LEARNING STABLE MATCHINGS UNDER SOCIAL INFLUENCE: A BANDIT APPROACH

Anonymous authors
Paper under double-blind review

ABSTRACT

In many matching markets, participants evaluate options individually, discuss them with peers, and revise their preferences before a final matching is made. The preferences governing stability therefore reflect both personal experience and social influence. However, existing learning models in matching markets abstract away this deliberation stage, treating preferences as fixed lists to be learned in isolation through direct interaction feedback. We study a pure-exploration setting where direct interactions reveal noisy intrinsic evaluations, but the preferences that determine stability emerge only after these evaluations propagate through a social network. Our goal is to identify the player-optimal socially stable matching with as few evaluations as possible. The central difficulty is a decoupling of evidence and decision: uncertainty in one player's final ranking often originates from the intrinsic evaluations of others. Consequently, observing an unresolved ranking does not reveal which player–arm pairs should be sampled next. We characterize this challenge by establishing an instance-dependent lower bound through a network-weighted information-allocation problem over intrinsic source pairs, capturing necessary evidence to rule out alternatives that make the target matching unstable or no longer player-optimal. Guided by this variance structure, we design Source-Aware Elimination with Quadratic Allocation (SAE-Q), an algorithm that coordinates evaluations across source pairs to resolve network-propagated uncertainties. We prove that SAE-Q identifies the target matching with probability at least $1 - \delta$, achieving instance-dependent sample complexity guarantees explicitly governed by network topology and social preference gaps. Our results show that social influence shapes not only which matching is stable, but also where evidence must be collected to identify it.

1 INTRODUCTION

In two-sided matching markets, such as employment (Kelso Jr & Crawford, 1982; Roth, 1984), school admissions (Roth, 2008), and online dating (Hitsch et al., 2010), finding a good match depends on how participants evaluate their options. In these scenarios, participants rarely decide what they prefer from direct experience alone. A job seeker evaluating multiple employers gains initial evaluations from interviews, yet also turns to colleagues and alumni for information about employers and career opportunities (Granovetter, 1973; Calvo-Armengol & Jackson, 2004). Families choosing schools likewise compare their impressions with the experiences of neighbors (Dustan, 2018; Campos, 2024). These conversations do more than supply missing details. Personal encounters provide a starting point, but exchanges with peers can lead participants to reconsider their priorities and adjust their choices (Campos, 2024). Consequently, the preferences that ultimately guide their decisions are not the isolated impressions formed at the outset, but opinions shaped and refined through social interactions before the final assignment is made.

Learning in matching markets is typically formulated around an unknown but fixed preference profile (Liu et al., 2020; 2021; Sankararaman et al., 2021; Basu et al., 2021; Kong et al., 2022; Zhang et al., 2022; Kong & Li, 2023; Kong et al., 2024). Agents on one side of the market interact with alternatives on the other, using noisy observations from matched pairs to estimate the underlying utilities that determine their rankings (Liu et al., 2020; Kong et al., 2022; Kong & Li, 2023). This direct link between interaction feedback and preference ranking is also shared by recent pure-exploration formulations, which seek to identify a stable matching using as few observations as possible (Hosseini

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al., 2024; Athanasopoulos et al., 2026; Pagare et al., 2025; Kong et al., 2026). When participants consult peers before matching, however, this direct correspondence no longer holds. An agent's own experience reflects only their intrinsic evaluation. The preferences that ultimately determine whether a matching is stable emerge only after participants exchange opinions across a social network.

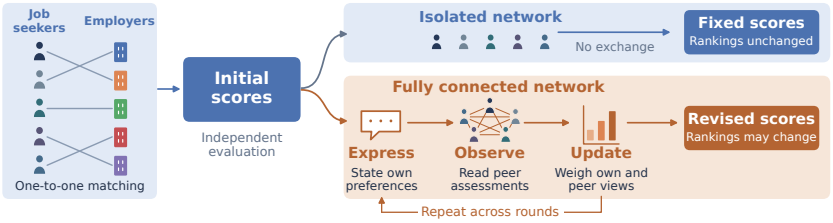


Figure 1: Preference-formation illustration. Agents initially evaluate employers independently, then repeatedly exchange evaluations with neighbors and update their scores based on personal priorities and peer feedback.

In this paper, we consider pure exploration in matching markets where personal evaluations take place before social interaction. A platform conducts preliminary evaluations to learn participants' intrinsic evaluations, after which participants communicate over a social network, holding to their initial evaluations while incorporating the views of their peers (DeGroot, 1974; Friedkin & Johnsen, 1990; Friedkin et al., 2016; 2019). The final matching is evaluated with respect to these settled, socially formed preferences. Our goal is to identify the player-optimal stable matching with high confidence, using as few evaluations as possible. Through the social network, an agent's critical preferences are shaped by the judgments of others. A decisive ranking may therefore remain uncertain simply because a peer's evaluation is imprecise—creating a source–destination mismatch between where uncertainty appears and where evidence is needed.

This source–destination mismatch turns pure exploration into a global sample allocation problem. Rather than directly sampling uncertain pairs, the platform must determine which intrinsic evaluations dictate the stability of the final matching and allocate observations among them. This allocation is coupled across the network, where an agent's final ranking aggregates noise from multiple peers, while a single intrinsic evaluation can clarify preferences for several participants at once. Furthermore, the algorithm must rule out not only unstable configurations that admit blocking pairs, but also competing stable matchings that could challenge the target matching in player-optimality.

To resolve this challenge, we first establish an instance-dependent sample complexity lower bound and then develop a source-aware elimination algorithm that systematically aligns observation allocation with the network structure. Our contributions are summarized as follows:

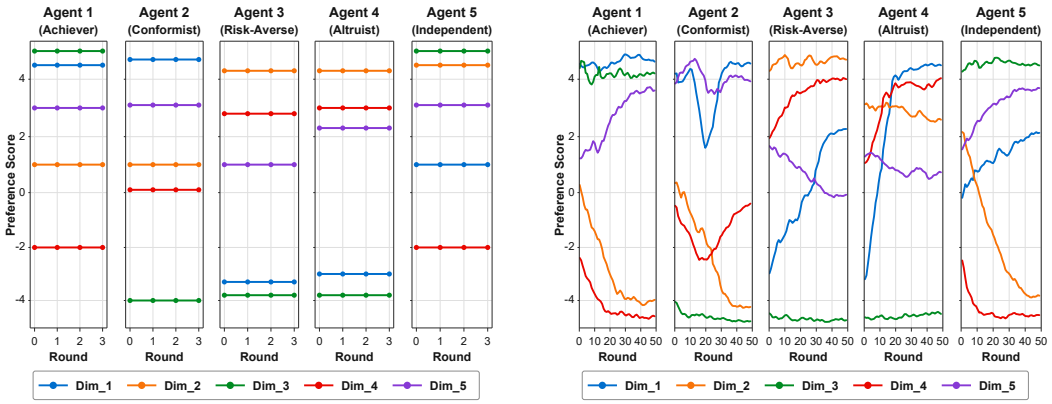
- Motivated by an LLM-agent experiment illustrating how peer communication alters preferences (Section 2), we formulate a pure-exploration problem in matching markets across individual evaluation, social deliberation, and final matching in Section 4.
- We formalize the resulting source-destination mismatch and characterize the fundamental learning difficulty it creates through an instance-dependent lower bound (Section 5). By accounting for both blocking pairs and competing stable matchings, the bound establishes a network-weighted constraint that couples sample requirements across the market.
- Guided by the propagated variance structure, we develop SAE-Q (Section 6), an elimination algorithm that globally coordinates source evaluations to certify social rankings and recover the player-optimal stable matching with high probability. Furthermore, we analyze the structural distinction between global preference certification and matching-specific identification (Appendix F), developing an adaptive refinement under serial dictatorship that sharply targets matching-critical comparisons, supported by numerical illustrations across various topologies (Appendix F.6).

2 Motivating Illustration of Social Preference Formation

Job seekers often discuss employers with peers while weighing salary, work-life balance, and career prospects. Such communication can reshape how they rank potential employers. To illustrate this

process, we use an LLM-based simulation of five job seekers evaluating five employers, following work on LLM-based social simulation (Argyle et al., 2023; Park et al., 2023; Gao et al., 2023; Xu et al., 2023; Ashery et al., 2025; Gao et al., 2024). The agents differ in their career priorities and receptiveness to social influence. As shown in Figure 1, they first evaluate employers independently, then repeatedly exchange assessments with their neighbors and revise their evaluations.

To illustrate how social structure shapes agents' preferences, we examine two contrasting network configurations: an isolated network, in which agents do not communicate with one another, and a fully connected network, in which every agent can observe all other agents' assessments. Each subplot in Figure 2 represents a job seeker, and each colored curve tracks the score assigned to an employer over successive interaction steps. In the isolated network (Figure 2a), scores remain at their initial values because agents receive no social information. By contrast, in the fully connected network (Figure 2b), scores evolve as agents exchange assessments, and several curves intersect, indicating reversals in employer rankings. For example, Agent 4 initially ranks Employer 2 (orange) above Employer 1 (blue), but their ordering reverses during social interaction, with Employer 1 ranked above Employer 2 at convergence. This contrast demonstrates that agents' preferences are not fixed; rather, they are reshaped through social interaction, with the network structure determining whether and how such preference changes occur. Simulation details are provided in Appendix A.



(a) Isolated topology

(b) Fully connected topology

Figure 2: Evolution of preference scores under isolated (a) and fully connected (b) networks. Each column represents an agent (job seeker), and colored curves track evaluations for employers (Dim 1–5) across rounds. Without social interaction (a), initial scores remain fixed; under full connectivity (b), peer influence induces ranking reversals across agents, stabilizing by round 50.

