{\pi}}(q^i, b^i)$.

$$
\begin{aligned}
&\left|v^i(b^i, \widetilde{\boldsymbol{\pi}}_N^{-i}) - U_{\widetilde{\pi}}(q^i, b^i)\right| \\
&\leq \left|v^i(b^i, \widetilde{\boldsymbol{\pi}}_N^{-i}) - \overline{v}^i(b^i, \widetilde{\boldsymbol{\pi}}_N^{-i})\right| + \left|\overline{v}^i(b^i, \widetilde{\boldsymbol{\pi}}_N^{-i}) - U_{\widetilde{\pi}}(q^i, b^i)\right| \\
&\overset{(a)}{\leq} \lambda N \exp(-\tfrac{3}{8}\lambda\tau) + \frac{\lambda\tau}{T} + 2\lambda\left(1 - \exp\left(-6\sqrt{\frac{4ce^{3/2}\tau}{T}}\right)\right) + \frac{14\lambda}{N} \\
&\overset{(b)}{=} \lambda N^{-5/4}T^{-9/8} + 3\frac{\log(N^2T)}{T} + 2\lambda\left(1 - \exp\left(-6\sqrt{\frac{12ce^{3/2}\log(N^2T)}{\lambda T}}\right)\right) + \frac{14\lambda}{N} \qquad \text{(EC.34)} \\
&\overset{(c)}{\leq} \lambda N^{-5/4}T^{-9/8} + 3\frac{\log(N^2T)}{T} + 12\lambda\sqrt{\frac{12ce^{3/2}\log(N^2T)}{\lambda T}} + \frac{14\lambda}{N} \\
&= \lambda N^{-5/4}T^{-9/8} + 3\frac{\log(N^2T)}{T} + 12\sqrt{12\lambda ce^{3/2}}\sqrt{\frac{\log(N^2T)}{T}} + \frac{14\lambda}{N} \\
&= \frac{\delta_1}{2}\frac{\log(N^2T)}{T} + \frac{\delta_2}{2}\frac{1}{N^{5/4}T^{9/8}} + \frac{\delta_3}{2}\sqrt{\frac{\log(N^2T)}{T}} + \frac{\delta_4}{2}\frac{1}{N} = \frac{\varepsilon}{2},
\end{aligned}
$$

where the inequality in (a) follows from Steps 1 and 2. In step (b), we plug in $\tau = 3\log(N^2T)/\lambda$. In step (c), we use $1-\exp(-x) \leq x$ for $x \geq 0$. Since the choices of $i \in [N]$ and $b^i \geq 0$ are arbitrary, the first statement of the theorem follows by letting $\delta_1 := 6, \delta_2 := 2\lambda, \delta_3 := 24\sqrt{12\lambda ce^{3/2}}, \delta_4 := 28\lambda$.

Step 4: Quantify the benefit of deviation. Building on the result in the previous step, we can bound the benefit of deviation from the NAE-implied intensity $\tilde{\pi}(q^i)$. Given any $b^i \geq 0$:

$$\begin{aligned} & v^i(b^i, \tilde{\boldsymbol{\pi}}_N^{-i}) - v^i(\tilde{\pi}(q^i), \tilde{\boldsymbol{\pi}}_N^{-i}) \\ = \; & v^i(b^i, \tilde{\boldsymbol{\pi}}_N^{-i}) - U_{\tilde{\pi}}(q^i, b^i) + U_{\tilde{\pi}}(q^i, b^i) - U_{\tilde{\pi}}(q^i, \tilde{\pi}(q^i)) + U_{\tilde{\pi}}(q^i, \tilde{\pi}(q^i)) - v^i(\tilde{\pi}(q^i), \tilde{\boldsymbol{\pi}}_N^{-i}) \\ \overset{(a)}{\leq} \; & \Big(v^i(b^i, \tilde{\boldsymbol{\pi}}_N^{-i}) - U_{\tilde{\pi}}(q^i, b^i)\Big) + 0 + \Big(U_{\tilde{\pi}}(q^i, \tilde{\pi}(q^i)) - v^i(\tilde{\pi}(q^i), \tilde{\boldsymbol{\pi}}_N^{-i})\Big) \\ \overset{(b)}{\leq} \; & 2 \times \tfrac{\varepsilon}{2} = \varepsilon, \end{aligned} \tag{EC.35}$$

where (a) holds because $U_{\tilde{\pi}}(q^i, b^i) \leq U_{\tilde{\pi}}(q^i, \tilde{\pi}(q^i))$ under the NAE; in (b), we apply the uniform bound (EC.34) twice for both components. ■

Proof of Lemma EC.3. We derive the following chain of inequalities.

$$\begin{aligned} & \big|F(q^i, \hat{\sigma}) - F(q^i, \dot{\sigma})\big| \\ = \; & \lambda q^i \left| \prod_{j \in S^i_{\hat{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big) - \prod_{j \in S^i_{\dot{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big) \right| \\ = \; & \lambda q^i \prod_{j \in S^i_{\hat{\sigma}} \cap S^i_{\dot{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big) \cdot \left| \prod_{j \in S^i_{\hat{\sigma}} \setminus S^i_{\dot{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big) - \prod_{j \in S^i_{\dot{\sigma}} \setminus S^i_{\hat{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big) \right| \\ \overset{(a)}{\leq} \; & \lambda q^i \left| \prod_{j \in S^i_{\hat{\sigma}} \setminus S^i_{\dot{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big) - \prod_{j \in S^i_{\dot{\sigma}} \setminus S^i_{\hat{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big) \right| \\ \overset{(b)}{\leq} \; & \lambda q^i \left(\left|1 - \prod_{j \in S^i_{\hat{\sigma}} \setminus S^i_{\dot{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big)\right| + \left|1 - \prod_{j \in S^i_{\dot{\sigma}} \setminus S^i_{\hat{\sigma}}} \Big(1 - \tfrac{q^j}{N}\Big)\Big(1 - \tfrac{1}{N}\Big)\right| \right) \\ \overset{(c)}{\leq} \; & \lambda q^i \Big(\Big|1 - (1 - \tfrac{1}{N})^{2|S^i_{\hat{\sigma}} \setminus S^i_{\dot{\sigma}}|}\Big| + \Big|1 - (1 - \tfrac{1}{N})^{2|S^i_{\dot{\sigma}} \setminus S^i_{\hat{\sigma}}|}\Big| \Big) \\ \leq \; & \lambda q^i \Big(2 - (1 - \tfrac{1}{N})^{2|S^i_{\hat{\sigma}} \setminus S^i_{\dot{\sigma}}|} - (1 - \tfrac{1}{N})^{2|S^i_{\dot{\sigma}} \setminus S^i_{\hat{\sigma}}|}\Big) \\ \overset{(d)}{\leq} \; & \lambda q^i \left(2 - \exp\Big(-\tfrac{3|S^i_{\hat{\sigma}} \setminus S^i_{\dot{\sigma}}|}{N}\Big) - \exp\Big(-\tfrac{3|S^i_{\dot{\sigma}} \setminus S^i_{\hat{\sigma}}|}{N}\Big)\right). \end{aligned}$$