This example motivates evaluating matching stability with respect to socially influenced preferences: a matching that is stable under initial evaluations may cease to be stable after social interaction. To model how these preferences form, we adopt anchored linear opinion dynamics (DeGroot, 1974; Friedkin & Johnsen, 1990), in which agents combine their intrinsic evaluations with their neighbors' current opinions, with individual weights capturing differences in receptiveness to influence. Section 4 formalizes this process and defines socially stable matching under the resulting preferences.

3 Related Work

Social Effects and Network Structure in Matching Markets. A classical literature studies matching with allocation externalities, where an agent's utility depends not only on its own partner but also on others' assignments. Such dependence may arise through the entire matching (Sasaki & Toda, 1996), peer or coworker composition (Dutta & Massó, 1997; Echenique & Yenmez, 2007; Dur & Wiseman, 2019; Bykhovskaya, 2020), or network-based peer externalities, where utilities depend on the assignments or payoffs of socially connected agents (Bodine-Baron et al., 2011; Baccara et al., 2012; Anshelevich et al., 2017). This literature studies how such interdependence alters stability, computation, and welfare (Bando et al., 2016). A second line instead uses network or platform structure to constrain the matching process, including partner discovery and local stability (Arcaute

& Vassilvitskii, 2009; Hoefer, 2013; Hoefer & Wagner, 2017), admissible blocking pairs (Askalidis et al., 2013), network-based participation (Cho et al., 2022; Li et al., 2026), and platform-mediated visibility or search (Halaburda et al., 2018; Arnosti et al., 2021; Kanoria & Saban, 2021).

Closer to our setting in timing, Yang et al. (2021); Cai et al. (2022) incorporate social-network information before matching to aggregate multi-attribute evaluations or satisfaction scores, and then solve two-sided matching as an optimization problem. In contrast, our network operates directly on the intrinsic evaluations of the agents being matched, inducing the preference rankings with respect to which two-sided stability is defined. Therefore, prior work primarily places social dependence in allocation utilities, matching opportunities, or evaluation aggregation, whereas we study network-mediated preference formation itself.

Opinion Dynamics on Social Networks. The DeGroot model describes repeated weighted averaging of neighbors' opinions, while the Friedkin–Johnsen (FJ) model additionally anchors agents to their initial opinions, allowing persistent disagreement (DeGroot, 1974; Friedkin & Johnsen, 1990). Multidimensional extensions allow agents to hold opinions on multiple topics and introduce a coupling matrix across topics; when this matrix is the identity, the dimensions evolve independently under the same FJ influence process (Friedkin et al., 2016; Parsegov et al., 2017). This independent-dimension form has also been applied to multidimensional resource-allocation preferences (Friedkin et al., 2019). Our preference-formation dynamics adopts this formulation, where dimensions correspond to potential partners and initial opinions represent intrinsic utilities. Our contribution focuses on the downstream learning problem: the limiting evaluations induce preference rankings that govern two-sided stability, which must be identified under bandit feedback.

Bandit learning in matching markets. A broad literature studies learning in matching markets under bandit feedback, including stable-regret minimization in centralized and decentralized settings (Liu et al., 2020; 2021; Sankararaman et al., 2021; Basu et al., 2021; Maheshwari et al., 2022; Kong et al., 2022; Zhang et al., 2022; Kong & Li, 2023; Kong et al., 2024), as well as variants with equilibrium learning (Jagadeesan et al., 2023), contextual or Markovian preferences (Min et al., 2022), two-sided uncertainty (Zhang & Fang, 2024; Basu, 2025), and ties or indifferences (Kong et al., 2025; Lin et al., 2025). Related pure-exploration work studies high-confidence identification of stable matchings and its sample complexity (Hosseini et al., 2024; Athanasopoulos et al., 2025; 2026; Pagare et al., 2025; Kong et al., 2026). These works typically learn the preferences that directly determine matching outcomes. In this work, we observe noisy intrinsic source-pair utilities, while stability is defined by socially propagated preferences; consequently, one source observation can affect several players' social comparisons, and the uncertainty driving a matching decision may lie at a different pair from the one being matched.

Bandit learning under network interference. Related work studies bandit learning under network interference, where agents' observed outcomes depend on the actions or exposures of others in the network, including stochastic, adversarial, and contextual settings (Agarwal et al., 2024; Jia et al., 2025; Xu et al., 2025; Jamshidi et al., 2026). Other work examines estimation–regret trade-offs under interference (Zhang & Wang, 2025; Wang et al., 2025) or learns heterogeneous peer influence (Faruk & Zheleva, 2025; Faruk et al., 2026). In our setting, by contrast, observations locally reveal fixed intrinsic source-pair utilities, while the network couples how these utilities enter downstream preference comparisons. By separating decision uncertainty from informative observations, this transforms stable matching into a network-coupled allocation problem.

4 Model and Learning Protocol

Building on the illustration in Section 2, we consider a centralized two-sided matching market with players $\mathcal{U} = [N]$ and arms $\mathcal{V} = [K]$. In an internal talent market, for example, a platform may arrange short-term placements for individual evaluation, followed by peer discussion and a final assignment (Cowgill et al., 2025). We refer to these stages as individual evaluation, social deliberation, and final allocation. We first describe preference formation and stability when intrinsic evaluations are known, and then study how to identify the resulting matching from noisy evaluation feedback.

Social Preference Formation. A player's preference for an arm reflects both their personal evaluation and the opinions they encounter through social interaction. We represent the player's own experience and priorities in the absence of social influence by an intrinsic utility matrix

$\mu \in (0,1]^{N\times K}$, where $\mu_{u,v}$ denotes player u's intrinsic evaluation of arm v. For each arm v, write $\mu_{\cdot,v} = (\mu_{1,v}, \ldots, \mu_{N,v})^\top$ for the vector of intrinsic evaluations across players.

How social interaction shapes these evaluations depends on whose opinions a player considers and how receptive they are to social influence. To describe the sources of influence, let $P \in \mathbb{R}^{N\times N}$ be a nonnegative column-stochastic matrix, where $P_{w,u}$ is the relative weight that player u assigns to player w's opinion within the social component. In particular, $P_{w,u} > 0$ indicates that player w influences player u. Players may differ in their reliance on personal judgment and their tendency to conform to others' views. To capture this heterogeneity, let $\alpha_u \in (0,1]$ denote the weight player u retains on their intrinsic evaluation. A player with $\alpha_u = 1$ relies entirely on their own evaluation, whereas a smaller α_u reflects greater receptiveness to socially transmitted opinions.

We first consider the complete-information scenario, in which the intrinsic evaluations μ are known. As players exchange opinions, they respond to peers whose evaluations may already reflect earlier discussions. We model this iterative process using anchored linear opinion dynamics (DeGroot, 1974; Friedkin & Johnsen, 1990). For each arm v, let $x_{u,v}^{(k)}$ denote player u's evaluation after k social interaction steps, initialized at $x_{\cdot,v}^{(0)} = \mu_{\cdot,v}$. At each step, player u retains weight α_u on their intrinsic evaluation and places the remaining weight on their peers' current evaluations, aggregated according to P. The resulting updates can be written in scalar and vector form as

$$x_{u,v}^{(k+1)} = \alpha_u \mu_{u,v} + (1-\alpha_u) \sum_{w\in\mathcal{U}} P_{w,u} x_{w,v}^{(k)}, \quad x_{\cdot,v}^{(k+1)} = A\mu_{\cdot,v} + (I-A)P^\top x_{\cdot,v}^{(k)}, \tag{1}$$

where $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_N)$. Anchoring each update to $\mu_{u,v}$ preserves the role of personal judgment throughout the exchange of opinions. Symmetrically, social preference formation on the arm side can be modeled by the same dynamics (Appendix B.1).

As players exchange opinions, their evaluations evolve beyond their initial personal judgments. After sufficiently many social interactions, players' evaluations approach a steady state, which we use to represent their preferences over arms. Positive anchoring weights guarantee convergence to a unique steady state. Specifically, let $\alpha_{\min} := \min_{u\in\mathcal{U}} \alpha_u > 0$. Since P is column-stochastic, $\|(I-A)P^\top\|_\infty = 1 - \alpha_{\min} < 1$. The update is therefore a contraction and converges to $x_{\cdot,v}^\star = \Gamma^\top \mu_{\cdot,v}$, where $\Gamma := A\big[I - P(I-A)\big]^{-1}$ is the social propagation operator and I is the identity matrix. For convenience, denote the resulting socially influenced utilities by

$$\theta_{u,v} := x_{u,v}^\star = \sum_{w\in\mathcal{U}} \Gamma_{w,u}\mu_{w,v}, \qquad \Theta = \Gamma^\top \mu. \tag{2}$$

Here, $\Gamma_{w,u}$ captures the overall influence of player w's intrinsic evaluation on player u's limiting evaluation. Operator properties, convergence, and robustness are detailed in Appendices B-C.

Social Stability. The socially influenced utilities determine how players rank potential partners and, consequently, whether they would prefer to leave a given matching. We therefore evaluate stability of the final allocation using the preferences induced by $\Theta = \Gamma^\top \mu$. For each player u, set $\theta_{u,\varnothing} = 0$, where $\varnothing$ denotes remaining unmatched, and define preferences over $\mathcal{V} \cup \{\varnothing\}$ by $a \succ_u^* b$ if and only if $\theta_{u,a} > \theta_{u,b}$, with $a \sim_u^* b$ when the two utilities are equal. For simplicity, we consider strict player-side preferences.

Players need direct experience to evaluate an arm, while arms can select among players using established criteria. Following previous works, we assume the arm-side preferences $\{\succ_v\}_{v\in\mathcal{V}}$ are strict and known to the platform (Liu et al., 2020; 2021; Kong et al., 2022; Sankararaman et al., 2021; Basu et al., 2021; Zhang et al., 2022; Kong & Li, 2023; Wang & Li, 2024; Kong et al., 2024; 2025). Two-sided learning for unknown arm-side preference is discussed in Appendix E.8.

For a matching M, let $M(u)$ and $M^{-1}(v)$ denote the partners of player u and arm v, respectively, using $\varnothing$ for an unmatched participant. A player and an arm may both prefer to match with each other rather than retain their current partners. Formally, a pair (u,v) *blocks* M if $v \succ_u^* M(u)$ and $u \succ_v M^{-1}(v)$. We call M *socially stable* [1] if it admits no blocking pair. Under strict preferences, the player-proposing Gale–Shapley algorithm (Gale & Shapley, 1962), applied to the induced preferences, returns the unique player-optimal socially stable matching, denoted by $M^\star := M^{\mathrm{opt}}(\mu;\Gamma)$.

[1] Unlike Askalidis et al. (2013), our social graph shapes preferences without restricting blocking pairs.

Information Structure. Practically, the influence players exert on one another may depend on the strength of their social ties. These established relationships, together with players' confidence in their own judgments, can often be specified before matching, motivating known knowledge of the influence network P and the anchoring parameters A. In contrast, players' preferences over unfamiliar arms are typically unknown and must be learned through matching experiences. Accordingly, our main analysis treats P and A, and hence Γ, as fixed and known before social deliberation begins, while the intrinsic utilities μ must be learned from individual evaluations. The more general setting where P is unknown is deferred to Appendix C.4.

Pure Exploration. We study fixed-confidence identification of $M^\star$, the player-optimal socially stable matching induced by the social utilities $\Theta = \Gamma^\top \mu$, using as few individual evaluations as possible. During the individual evaluation phase, players evaluate their assigned arms and report feedback to the platform without exchanging evaluations with one another. At each round t, the platform uses the known social network, past feedback, and any internal randomization to select an evaluation matching $M_t \subseteq U \times V$, assigning each player to at most one arm and each arm to at most one player. These evaluation matchings need not be socially stable (Athanasopoulos et al., 2026). For each $(w, v) \in M_t$, player w observes $Y_t(w, v) = \mu_{w,v} + \eta_t(w, v)$, where the noise variables $\eta_t(w, v)$ are independent, mean-zero, and 1-sub-Gaussian. Let $N_{w,v}(t)$ denote the number of evaluations of (w, v) through round t, and let $\hat{\mu}_{w,v}(t)$ be the empirical mean of the corresponding observations, set to zero when $N_{w,v}(t) = 0$.

The platform ends the individual evaluation phase at a stopping time τ determined by the evaluation history and its internal randomization. Players then enter social deliberation, using $\hat{\mu}(\tau)$ as both initial evaluations and fixed anchors in the dynamics introduced above. Through interactions over the known social network, their evaluations converge to $\hat{\Theta}_\tau = \Gamma^\top \hat{\mu}(\tau)$. Then the final matching $\hat{M}_\tau$ is obtained by applying the Gale–Shapley algorithm (Gale & Shapley, 1962) to the preferences.

For a prescribed error tolerance $\delta \in (0, 1/2)$, the procedure is δ-correct if for every intrinsic utility matrix $\mu \in (0, 1]^{N \times K}$ inducing strict social preferences, the evaluation phase terminates almost surely and the final matching satisfies $\Pr_\mu\left(\hat{M}_\tau = M^\star\right) \geq 1 - \delta$. We measure exploration cost by the total number of individual evaluations,

$$C_\tau := \sum_{t=1}^{\tau} |M_t| = \sum_{w \in \mathcal{U}} \sum_{v \in \mathcal{V}} N_{w,v}(\tau).$$

The objective is to minimize $\mathbb{E}_\mu[C_\tau]$ over δ-correct procedures. Probability and expectation are taken over the evaluation noise and the platform's internal randomization.

5 Learning Challenges

5.1 Source-Destination Mismatch

During the individual evaluation phase, feedback is used to estimate intrinsic preferences. Identifying a socially stable matching may require observations from different pairs, or greater precision on the same pairs, than just identifying a stable matching under intrinsic preferences. For a player u, the social comparison between arms a and b satisfies $\theta_{u,a} - \theta_{u,b} = \sum_{w \in \mathcal{U}} \Gamma_{w,u}\big(\mu_{w,a} - \mu_{w,b}\big)$. Social influence therefore creates a dependence across players: resolving one player's social preference may require refining another player's intrinsic evaluations. The following example shows how this dependence can leave the socially stable matching unresolved even after all intrinsic rankings have been determined.

Example 1 (Source-destination mismatch)**.** *Consider a market with two players* $1, 2$ *and two arms* a, b*, both of which prefer player* 1 *to player* 2*. The social propagation operator is*

$$\Gamma = \begin{pmatrix} 1/2 & 0 \\ 1/2 & 1 \end{pmatrix}.$$

Suppose that all intrinsic means are known except $\mu_{2,a}$*. The current estimates and error bound are*

$$\hat{\mu} = \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.5 \end{pmatrix}, \qquad |\mu_{2,a} - 0.3| \leq \epsilon, \quad \textit{where } \epsilon \in (0, 0.1).$$

Throughout this interval, the unique player-optimal stable matching under intrinsic preferences is $\{(1,a),(2,b)\}$. *Under strict social preferences, the unique player-optimal socially stable matching is*

$$M^\star = \begin{cases} \{(1,a),(2,b)\}, & \mu_{2,a} > 0.3, \\ \{(1,b),(2,a)\}, & \mu_{2,a} < 0.3. \end{cases}$$

In Example 1, the estimate of $\mu_{2,a}$ is already precise enough to determine player 2's intrinsic preference. Its remaining uncertainty, however, can reverse player 1's social comparison: $\theta_{1,a} - \theta_{1,b} = \frac{1}{2}(\mu_{2,a} - 0.3) \in [-\epsilon/2,\, \epsilon/2]$. Simply responding to this unresolved comparison by collecting more evaluations of $(1,a)$ and $(1,b)$ cannot reduce the remaining uncertainty, since both intrinsic means are already known exactly. Resolving the comparison requires further evaluation of $(2,a)$, even though player 2's intrinsic preference is already determined.

Here, player 1 is the destination whose social preference remains unresolved, while player 2 is the source of the required information. This *source–destination mismatch* means that locating an uncertain social comparison does not by itself identify the pairs that should be explored. The sampling strategy must also account for how source evaluations contribute to that comparison through the network. This leads to the central questions of which source pairs are needed to identify the target matching and how many evaluations each requires.

5.2 Sample Complexity Lower Bound

The source–destination mismatch discussed in Section 5.1 raises the question of which intrinsic pairs must be sampled, and to what precision, to identify the player-optimal socially stable matching. We study this question by establishing a sample complexity lower bound.

To derive the lower bound, we first consider comparisons that are necessary for the stability of the target matching. If an arm prefers a player to its current partner, that player must prefer their assigned arm to the alternative; otherwise, the two would form a blocking pair. Identifying the target matching therefore requires sufficient observations of the intrinsic utilities that determine these comparisons. We refer to such comparisons as *blocking-critical*, as defined below.

Definition 1 (Blocking-critical comparisons)**.** *Let* $a_u := M^{\mathrm{opt}}(\mu)(u)$ *denote the partner assigned to player* u *in the player-optimal socially stable matching. A pair* $(u,b) \in \mathcal{U} \times \mathcal{V}$ *with* $b \neq a_u$ *is* blocking-critical *if arm* b *strictly prefers* u *to its partner under* $M^{\mathrm{opt}}(\mu)$. *Denote the set of such pairs by* $\mathcal{C}_{\mathrm{blk}}(\mu)$. *Under strict social preferences, stability requires* $\theta_{u,a_u}(\mu) > \theta_{u,b}(\mu)$ *for every* $(u,b) \in \mathcal{C}_{\mathrm{blk}}(\mu)$. *We define the gap as* $\Delta_{u,b} := \theta_{u,a_u}(\mu) - \theta_{u,b}(\mu) > 0$, *where* $(u,b) \in \mathcal{C}_{\mathrm{blk}}(\mu)$.

Reversing a blocking-critical comparison makes the target matching unstable. An alternative instance may also leave this matching stable while it is no longer player-optimal. Identifying the player-optimal socially stable matching requires distinguishing the true instance from both types of alternatives. Let $\mathcal{D}_\Gamma$ denote the intrinsic utility matrices in $(0,1]^{N\times K}$ that induce strict social preferences, with the social operator and arm preferences fixed, write $\mathrm{Alt}_{\mathrm{opt}}(\mu) := \{\lambda \in \mathcal{D}_\Gamma : M^{\mathrm{opt}}(\lambda) \neq M^{\mathrm{opt}}(\mu)\}$.