In the derivations above, part (a) is because the prefactor $\prod_{j \in S^i_{\hat{\sigma}} \cap S^i_{\dot{\sigma}}}(1 - \frac{q^j}{N})(1 - \frac{1}{N})$ is a product of numbers strictly between 0 and 1; hence it is at most 1. Part (b) is because of the inequality $|A - B| = |(1 - A) - (1 - B)| \leq |1 - A| + |1 - B|$ for any two numbers A and B. Assigning $A = \prod_{j \in S^i_{\hat{\sigma}} \setminus S^i_{\dot{\sigma}}}(1 - \frac{q^j}{N})(1 - \frac{1}{N})$ and the analogous expression for B yields the desired bound. Part (c) is because $0 \leq q^j \leq 1$, and hence we have $(1 - \frac{q^j}{N})(1 - \frac{1}{N}) \geq (1 - \frac{1}{N})^2$. In part (d), note that for $x \in (0, 1)$ we have $\log(1 - x) \geq -x - \frac{x^2}{2(1-x)}$. Substituting $x = \frac{1}{N}$ gives $\log(1 - \frac{1}{N}) \geq -\frac{1}{N} - \frac{1}{2N(N-1)} = -\frac{1}{N}(1 + \frac{1}{2(N-1)}) \geq -\frac{1.5}{N}$ for all $N \geq 2$. It follows that $(1 - \frac{1}{N})^{2k} = \exp\big(2k \log(1 - \frac{1}{N})\big) \geq \exp\big(-\frac{3k}{N}\big)$. ■

Proof of Lemma EC.4. For shorthand notation, let

$$x := \tilde{\pi}^{-1}(b^i), \quad \text{and} \quad \varepsilon_x := \lceil (N-1)x \rceil - (N-1)x,$$

so that $\lceil (N-1)x \rceil = (N-1)x + \varepsilon_x$. Here, x denotes the agent's ranking position, and ε_x captures the corresponding rounding error. Later, we will use the fact that $0 \le \varepsilon_x < 1$, $0 \le x \le 1$, and $|\varepsilon_x - x| \le 1$. Because $\tilde{\pi}(q^j) > b^i$ iff $q^j > \tilde{\pi}^{-1}(b^i)$ iff $j \ge \lceil (N-1)x \rceil + 1$, define

$$\begin{aligned} G_{\text{disc}} = \sum_{j:\tilde{\pi}(q^j)>b^i} \left(-\tfrac{N+j-1}{N^2}\right) &:= -\tfrac{1}{N^2} \sum_{j=\lceil (N-1)x \rceil + 1}^{N} (N-1+j) \\ &= \tfrac{1}{2N} + \tfrac{(N-1)x+\varepsilon_x}{N} - \tfrac{(N-1)x+\varepsilon_x}{2N^2} + \tfrac{\left((N-1)x+\varepsilon_x\right)^2}{2N^2} - \tfrac{3}{2}. \end{aligned}$$

Also, define

$$G_{\text{cont}} := \frac{H^2(b^i \mid \tilde{\pi})}{2} + H(b^i \mid \tilde{\pi}) - \frac{3}{2} = \frac{(\tilde{\pi}^{-1}(b^i))^2}{2} + \tilde{\pi}^{-1}(b^i) - \frac{3}{2} = \frac{x^2}{2} + x - \frac{3}{2}$$

Then

$$|G_{\text{disc}} - G_{\text{cont}}| = \left| \tfrac{1}{2N} + \tfrac{\varepsilon_x - x}{N} + \tfrac{(\varepsilon_x - x)\left((N-1)x+\varepsilon_x+Nx\right)}{2N^2} - \tfrac{(N-1)x+\varepsilon_x}{2N^2} \right| \le \tfrac{3}{N},$$

since we can bound the gap term by term: $\left|\tfrac{\varepsilon_x - x}{N}\right| \le \tfrac{1}{N}$, $\left|\tfrac{(\varepsilon_x - x)\left((2N-1)x+\varepsilon_x\right)}{2N^2}\right| \le \tfrac{1 \cdot 2N}{2N^2} = \tfrac{1}{N}$, and $\left|\tfrac{(N-1)x+\varepsilon_x}{2N^2}\right| \le \tfrac{N}{2N^2} = \tfrac{1}{2N}$. ■

Proof of Theorem 8. Step 1. Optimality of quality-based ranking under cascade model. In this step, we show that the clairvoyant wishes to use the quality-based ranking $\boldsymbol{e}$ to maximize the total reward, i.e., the analogue of Assumption (i) holds under the cascade model. To be more precise, we show that

$$\sum_{i \in [N]} F(q^i, \hat{\sigma}) \le \sum_{i \in [N]} F(q^i, \boldsymbol{e}), \quad \text{for every ranking } \hat{\sigma}.$$

While this property is fairly well known, we verify it for completeness. Without loss of generality, pick a ranking $\hat{\sigma} \ne \boldsymbol{e}$. Then there exists a pair of agents $\{m, n\}$ such that $m > n$ but $\hat{\sigma}(m) = \hat{\sigma}(n) - \frac{1}{N-1}$, meaning that agent m has higher quality than agent n, yet is ranked one position lower in the ranking $\hat{\sigma}$. It suffices to show that the total expected reward can be improved by swapping the ranking positions of agents m and n. To that end, define $\mathcal{P}(i) := \left(1 - \frac{q^i}{N}\right)(1-\zeta)$ for notational convenience. We decompose the total reward under $\hat{\sigma}$ as follows:

$$\begin{aligned} \sum_{i \in [N]} F(q^i, \hat{\sigma}) &= \sum_{i \in [N] \setminus \{n,m\}} F(q^i, \hat{\sigma}) + \lambda N \prod_{j:\hat{\sigma}(j) > \hat{\sigma}(n)} \mathcal{P}(j) \left[r(q^n) + \mathcal{P}(n) r(q^m) \right] \\ &\overset{(a)}{=} \sum_{i \in [N] \setminus \{n,m\}} F(q^i, \hat{\sigma}) + \lambda N \prod_{j:\hat{\sigma}(j) > \hat{\sigma}(n)} \mathcal{P}(j) \left[r(q^n) + (r(q^m) - r(q^n))(1-\zeta) + \mathcal{P}(m) r(q^n) \right] \\ &\overset{(b)}{\le} \sum_{i \in [N] \setminus \{n,m\}} F(q^i, \hat{\sigma}) + \lambda N \prod_{j:\hat{\sigma}(j) > \hat{\sigma}(n)} \mathcal{P}(j) \left[r(q^m) + \mathcal{P}(m) r(q^n) \right], \end{aligned}$$

where step (a) uses the identity $r(q^j)\mathcal{P}(i) - r(q^i)\mathcal{P}(j) = [r(q^j) - r(q^i)](1-\zeta)$, and step (b) follows from the fact that $\zeta \in (0,1)$, $n < m$, and the monotonicity of $r(\cdot)$, which implies $r(q^n) < r(q^m)$. The proof is thus finished by observing that the right-hand side corresponds to the total expected reward under a new ranking that is identical to $\hat{\sigma}$ except that the positions of agents m and n are swapped.

Step 2: Establish the learner's regret. Suppose that all agents follow the NAE and consider the action profile $\tilde{\boldsymbol{\pi}}_N$. Since $\tilde{\pi}$ is strictly increasing, the corresponding inflation-intensity based ranking satisfies $H^N(\tilde{\boldsymbol{\pi}}_N^i, \tilde{\boldsymbol{\pi}}_N^{-i}) = \boldsymbol{e}$, which, as shown in Step 1, maximizes the learner's total expected reward.

We refine Step 1-2 in the proof of Theorem 7 by retaining the dependence on the quality q^i. Throughout this step, we use $b^i = \widetilde{\pi}(q^i)$ and $s_\tau = (2\lambda\tau)/T$. Since $\widetilde{\pi}'(q) = \frac{\lambda}{c} q(q+1)\exp(q^2/2+q-3/2)$, it follows that $\widetilde{\pi}'(q) \geq \frac{\lambda}{c} e^{-3/2} q$ for every $q \in [0,1]$. Thus, for any $0 \leq q < q' \leq 1$, we have $\widetilde{\pi}(q') - \widetilde{\pi}(q) \geq \frac{\lambda}{2c} e^{-3/2}((q')^2 - q^2)$. Consequently, if $|\widetilde{\pi}(q^j) - b^i| \leq s_\tau$, then $|(q^j)^2 - (q^i)^2| \leq \frac{2ce^{3/2}}{\lambda} s_\tau = \frac{4ce^{3/2}\tau}{T}$.