Theorem 1 (Source-level sample complexity lower bound)**.** *Fix* $\mu \in \mathcal{D}_\Gamma$ *and assume independent Gaussian observations* $X_{w,v} \sim \mathcal{N}(\mu_{w,v}, 1)$. *Let* $\delta \in (0, 1/2)$. *Consider any algorithm that stops almost surely and identifies* $M^{\mathrm{opt}}(\lambda)$ *with probability at least* $1-\delta$ *for every* $\lambda \in \mathcal{D}_\Gamma$. *Let* $n_{w,v} := \mathbb{E}_\mu[N_{w,v}(\tau)]$ *denote the expected number of observations from source pair* (w,v). *If these expected counts are finite, then*

$$\frac{1}{2}\sum_{w\in\mathcal{U}}\sum_{v\in\mathcal{V}} n_{w,v}(\mu_{w,v} - \lambda_{w,v})^2 \geq \mathrm{kl}(1-\delta,\delta), \qquad \lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu), \tag{3}$$

where kl *denotes binary relative entropy. Consequently,*

$$\mathbb{E}_\mu\left[\sum_{w,v} N_{w,v}(\tau)\right] \geq T^\star_{\mathrm{opt}}(\mu,\delta),$$

where $T^\star_{\mathrm{opt}}(\mu,\delta)$ *is the infimum total allocation* $\sum_{w,v} x_{w,v}$ *over all* $x \geq 0$ *satisfying equation 3 with* $n = x$ *for every* $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$. *The total-cost bound also holds when the expected cost is infinite.*

For a blocking-critical comparison, the information constraint gives an explicit relation between source counts and the social gap. For player u and distinct arms a, b, define $V_{u,a,b}(x) := \sum_{w\in\mathcal{U}} \Gamma_{w,u}^2 \left(x_{w,a}^{-1} + x_{w,b}^{-1}\right)$. If the expected counts are finite and positive on the influence support, and the unconstrained minimum-cost reversal satisfies the box constraints specified in Corollary 1, then

$$V_{u,a_u,b}(n) \le \left(\Delta_{u,b}^2 / 2\,\mathrm{kl}(1-\delta,\delta)\right). \tag{4}$$

The dependence on $\Gamma_{w,u}$ reflects the contribution of source w to player u's social comparison. The full bound also accounts for alternatives that preserve stability but change player-optimality. The detailed characterization and the proofs are given in Appendix D.1.

6 Source-Aware Learning

For blocking-critical comparison between arms a and b, the information constraint in equation 4 shows that certifying player u's preference requires reducing the propagated variance $\sum_w \Gamma_{w,u}^2 (n_{w,a}^{-1} + n_{w,b}^{-1})$ below the squared gap $\Delta_{u,b}^2/(2\mathrm{kl}(1-\delta,\delta))$. Since preferences are mediated by the social network, resolving an uncertain comparison for player u often hinges on exploring the intrinsic utilities of other players w with large $\Gamma_{w,u}$. Consequently, exploration cannot simply sample pairs that appear locally competitive; it must actively collect observations from the source pairs that govern these weighted variances. Moreover, since evaluating a pair (w, v) simultaneously reduces variance for every player influenced by w, sample allocation must be coordinated globally across the network. We implement this joint allocation in Source-Aware Elimination with Quadratic Allocation [2] (SAE-Q, Algorithm 1). Let $d := NK$ and $\mathcal{C} := \{(u, a, b) : u \in \mathcal{U},\ a, b \in \mathcal{V},\ a < b\}$, which contains all unordered arm comparisons for every player. For $c = (u, a, b) \in \mathcal{C}$, define $g_c := \theta_{u,a} - \theta_{u,b}$ and $\Delta_c := |g_c|$. Write $V_c(x) := V_{u,a,b}(x)$ for $c = (u, a, b)$. Also write N^m and $\widehat{\mu}^m$ for the matrices of cumulative evaluation counts and empirical means after epoch m.

Algorithm 1 proceeds in epochs $m = 1, 2, \ldots$. At each epoch, it (i) tightens the target accuracy for the unresolved comparisons, (ii) computes the cheapest additional source allocation achieving these accuracy requirements, (iii) collects the prescribed evaluations, and (iv) eliminates comparisons whose signs can be certified. Let $\mathcal{A}_m$ denote the active comparisons and set $\epsilon_m := 2^{-m}$. Active comparisons are therefore required to reach precision ϵ_m.

A small complication arises because all social gaps are recomputed from cumulative empirical means. Even after a comparison has been certified, future samples collected for other comparisons may change its empirical gap. Hence, unlike classical elimination schemes that can discard a resolved comparison entirely, SAE-Q retains its last certified precision. If comparison c is eliminated at epoch j_c, we set $\rho_c^m := \epsilon_m$ when $c \in \mathcal{A}_m$; otherwise $\rho_c^m := \epsilon_{j_c}$. Given cumulative counts N^{m-1}, the algorithm then computes the minimum additional source demand

$$n^m \in \arg\min_{n\ge 0} \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} n_{w,v} \quad \text{s.t.} \quad V_c(N^{m-1} + n) \le \frac{(\rho_c^m)^2}{2\beta_m}, \quad \forall c \in \mathcal{C}, \tag{5}$$

where $\beta_m = B_0(m+1)$ is a deterministic confidence threshold ensuring simultaneous coverage across comparisons and epochs. The coefficient B_0 depends only on N, K, Γ, and δ, with its explicit form given in Appendix E. The resulting integer demand $\lceil n^m \rceil$ is realized by Schedule (Algorithm 2) as a sequence of collision-free evaluation matchings. After updating the empirical utilities, SAE-Q eliminates every active comparison whose estimated gap is sufficiently separated from zero, and repeats until no active comparisons remain.

We characterize the instance difficulty through the full-ranking allocation complexity

$$H_2(\mu) := \min_{x>0} \left\{ \sum_{w,v} x_{w,v} : V_c(x) \le \Delta_c^2,\ \forall c \in \mathcal{C} \right\}, \tag{6}$$

[2] Here, "quadratic" refers to the squared social influence weights in the propagated variance constraints.

Algorithm 1 SAE-Q: Source-Aware Elimination with Quadratic Allocation

Require: Confidence level $\delta \in (0, 1/2)$, known Γ, and known arm-side preferences.

1: Evaluate every source pair once using SCHEDULE; set $N^0 = \mathbf{1}$ and $\mathcal{A}_1 = \mathcal{C}$.
2: **for** $m = 1, 2, \ldots$ **do**
3: Set $\epsilon_m = 2^{-m}$ and $\beta_m = B_0(m+1)$.
4: For every $c \in \mathcal{C}$, set

$$\rho_c^m = \begin{cases} \epsilon_m, & c \in \mathcal{A}_m, \\ \epsilon_{j_c}, & c \notin \mathcal{A}_m, \end{cases}$$

where j_c is the elimination epoch of c.
5: Choose n^m by solving

$$\min_{n \geq 0} \sum_{w,v} n_{w,v} \quad \text{s.t.} \quad V_c(N^{m-1} + n) \leq \frac{(\rho_c^m)^2}{2\beta_m}, \quad c \in \mathcal{C}.$$

6: Execute SCHEDULE($\lceil n^m \rceil$) and update the cumulative counts N^m and empirical means $\widehat{\mu}^m$.
7: For each $c = (u, a, b)$, compute $\widehat{g}_c^m = \sum_w \Gamma_{w,u}(\widehat{\mu}_{w,a}^m - \widehat{\mu}_{w,b}^m)$.
8: Update $\mathcal{A}_{m+1} = \{c \in \mathcal{A}_m : |\widehat{g}_c^m| \leq 2\epsilon_m\}$, $j_c = m$ for every newly eliminated comparison.
9: **if** $\mathcal{A}_{m+1} = \varnothing$ **then**
10: Form $\widehat{\Theta} = \Gamma^\top \widehat{\mu}^m$.
11: **return** player-proposing Gale–Shapley applied to $\widehat{\Theta}$, with the outside option last.
12: **end if**
13: **end for**

which captures the minimum source-level budget required to resolve all social comparisons at their gap scales while accounting for sample sharing through the network.

Theorem 2 (SAE-Q Sample Complexity)**.** *Assume $K \geq 2$, independent 1-sub-Gaussian noise and strict social preferences under $\mu \in \mathcal{D}_\Gamma$. Let $\Delta_{\min} := \min_{c \in \mathcal{C}} \Delta_c$, $M_\Delta := \lceil \log_2(6/\Delta_{\min}) \rceil$, and*

$$F_Q := d(M_\Delta + 1) + 36 B_0 H_2(\mu) M_\Delta (M_\Delta + 3),$$

where B_0 is a deterministic constant depending only on N, K, Γ and δ; its explicit form is given in Appendix E. Then Algorithm 1 stops almost surely and is δ-correct. With probability at least $1 - \delta$,

$$\widehat{M}_\tau = M^\star, \qquad C_\tau \leq F_Q.$$

Moreover, $\mathbb{E}_\mu[C_\tau] \leq 2F_Q$.

Theorem 2 shows that all comparisons are certified within $O(\log(1/\Delta_{\min}))$ epochs, while the dominant instance-dependent cost is governed by $H_2(\mu)$. Thus, SAE-Q exploits the network by allocating samples to source pairs that jointly reduce uncertainty across multiple social comparisons.

Remark 1 (Full ranking vs. Matching Identification)**.** *SAE-Q certifies all pairwise social comparisons, which can require substantially more samples than identifying $M^\star$. Appendix F formalizes this discrepancy in sample requirements and characterizes instances where matching identification requires strictly fewer samples than global ranking resolution. To exploit this structural difference, we develop a matching-critical variant (SAE-Q-SD) under a serial-dictatorship structure, proving that sample complexity can be decoupled from irrelevant gaps and governed solely by matching-critical comparisons. Numerical simulations in Appendix F.6 corroborate this gap decoupling alongside the sample efficiency of quadratic allocation across diverse networks.*

7 CONCLUSION

We studied pure-exploration stable matching under social influence, formalizing the source–destination mismatch that decouples decision uncertainty from informative observations. We established an instance-dependent lower bound under network variance constraints and designed SAE-Q, an elimination algorithm that provably recovers the target matching. Furthermore, we demonstrated that matching-specific exploration under common priorities can decouple sample complexity from irrelevant gaps. Future work includes developing polynomial-sized certificates for general priorities and exploring decentralized market protocols under network interference.

AI USE STATEMENT

We used generative AI tools for three purposes: (i) running the LLM-agent simulation in Section 2 and Appendix A, which produces synthetic preference trajectories used only as a motivating illustration; (ii) language polishing. The authors reviewed all AI-assisted outputs and take full responsibility for the content of the paper.