Therefore, any such q^j must lie in the interval

$$\left[\sqrt{\left((q^i)^2 - \frac{4ce^{3/2}\tau}{T}\right)_+}, \sqrt{\min\left\{1, (q^i)^2 + \frac{4ce^{3/2}\tau}{T}\right\}}\right].$$

The number of agents $j \neq i$ satisfying $|\widetilde{\pi}(q^j) - b^i| \leq s_\tau$ is therefore bounded by

$$(N-1)\left(\sqrt{\min\left\{1, (q^i)^2 + \frac{4ce^{3/2}\tau}{T}\right\}} - \sqrt{\left((q^i)^2 - \frac{4ce^{3/2}\tau}{T}\right)_+}\right).$$

On the event Ξ_i, the only possible discrepancy between $S^i_{\overline{\sigma}}$ and $S^i_{H^N(b^i, \widetilde{\boldsymbol{\pi}}_N^{-i})}$ comes from these agents. Hence, by Lemma EC.3, together with the identity $F(q^i, \boldsymbol{e}) = F(q^i, H^N(b^i, \widetilde{\boldsymbol{\pi}}_N^{-i}))$, we obtain

$$\begin{aligned}
&\mathbb{E}\big[F(q^i, \boldsymbol{e}) - F(q^i, \overline{\sigma}) \,\big|\, \Xi_i\big] \\
&\leq 2\lambda q^i \left(1 - \exp\left(-\frac{3(N-1)}{N}\left(\sqrt{\min\left\{1, (q^i)^2 + \frac{4ce^{3/2}\tau}{T}\right\}} - \sqrt{\left((q^i)^2 - \frac{4ce^{3/2}\tau}{T}\right)_+}\right)\right)\right).
\end{aligned}$$

The following elementary lemma will be used to aggregate the preceding bound over i.

LEMMA EC.5. *For any $A \geq 0$ and any $q \in [0,1]$,*

$$q\left(\sqrt{\min\{1, q^2 + A\}} - \sqrt{(q^2 - A)_+}\right) \leq \sqrt{2}\, A.$$

We now sum over all agents. Using $1 - e^{-x} \leq x$, the preceding display implies that the total contribution from the events Ξ_i is at most

$$\frac{6\lambda(N-1)}{N} \sum_{i \in [N]} q^i \left(\sqrt{\min\left\{1, (q^i)^2 + \frac{4ce^{3/2}\tau}{T}\right\}} - \sqrt{\left((q^i)^2 - \frac{4ce^{3/2}\tau}{T}\right)_+}\right).$$

Applying Lemma EC.5 with $A = \frac{4ce^{3/2}\tau}{T}$ and $q = q^i$, each summand is bounded by $\frac{4\sqrt{2}\, ce^{3/2}\tau}{T}$. Hence the contribution from the event Ξ_i is bounded by

$$\frac{6\lambda(N-1)}{N} \cdot N \cdot \frac{4\sqrt{2}\, ce^{3/2}\tau}{T} = 24\sqrt{2}\lambda ce^{3/2}(N-1)\frac{\tau}{T}.$$

During the first τ periods, namely the experimentation phase, the expected loss in reward is at most $\lambda N \tau$. It remains to bound the expected loss in the commitment phase. Using the same event Ξ_i as in (EC.31), we decompose

$$\begin{aligned}
&(T-\tau) \sum_{i \in [N]} \mathbb{E}\big[F(q^i, \boldsymbol{e}) - F(q^i, \overline{\sigma})\big] \\
&= (T-\tau) \sum_{i \in [N]} \Big[\mathbb{E}\big[F(q^i, \boldsymbol{e}) - F(q^i, \overline{\sigma}) \,\big|\, \Xi_i^C\big]\, \mathbb{P}(\Xi_i^C) + \mathbb{E}\big[F(q^i, \boldsymbol{e}) - F(q^i, \overline{\sigma}) \,\big|\, \Xi_i\big]\, \mathbb{P}(\Xi_i)\Big] \\
&\leq (T-\tau)\left[\frac{\lambda N^2}{2}\exp\left(-\frac{3}{8}\lambda\tau\right) + 24\sqrt{2}\lambda ce^{3/2}(N-1)\frac{\tau}{T}\right].
\end{aligned}$$