ETHICS STATEMENT

This work is primarily theoretical and studies pure-exploration learning of stable matchings under social influence. It involves no human subjects, crowdsourcing, or personal data. The motivating simulation in Section 2 and Appendix A uses LLM agents with synthetic personas and fictitious companies. Its outputs are synthetic trajectories that serve only as illustration and are not empirical evidence about human behavior. LLMs were accessed through a commercial API under the provider's terms of use. Because our target is stability with respect to socially influenced preferences, applying the method to real matching markets (e.g., hiring or school admission) calls for care: stability does not imply fairness, and social influence can amplify herding or existing biases. Practitioners should therefore audit the influence model and the resulting matchings for disparate impact. We are not aware of any conflicts of interest or sponsorship that bear on this work.

REPRODUCIBILITY STATEMENT

The model, assumptions, and algorithms are stated in Sections 3–5, with pseudocode in Algorithm 1 (SAE-Q) and Algorithm 3 (SAE-Q-SD). Complete proofs are in the appendix: properties of the social operator in Appendix B, convergence, robustness, and calibration in Appendix C, the lower bounds in Appendix D, the guarantees for SAE-Q in Appendix E, and the guarantees for SAE-Q-SD in Appendix F. Appendix F.6 describes the numerical setup, including instance generation, network topologies, baselines, and the practical confidence schedules. The anonymized supplementary material provides the code, instance seeds, raw result files, and scripts that regenerate Figure 4. For the LLM simulation, which serves only as a motivating illustration, Appendix A gives the company descriptions, personas, and all prompt templates.

REFERENCES

Abhineet Agarwal, Anish Agarwal, Lorenzo Masoero, and Justin Whitehouse. Multi-armed bandits with network interference. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Elliot Anshelevich, Onkar Bhardwaj, and Martin Hoefer. Stable matching with network externalities. *Algorithmica*, 78(3):1067–1106, 2017.

Esteban Arcaute and Sergei Vassilvitskii. Social networks and stable matchings in the job market. In *International Workshop on Internet and Network Economics*, pp. 220–231. Springer, 2009.

Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351, 2023.

Nick Arnosti, Ramesh Johari, and Yash Kanoria. Managing congestion in matching markets. *Manufacturing & Service Operations Management*, 23(3):620–636, 2021.

Ariel Flint Ashery, Luca Maria Aiello, and Andrea Baronchelli. Emergent social conventions and collective bias in LLM populations. *Science Advances*, 11(20):eadu9368, 2025.

Georgios Askalidis, Nicole Immorlica, and Emmanouil Pountourakis. Socially stable matchings. *arXiv preprint arXiv:1302.3309*, 2013.

Andreas Athanasopoulos, Anne-Marie George, and Christos Dimitrakakis. Probably correct optimal stable matching for two-sided market under uncertainty. In *Proceedings of the 24th International*

Conference on Autonomous Agents and Multiagent Systems, AAMAS '25, pp. 152–160, Richland, SC, 2025. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9798400714269.

Andreas Athanasopoulos, Anne-Marie George, and Christos Dimitrakakis. Probably correct optimal stable matching under two-sided uncertainty. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 228–254. PMLR, 2026.

Mariagiovanna Baccara, Ayşe İmrohoroğlu, Alistair J Wilson, and Leeat Yariv. A field study on matching with network externalities. *American Economic Review*, 102(5):1773–1804, 2012.

Keisuke Bando, Ryo Kawasaki, and Shigeo Muto. Two-sided matching with externalities: A survey. *Journal of the Operations Research Society of Japan*, 59(1):35–71, 2016.

Soumya Basu. Competing bandits in matching markets via super stability. In *Forty-second International Conference on Machine Learning*, 2025.

Soumya Basu, Karthik Abinav Sankararaman, and Abishek Sankararaman. Beyond $\log^2(t)$ regret for decentralized bandits in matching markets. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 705–715. PMLR, 2021.

Elizabeth Bodine-Baron, Christina Lee, Anthony Chong, Babak Hassibi, and Adam Wierman. Peer effects and stability in matching markets. In *Proceedings of the 4th international conference on Algorithmic game theory*, pp. 117–129, 2011.

Anna Bykhovskaya. Stability in matching markets with peer effects. *Games and Economic Behavior*, 122:28–54, 2020.

Mei Cai, Suqiong Hu, Ya Wang, and Jingmei Xiao. A dynamic social network matching model for virtual power plants and distributed energy resources with probabilistic linguistic information. *Sustainability*, 14(22):14920, 2022.

Antoni Calvo-Armengol and Matthew O Jackson. The effects of social networks on employment and inequality. *American economic review*, 94(3):426–454, 2004.

Christopher Campos. Social interactions, information, and preferences for schools: Experimental evidence from Los Angeles. NBER Working Paper 33010, National Bureau of Economic Research, 2024. URL https://doi.org/10.3386/w33010.

Sung-Ho Cho, Taiki Todo, and Makoto Yokoo. Two-sided matching over social networks. In Luc De Raedt (ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*, pp. 186–193. International Joint Conferences on Artificial Intelligence Organization, July 2022. doi: 10.24963/ijcai.2022/27. URL https://doi.org/10.24963/ijcai.2022/27. Main Track.

Bo Cowgill, Jonathan MV Davis, B Pablo Montagnes, and Patryk Perkowski. Stable matching on the job? theory and evidence on internal talent markets. *Management Science*, 71(3):2508–2526, 2025.

Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical association*, 69(345): 118–121, 1974.

Umut Mert Dur and Thomas Wiseman. School choice with neighbors. *Journal of Mathematical Economics*, 83:101–109, 2019.

Andrew Dustan. Family networks and school choice. *Journal of Development Economics*, 134: 372–391, 2018.

Bhaskar Dutta and Jordi Massó. Stability of matchings when individuals have preferences over colleagues. *journal of Economic Theory*, 75(2):464–475, 1997.

Federico Echenique and M Bumin Yenmez. A solution to matching with preferences over colleagues. *Games and Economic Behavior*, 59(1):46–71, 2007.

Ahmed Sayeed Faruk and Elena Zheleva. Leveraging heterogeneous spillover in maximizing contextual bandit rewards. In *Proceedings of the ACM Web Conference 2025*, pp. 3049–3060. Association for Computing Machinery, 2025. doi: 10.1145/3696410.3714706.

Ahmed Sayeed Faruk, Mohammad Shahverdikondori, and Elena Zheleva. Learning peer influence probabilities with linear contextual bandits. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1073–1084. Association for Computing Machinery, 2026. doi: 10.1145/3770855.3817733.

Noah E Friedkin and Eugene C Johnsen. Social influence and opinions. *Journal of mathematical sociology*, 15(3-4):193–206, 1990.

Noah E Friedkin, Anton V Proskurnikov, Roberto Tempo, and Sergey E Parsegov. Network science on belief system dynamics under logic constraints. *Science*, 354(6310):321–326, 2016.

Noah E Friedkin, Anton V Proskurnikov, Wenjun Mei, and Francesco Bullo. Mathematical structures in group decision-making on resource allocation distributions. *Scientific Reports*, 9(1):1377, 2019.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American mathematical monthly*, 69(1):9–15, 1962.

Chen Gao, Xiaochong Lan, Zhihong Lu, Jinzhu Mao, Jinghua Piao, Huandong Wang, Depeng Jin, and Yong Li. S3: Social-network simulation system with large language model-empowered agents. *arXiv preprint arXiv:2307.14984*, 2023.

Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. Large language models empowered agent-based modeling and simulation: a survey and perspectives. *Humanities and Social Sciences Communications*, 11:1259, 2024. doi: 10.1057/s41599-024-03611-3.

Mark S Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973.

Hanna Halaburda, Mikołaj Jan Piskorski, and Pınar Yıldırım. Competing by restricting choice: The case of matching platforms. *Management Science*, 64(8):3574–3594, 2018.

Günter J. Hitsch, Ali Hortaçsu, and Dan Ariely. Matching and sorting in online dating. *American Economic Review*, 100(1):130–163, 2010. doi: 10.1257/aer.100.1.130.

Martin Hoefer. Local matching dynamics in social networks. *Information and Computation*, 222: 20–35, 2013.

Martin Hoefer and Lisa Wagner. Locally stable marriage with strict preferences. *SIAM Journal on Discrete Mathematics*, 31(1):283–316, 2017. doi: 10.1137/151003854. URL https://doi.org/10.1137/151003854.

Hadi Hosseini, Sanjukta Roy, and Duohan Zhang. Putting Gale & Shapley to work: guaranteeing stability through learning. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9798331314385.

Meena Jagadeesan, Alexander Wei, Yixin Wang, Michael I Jordan, and Jacob Steinhardt. Learning equilibria in matching markets with bandit feedback. *Journal of the ACM*, 70(3):1–46, 2023.

Fateme Jamshidi, Mohammad Shahverdikondori, and Negar Kiyavash. Graph-dependent regret bounds in multi-armed bandits with interference. In *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 2442–2460. PMLR, 2026.

Su Jia, Peter Frazier, and Nathan Kallus. Multi-armed bandits with interference: bridging causal inference and adversarial bandits. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025.

Yash Kanoria and Daniela Saban. Facilitating the search for partners on matching platforms. *Management Science*, 67(10):5990–6029, 2021.

Alexander S Kelso Jr and Vincent P Crawford. Job matching, coalition formation, and gross substitutes. *Econometrica: Journal of the Econometric Society*, pp. 1483–1504, 1982.

Fang Kong and Shuai Li. Player-optimal stable regret for bandit learning in matching markets. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1512–1522. SIAM, 2023.

Fang Kong, Junming Yin, and Shuai Li. Thompson sampling for bandit learning in matching markets. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pp. 3164–3170. International Joint Conferences on Artificial Intelligence Organization, 2022.

Fang Kong, Zilong Wang, and Shuai Li. Improved analysis for bandit learning in matching markets. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, pp. 91904–91929, 2024.

Fang Kong, Jingqi Tang, Mingzhu Li, Pinyan Lu, John C.S. Lui, and Shuai Li. Bandit learning in matching markets with indifference. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=7ENakslm9J.

Fang Kong, Xiaoxi Zhang, Xiao Huang, Lijun Zhang, and Shuai Li. Bandit learning in matching markets with relative feedback. *Theoretical Computer Science*, 1071:115846, 2026. ISSN 0304-3975. doi: https://doi.org/10.1016/j.tcs.2026.115846. URL https://www.sciencedirect.com/science/article/pii/S0304397526001052.

Miao Li, Xinwei Song, and Dengji Zhao. Stable marriage on networks. In Christopher Amato, Louise Dennis, Viviana Mascardi, and John Thangarajah (eds.), *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026)*, pp. 1910–1918, Paphos, Cyprus, May 2026. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 979-8-4007-2317-9. doi: 10.65109/PKVJ4128. URL https://doi.org/10.65109/PKVJ4128.

Shiyun Lin, Simon Mauras, Nadav Merlis, and Vianney Perchet. Stable matching with ties: Approximation ratios and learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Lydia T. Liu, Horia Mania, and Michael I. Jordan. Competing bandits in matching markets. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 1618–1628. PMLR, 2020.

Lydia T Liu, Feng Ruan, Horia Mania, and Michael I Jordan. Bandit learning in decentralized matching markets. *Journal of Machine Learning Research*, 22(211):1–34, 2021.

Chinmay Maheshwari, Shankar Sastry, and Eric Mazumdar. Decentralized, communication-and coordination-free learning in structured matching markets. *Advances in Neural Information Processing Systems*, 35:15081–15092, 2022.

Yifei Min, Tianhao Wang, Ruitu Xu, Zhaoran Wang, Michael I Jordan, and Zhuoran Yang. Learn to match with no regret: reinforcement learning in markov matching markets. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, pp. 19956–19970, 2022.

Tejas Pagare, Agniv Bandyopadhyay, and Sandeep Juneja. Optimal algorithms for bandit learning in matching markets. *arXiv preprint arXiv:2509.14466*, 2025.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.

Sergey E. Parsegov, Anton V. Proskurnikov, Roberto Tempo, and Noah E. Friedkin. Novel multidimensional models of opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, 62(5):2270–2285, 2017. doi: 10.1109/TAC.2016.2613905.

Alvin E Roth. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of political Economy*, 92(6):991–1016, 1984.

Alvin E Roth. What have we learned from market design? *The economic journal*, 118(527):285–310, 2008.

Abishek Sankararaman, Soumya Basu, and Karthik Abinav Sankararaman. Dominate or delete: Decentralized competing bandits in serial dictatorship. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 1252–1260. PMLR, 2021.

Hiroo Sasaki and Manabu Toda. Two-sided matching problems with externalities. *Journal of Economic Theory*, 70(1):93–108, 1996.

Zichen Wang, Haoyang Hong, Chuanhao Li, Haoxuan Li, Zhiheng Zhang, and Huazheng Wang. Design-based bandits under network interference: Trade-off between regret and statistical inference. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Zilong Wang and Shuai Li. Optimal analysis for bandit learning in matching markets with serial dictatorship. *Theoretical Computer Science*, 1010:114703, 2024.

Yang Xu, Wenbin Lu, and Rui Song. Linear contextual bandits with interference. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 69545–69582. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/xu25v.html.

Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, and Yang Liu. Exploring large language models for communication games: An empirical study on werewolf. *arXiv preprint arXiv:2309.04658*, 2023.

Qing Yang, Xinshang You, and Yiye Zhang. Two-sided matching based on I-BTM and LSGDM applied to high-level overseas talent and job fit problems. *Scientific Reports*, 11(1):12723, 2021.

YiRui Zhang and Zhixuan Fang. Decentralized two-sided bandit learning in matching market. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024.

Yirui Zhang, Siwei Wang, and Zhixuan Fang. Matching in multi-arm bandit with collision. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, pp. 9552–9563, 2022.

Zhiheng Zhang and Zichen Wang. Online experimental design with estimation-regret trade-off under network interference. In *Advances in Neural Information Processing Systems*, volume 38, pp. 17946–17986, 2025. doi: 10.52202/085713-0538.

APPENDIX CONTENTS

A LLM Simulation Setup and Prompts **16**

A.1 Company Descriptions 16

A.2 Player Configurations 16

A.3 Simulation Prompts 17

A.3.1 Initial Preference Generation 17

A.3.2 Comment Generation (Express) 18

A.3.3 Format Observation (Observe) 18

A.3.4 Preference Update (Update) 18

B Arm-Side Model and Social Influence Operator **19**

B.1 Symmetric Arm-Side Construction 19

B.2 Proof of Γ's Properties 19

B.3 Auxiliary properties of Γ 20

B.4 Discussion of modeling choices 21

C Convergence, Robustness, and Calibration of Social Preferences **21**

C.1 Convergence and Induced Stable Matchings 22

C.2 Convergence in Special Network Topologies 22

C.3 Robustness to Misspecification of (P, A) 23

C.4 Calibrating an Unknown Social Operator 24

D Information Lower Bounds for Socially Stable Matching **27**

D.1 Lower bound: proof and characterization 27

D.1.1 Setup and information complexity 27

D.1.2 Proof of Theorem 1 29

D.1.3 Stability-Loss and Optimality-Loss Alternatives 32

D.1.4 Exact characterization by bounded quadratic programs 33

D.1.5 Explicit constraints for blocking comparisons 37

D.1.6 Blocking, Matching, and Ranking Complexities 39

D.2 Extension to matching-round complexity 39

E Analysis and Extensions of SAE-Q **40**

E.1 Bipartite scheduling 41

E.2 Existence of the allocation optimizers 42

E.3 Uniform concentration under adaptive evaluation 42

E.4 Deterministic count and confidence-budget bounds 43

E.5 Persistence of Eliminated Comparisons 44

E.6 Termination and evaluation cost on bounded-error paths 45

E.7 Proof of Theorem 2 45
E.8 Extension to Two-Sided Unknown Preferences 46

F Matching-Critical Exploration and Ranking Overhead **48**
F.1 Overview and Notation 48
F.2 Why complete rankings can be arbitrarily wasteful 49
F.3 The comparisons that determine the matching 50
F.4 SAE-Q-SD: discovery, compression, and persistence 51
F.5 SAE-Q-SD Guarantee and the Separation Example 53
F.6 Numerical Illustrations 54
F.7 Proof of the main guarantee 57
F.8 When the certificate matches the information lower bound 59
F.9 Scope, Overhead, and General Priorities 60

A LLM Simulation Setup and Prompts

This appendix provides the supplementary details for the LLM-based agent simulation. It includes the specific parameters defining the company environments, the distinct personality profiles of the simulated agents, and the exact prompt templates used to drive the iterative simulation pipeline.

A.1 Company Descriptions

In our simulation, job offers are modeled as five distinct company "arms". Each arm is characterized by five core dimensions: Salary, Prestige, Work-life Balance, Risk, and Growth. As shown in Table 1, these dimensions are specifically designed with inherent trade-offs (e.g., high salary and prestige but low work-life balance) to force agents to make non-trivial decisions that reflect their underlying preferences.

Table 1: Company Arm Descriptions

Arm	Salary	Prestige	Work-life Balance	Risk	Growth
Arm 1	very high	very high	low	low	high
Arm 2	medium	medium	very high	very low	medium
Arm 3	low	low	low	very high	very high
Arm 4	medium	medium	high	very low	low
Arm 5	medium	high	medium	low	high

A.2 Player Configurations

To create a diverse and realistic applicant pool, we construct five heterogeneous agent personas. Each persona is defined not only by their primary and secondary career priorities, but also by their social susceptibility and communication styles. These configurations, detailed in Table 2, act as the core instructions for the LLMs, dictating how they independently evaluate companies and interact with their network neighbors.

Table 2: Player Configurations and Profiles

Type	Profile Setup	Social Behavior	Decision Style	Comm. Style
The Achiever	Primary: career growth Secondary: salary, innovation Risk: high	1. Relies mostly on own judgment 2. Not easily influenced by others	1. Willing to accept high pressure for faster growth 2. Focuses on long-term career benefits	1. Expresses strong opinions 2. Can influence others in discussions
The Conformist	Primary: prestige Secondary: salary, stability Risk: low	1. Strongly influenced by majority opinion 2. Adjusts preferences based on peers	1. Follows popular or well-known choices 2. Avoids unconventional options	1. Rarely challenges others 2. Aligns with group consensus
The Risk-Averse	Primary: stability Secondary: work-life balance, benefits Risk: very low	1. Cautious about others' opinions 2. Prefers safe and proven choices	1. Avoids high-risk, high-uncertainty options 2. Prioritizes security over growth	1. Reserved in discussions 2. Does not strongly influence others
The Altruist	Primary: social impact Secondary: ethics, company mission Risk: medium	1. Considers others' perspectives 2. Values collective well-being	1. Prefers ethical and socially responsible companies 2. Willing to sacrifice salary for impact	1. Shares values openly 2. Encourages others to consider ethics
The Independent	Primary: personal interest Secondary: autonomy, creativity Risk: medium-high	1. Ignores peer pressure 2. Makes decisions independently	1. Follows intrinsic motivation 2. Values freedom over structure	1. Speaks honestly 2. Does not try to influence others

A.3 Simulation Prompts

The simulation operates through a structured pipeline governed by specific prompt templates. The process follows an initialization step followed by an iterative "Express–Observe–Update" cycle.