On the complement event, we use the universal bound $F(q^i, \boldsymbol{e}) - F(q^i, \overline{\sigma}) \le \lambda q^i$, together with $\mathbb{P}(\Xi_i^C) \le N \exp(-3\lambda\tau/8)$. Since $\sum_{i\in[N]} q^i = N/2$, this contribution is at most $\frac{\lambda N^2}{2} \exp\left(-\frac{3}{8}\lambda\tau\right)$.

Combining the experimentation and commitment phases, and using $T - \tau \le T$, we obtain

$$\mathcal{R}_T \le \lambda N\tau + \frac{\lambda N^2 T}{2} \exp\left(-\frac{3}{8}\lambda\tau\right) + 24\sqrt{2}\lambda c e^{3/2}(N-1)\tau.$$

Now choose $\tau = 3\log(N^2T)/\lambda$. Then $\lambda N\tau = 3N\log(N^2T)$, and

$$\frac{\lambda N^2 T}{2} \exp\left(-\frac{3}{8}\lambda\tau\right) = \frac{\lambda N^2 T}{2} \exp\left(-\frac{9}{8}\log(N^2T)\right) = \frac{\lambda}{2}(N^2T)^{-1/8} \le \lambda.$$

Moreover,

$$24\sqrt{2}\lambda c e^{3/2}(N-1)\tau = 72\sqrt{2}\, c e^{3/2}(N-1)\log(N^2T) \le 72\sqrt{2}\, c e^{3/2} N \log(N^2T).$$

Thus,

$$\mathcal{R}_T \le \left(3 + 72\sqrt{2}\, c e^{3/2}\right) N \log(N^2T) + \lambda.$$

Equivalently, $\mathcal{R}_T \le \phi_1 N \log(N^2T) + \phi_2$, where $\phi_1 := 3 + 72\sqrt{2}\, c e^{3/2}$ and $\phi_2 := \lambda$. ■

Proof of Lemma EC.5. If $A = 0$, the claim is immediate. Suppose $A > 0$. We consider two cases. First, if $q^2 \le A$, then $(q^2 - A)_+ = 0$, and therefore $q\left(\sqrt{\min\{1, q^2 + A\}} - \sqrt{(q^2 - A)_+}\right) \le q\sqrt{q^2 + A}$. Since $q \le \sqrt{A}$ and $q^2 + A \le 2A$, the right-hand side is at most $\sqrt{A}\sqrt{2A} = \sqrt{2}\, A$.

Second, suppose $q^2 > A$. Then $(q^2 - A)_+ = q^2 - A$, and replacing $\sqrt{\min\{1, q^2 + A\}}$ by the larger quantity $\sqrt{q^2 + A}$ can only increase the expression. Hence the left-hand side is at most $q(\sqrt{q^2 + A} - \sqrt{q^2 - A})$. Rationalizing the difference gives $q(\sqrt{q^2 + A} - \sqrt{q^2 - A}) = \frac{2qA}{\sqrt{q^2+A} + \sqrt{q^2-A}}$. Moreover, $\sqrt{q^2 + A} + \sqrt{q^2 - A} \ge \sqrt{2}\, q$, because the square of the left-hand side is $2q^2 + 2\sqrt{q^4 - A^2} \ge 2q^2$. Thus $\frac{2qA}{\sqrt{q^2+A} + \sqrt{q^2-A}} \le \frac{2qA}{\sqrt{2}\, q} = \sqrt{2}\, A$. This proves the claim.

■