A.3.1 Initial Preference Generation

Before placing agents into the social network, this prompt requires the LLM to generate a baseline preference vector. The prompt strictly enforces that the agent scores the companies based solely on their defined persona rather than generic human preferences.

```
Task: Simulate a job-seeker's initial preference scores for 5 different
    companies based on their personality profile and the companies'
    attributes.

{profile_text}

[Companies]
{chr(10).join(arm_descriptions)}

[Requirement]
- Compare companies dimension by dimension
```

```
- Focus on primary priority
- Make clear distinctions
- Use the FULL range from -5 to 5.
- Strong matches should be above 3.
- Strong mismatches should be below -3.
- Avoid clustering scores in a narrow range.
- Avoid tie scores for different arms.
- The agent MUST prioritize its primary value over all other factors,
    even if it leads to lower salary or prestige. Do not behave like an
    average person. Behave in a way that strongly reflects the
    personality type. Ignore general human preferences like comfort,
    stability, or salary unless they are explicitly part of the
    personality.

[Output]
Return ONLY JSON:
{{"scores": [float, float, float, float, float]}}
The agent MUST prioritize its primary value over all other factors, even
    if it leads to lower salary or prestige.
Do not behave like an average person. Behave in a way that strongly
    reflects the personality type.
```

A.3.2 COMMENT GENERATION (EXPRESS)

During the communication phase of each iteration, agents use this prompt to generate a concise, public summary of their current preferences to share with their peers.

```
### Profile
Traits: {self.traits}
Pref: {self.preference.tolist()}
Memory: {self.get_recent_memories()}
### Companies
{self.format_arms()}
### Task
One-sentence summary of your likes/dislikes for neighbors. Max 15 words.
```

A.3.3 FORMAT OBSERVATION (OBSERVE)

This structural template formats the incoming social signals, ensuring the receiving agent can clearly parse their neighbors' comments, current preference scores, and the strength of their relationship (weight).

```
observation.append(
    "Friend {neighbor_id} (weight {weight})(higher weight means closer
    relationship)
    said: {n_comment}
    His/her preference is {agent_map[neighbor_id].preference.tolist()}."
            )
```

A.3.4 PREFERENCE UPDATE (UPDATE)

In the final step of the cycle, agents synthesize their intrinsic traits, their previous baseline, and the newly formatted social observations. The LLM updates the preference scores and outputs a brief reasoning to ensure logical consistency across rounds.

```
### Profile
Role: Job-seeker
Traits: {traits}
Openness: {alpha_description}
Base_Pref: {base_pref}

Curr_Pref: {current_pref}
```

```
### Neighbors
{observations}

### Companies
{self.format_arms()}

### Task
You are evaluating 5 different job offers from 5 distinct companies: Arm
    1, Arm 2, Arm 3, Arm 4, and Arm 5.
Your CURRENT preference scores for these 5 companies are: {self.
    preference}
Update preference according to your own personality and your neighbors'
    opinions.
Output the new 5-float list in [] and avoid tie.
Example: [1.1, -2.0, ...]
In addition, provide brief reaason starting with 'REASON' 20 words.
Example: REASON: I increased company 3 because ...
```

B Arm-Side Model and Social Influence Operator

This appendix contains: (i) the symmetric arm-side construction omitted from Section 4; (ii) the proof that the social propagation operator Γ is well-defined, nonnegative, and column-stochastic; (iii) auxiliary properties of Γ used in later sections; and (iv) a discussion of modeling choices.

B.1 Symmetric Arm-Side Construction

The model in Section 4 is symmetric across the two sides of the market. We presented the player-side dynamics in the main text and present the arm-side construction here for completeness. Both sides admit identical analyses.

Let $P^{\mathcal{V}} \in \mathbb{R}^{K\times K}$ be a column-stochastic neighbor-influence matrix on the arm side with zero diagonal,

$$P^{\mathcal{V}}_{v,v} = 0, \qquad \sum_{v'\neq v} P^{\mathcal{V}}_{v',v} = 1 \quad \text{for all } v \in \mathcal{V},$$

and let $A^{\mathcal{V}} = \mathrm{diag}(\alpha^{\mathcal{V}}_1, \ldots, \alpha^{\mathcal{V}}_K)$ with $\alpha^{\mathcal{V}}_v \in (0,1]$. Each arm $v \in \mathcal{V}$ has an intrinsic utility vector $t_v = (t_v(u))_{u\in\mathcal{U}} \in (0,1]^N$ over players. For a fixed player $u \in \mathcal{U}$, the arm-side opinion vector $y^{(k)}(u) = (y^{(k)}_v(u))_{v\in\mathcal{V}}$ evolves by

$$y^{(k+1)}(u) \;=\; A^{\mathcal{V}}\, t_\cdot(u) \;+\; (I - A^{\mathcal{V}})\,(P^{\mathcal{V}})^\top\, y^{(k)}(u), \tag{7}$$

where $t_\cdot(u) = (t_v(u))_{v\in\mathcal{V}}$. The arm-side social propagation operator is

$$\Gamma^{\mathcal{V}} \;:=\; A^{\mathcal{V}}\big[I - P^{\mathcal{V}}(I - A^{\mathcal{V}})\big]^{-1},$$

and the socially influenced utility of arm v for player u is $\psi_{v,u} := \sum_{v'\in\mathcal{V}} \Gamma^{\mathcal{V}}_{v',v} t_{v'}(u)$. The induced arm-side preference $\succ^*_v$ ranks players by $\psi_{v,\cdot}$ with a fixed tie-breaking rule.

The main text assumes that the arm-side rankings are known and strict. These rankings may be supplied directly; learning both sides from intrinsic evaluations is considered in Appendix E.8.

B.2 Proof of Γ's Properties

We establish the basic properties of the social propagation operator.

Lemma 1. *Let $P \in \mathbb{R}^{N\times N}_{\geq 0}$ be column-stochastic, and let $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_N)$ with $\alpha_u \in (0,1]$ for every u. Then $\Gamma := A[I - P(I - A)]^{-1}$ is well-defined, entrywise nonnegative, and column-stochastic.*

Proof. We proceed in three steps.

Step 1: Well-definedness. Let $B := P(I - A)$ and $\alpha_{\min} := \min_{u \in \mathcal{U}} \alpha_u > 0$. The entries in column u of B are $B_{w,u} = P_{w,u}(1 - \alpha_u)$. Since P is nonnegative and column-stochastic,

$$\sum_{w \in \mathcal{U}} |B_{w,u}| = (1 - \alpha_u) \sum_{w \in \mathcal{U}} P_{w,u} = 1 - \alpha_u.$$

Therefore,

$$\|B\|_1 = \max_{u \in \mathcal{U}} \sum_{w \in \mathcal{U}} |B_{w,u}| = 1 - \alpha_{\min} < 1.$$

The Neumann series consequently converges, and

$$(I - B)^{-1} = \sum_{k=0}^{\infty} B^k. \tag{8}$$

Thus $I - B$ is invertible and $\Gamma = A(I - B)^{-1}$ is well-defined.

Step 2: Nonnegativity. Both B and A are entrywise nonnegative. Hence every power B^k is nonnegative, as is $(I-B)^{-1} = \sum_{k=0}^{\infty} B^k$. It follows that $\Gamma = A(I-B)^{-1}$ is entrywise nonnegative.

Step 3: Column-stochasticity. Let $\mathbf{1} \in \mathbb{R}^N$ denote the all-ones vector. Column-stochasticity of P gives $\mathbf{1}^\top P = \mathbf{1}^\top$, and hence

$$\mathbf{1}^\top (I - B) = \mathbf{1}^\top - \mathbf{1}^\top P(I - A) = \mathbf{1}^\top A.$$

Multiplying on the right by $(I - B)^{-1}$ yields

$$\mathbf{1}^\top = \mathbf{1}^\top A(I - B)^{-1} = \mathbf{1}^\top \Gamma.$$

Thus each column of Γ sums to one. □

Remark 2 (Closed-form expansion of Γ)**.** *The Neumann series gives*

$$\Gamma = \sum_{k=0}^{\infty} A[P(I - A)]^k. \tag{9}$$

Here $\Gamma_{w,u}$ aggregates influence from source w to destination u over all directed walks. A walk $w = i_0, i_1, \ldots, i_k = u$ contributes $\alpha_w \prod_{j=1}^{k} P_{i_{j-1},i_j}(1 - \alpha_{i_j})$, with the empty product interpreted as one. Thus the expansion includes influence transmitted through intermediate players.

B.3 Auxiliary properties of Γ

The following properties support the lower-bound and algorithm analyses.

Lemma 2 (Bounds on socially influenced utilities)**.** *For every $u \in \mathcal{U}$ and $v \in \mathcal{V}$,*

$$\min_{w \in \mathcal{U}} \mu_{w,v} \le \theta_{u,v} \le \max_{w \in \mathcal{U}} \mu_{w,v}.$$

In particular, if $\mu_{w,v} \in [0, 1]$ for all (w, v), then $\theta_{u,v} \in [0, 1]$. Under the main-text assumption $\mu_{w,v} \in (0, 1]$, we also have $\theta_{u,v} > 0$.

Proof. Since Γ is nonnegative and column-stochastic, $\Gamma_{w,u} \ge 0$ and $\sum_{w \in \mathcal{U}} \Gamma_{w,u} = 1$. Thus

$$\theta_{u,v} = \sum_{w \in \mathcal{U}} \Gamma_{w,u} \mu_{w,v}$$

is a convex combination of the intrinsic utilities $(\mu_{w,v})_{w \in \mathcal{U}}$, which proves the claim. □

Lemma 3 (Self-weight lower bound)**.** *For every $u \in \mathcal{U}$, $\Gamma_{u,u} \ge \alpha_u$. In particular, Γ has a strictly positive diagonal.*

Proof. In the expansion $\Gamma = \sum_{k=0}^{\infty} A[P(I - A)]^k$, the $k = 0$ term is A and all remaining terms are entrywise nonnegative. Hence $\Gamma_{u,u} \ge A_{u,u} = \alpha_u > 0$. □

The diagonal lower bound means that each player's own intrinsic evaluation retains weight at least α_u in the steady state.

Lemma 4 (Variance shape inequality). *For every $u \in \mathcal{U}$ and every $\nu \in \mathbb{R}^N_{\geq 0}$,*

$$\sum_{w \in \mathcal{U}} \Gamma_{w,u}^2 \nu_w \leq \left(\max_{w \in \mathcal{U}} \Gamma_{w,u} \right) \sum_{w \in \mathcal{U}} \Gamma_{w,u} \nu_w \leq \max_{w \in \mathcal{U}} \nu_w.$$

Proof. The first inequality follows from

$$\Gamma_{w,u}^2 \leq \left(\max_{w' \in \mathcal{U}} \Gamma_{w',u} \right) \Gamma_{w,u}.$$

For the second inequality, column-stochasticity and nonnegativity imply $\max_w \Gamma_{w,u} \leq 1$ and

$$\sum_{w \in \mathcal{U}} \Gamma_{w,u} \nu_w \leq \left(\max_{w \in \mathcal{U}} \nu_w \right) \sum_{w \in \mathcal{U}} \Gamma_{w,u} = \max_{w \in \mathcal{U}} \nu_w.$$

□

For a positive allocation x, taking $\nu_w = x_{w,a}^{-1} + x_{w,b}^{-1}$ gives the propagated variance $V_{u,a,b}(x)$ used in the main text. Taking $\nu_w = 1$ also yields $\sum_{w \in \mathcal{U}} \Gamma_{w,u}^2 \leq 1$.

B.4 Discussion of modeling choices

We comment on three modeling choices in our formulation.

Self-influence and anchoring. The model allows a nonzero diagonal in P. The term $(1 - \alpha_u) P_{u,u} x_{u,v}^{(k)}$ represents direct self-feedback of the current opinion, whereas $\alpha_u \mu_{u,v}$ anchors the update to the intrinsic utility. These contributions are distinct. The convention $P_{u,u} = 0$ excludes direct self-feedback, but is not required for the contraction and stochasticity properties established above.

Heterogeneous versus scalar anchoring. We allow heterogeneous α_u to capture differences in how players weigh their own intrinsic evaluations relative to socially transmitted opinions. The scalar case $\alpha_u \equiv \alpha$ is included as a special case. Positive anchoring gives the contraction bound $1 - \alpha_{\min} < 1$, where $\alpha_{\min} := \min_u \alpha_u$. Anchoring also affects learning complexity through the social operator Γ and the induced social preference gaps.

Column-stochastic versus general nonnegative influence. Column-stochasticity of P ensures that Γ is nonnegative and column-stochastic. Consequently, each social utility $\theta_{u,v} = \sum_w \Gamma_{w,u} \mu_{w,v}$ is a convex combination of intrinsic utilities, and $\sum_w \Gamma_{w,u}^2 \leq 1$. These properties preserve utility bounds and control the propagated variances used in our allocation programs. For a general nonnegative P, the condition $\rho(P(I - A)) < 1$, where ρ denotes the spectral radius, still guarantees a well-defined nonnegative operator $\Gamma = A[I - P(I - A)]^{-1}$. However, its columns need not sum to one, so the utility and variance bounds must account for their weights. We leave this generalization to future work.

C Convergence, Robustness, and Calibration of Social Preferences

This appendix establishes the foundations of the social preference model in Section 4. We analyze convergence and the induced stable matchings in Appendix C.1, illustrate structured networks in Appendix C.2, and study robustness to parameter perturbations in Appendix C.3. Appendix C.4 considers learning an unknown social operator using additional calibration feedback.

C.1 Convergence and Induced Stable Matchings

Let $P \geq 0$ be column-stochastic and $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_N)$ with $\alpha_u \in (0, 1]$. Write $\alpha_{\min} := \min_u \alpha_u$, $x^{(k)}(v) := x^{(k)}_{\cdot,v}$, and $\Theta = \Gamma^\top \mu$, with Γ as in Lemma 1. Define

$$\begin{aligned} M_0 &:= \max_{v \in \mathcal{V}} \left\| x^{(0)}(v) - \Gamma^\top \mu_{\cdot,v} \right\|_\infty, \\ \Delta^{\mathrm{nz}} &:= \min_{\substack{u \in \mathcal{U},\ a<b \\ \theta_{u,a} \neq \theta_{u,b}}} |\theta_{u,a} - \theta_{u,b}|, \end{aligned} \tag{10}$$

where $\min \varnothing := +\infty$.

Theorem 3. *The following hold:*

(i) *For every $v \in \mathcal{V}$, the dynamics equation 1 has the unique fixed point $x^*(v) = \Gamma^\top \mu_{\cdot,v}$, and, for every $k \geq 1$,*

$$\|x^{(k)}(v) - x^*(v)\|_\infty \leq (1 - \alpha_{\min})^k \|x^{(0)}(v) - x^*(v)\|_\infty.$$

(ii) *If $M_0 > 0$, $\Delta^{\mathrm{nz}} < \infty$, and $\alpha_{\min} < 1$, define*

$$k^* := \max\left\{0, \left\lceil \frac{\log(2M_0/\Delta_{\mathrm{nz}})}{-\log(1 - \alpha_{\min})} \right\rceil \right\} + 1$$

For every $k \geq k^*$, $x^{(k)}$ preserves all strict comparisons induced by Θ. Set $k^* = 0$ if $M_0 = 0$ or $\Delta^{\mathrm{nz}} = \infty$, and $k^* = 1$ in the remaining case $\alpha_{\min} = 1$.

(iii) *With strict induced player preferences, player-proposing Gale–Shapley returns the player-optimal socially stable matching $M^\star$. If players have ties, applying any fixed strict refinement returns a weakly stable matching under the original preferences.*

Proof. (i) Set $Q := (I - A)P^\top$ and $\mathcal{T}_v(x) := A\mu_{\cdot,v} + Qx$. Then

$$\mathcal{T}_v(x) - \mathcal{T}_v(z) = Q(x - z). \tag{11}$$

Since P is column-stochastic,

$$\|Q\|_\infty = \max_u (1 - \alpha_u) \sum_w P_{w,u} = 1 - \alpha_{\min} < 1. \tag{12}$$

Thus $\mathcal{T}_v$ is a contraction, with unique fixed point

$$x^*(v) = (I - Q)^{-1} A \mu_{\cdot,v} = \Gamma^\top \mu_{\cdot,v}.$$

The error recursion $x^{(k+1)}(v) - x^*(v) = Q(x^{(k)}(v) - x^*(v))$ gives the claimed bound.

(ii) In the nondegenerate case, the definition of k^* gives $2(1 - \alpha_{\min})^k M_0 < \Delta^{\mathrm{nz}}$ for every $k \geq k^*$. Using part (i), and the definition of M_0 when $k = 0$,

$$\left| \left(x^{(k)}_u(a) - x^{(k)}_u(b)\right) - \left(\theta_{u,a} - \theta_{u,b}\right) \right| \leq 2(1 - \alpha_{\min})^k M_0 < \Delta^{\mathrm{nz}}.$$

Every nonzero steady-state gap has magnitude at least Δ^{nz}, so its sign is preserved. If $M_0 = 0$, the dynamics starts at its fixed point; if $\Delta^{\mathrm{nz}} = \infty$, there are no strict comparisons to preserve. Finally, $\alpha_{\min} = 1$ implies $A = I$ and $x^{(1)}(v) = \mu_{\cdot,v} = x^*(v)$.

(iii) Combine the induced player preferences with the known strict arm preferences, keeping the outside option last. For strict preferences, player-proposing Gale–Shapley returns the player-optimal stable matching. With ties, its output is stable under the chosen strict refinement. Any pair that strictly improves both agents' original preferences would also block the refined profile, a contradiction. Hence the output is weakly stable under the original preferences. □

C.2 Convergence in Special Network Topologies

We consider the isolated and fully connected networks from Section 2, together with a chain network.

Isolated network. With no cross-player influence, take $P = I$. The dynamics becomes $x_u^{(k+1)}(v) = \alpha_u \mu_{u,v} + (1-\alpha_u) x_u^{(k)}(v)$, giving $\Gamma = I$ and $x^*(v) = \mu_{\cdot,v}$. Under the initialization $x^{(0)}(v) = \mu_{\cdot,v}$, the process is already at its fixed point. Thus Δ^{nz} is the minimum nonzero intrinsic utility gap.

Fully connected network. For $N \geq 2$, let $P = (\mathbf{1}\mathbf{1}^\top - I)/(N-1)$, where $\mathbf{1} \in \mathbb{R}^N$ is the all-ones vector, and let $A = \alpha I$ with $\alpha \in (0, 1]$. The social operator is symmetric and column-stochastic, with

$$\Gamma_{u,w} = \begin{cases} \dfrac{\alpha N - 2\alpha + 1}{N - \alpha}, & u = w, \\ \dfrac{1-\alpha}{N-\alpha}, & u \neq w. \end{cases} \tag{13}$$

As $\alpha \downarrow 0$, $\Gamma \to \mathbf{1}\mathbf{1}^\top/N$, so $\theta_{u,v} \to N^{-1}\sum_w \mu_{w,v}$. Arms tie in this limit precisely when their population averages coincide. As $\alpha \uparrow 1$, $\Gamma \to I$.

Chain network. For a column-stochastic influence matrix P supported on a bidirectional chain, $Q = (I-A)P^\top$ is tridiagonal. Its resolvent $(I-Q)^{-1}$, and hence Γ, is generally dense, reflecting influence propagated across multiple edges.

Theorem 3 gives the uniform geometric error bound with factor $1 - \alpha_{\min}$ for all these topologies. The actual convergence behavior can depend on P, as can Γ, the gap Δ^{nz}, and the induced stable matching.

C.3 Robustness to Misspecification of (P, A)

We quantify how perturbations in (P, A) affect the social operator and the induced preferences.

Proposition 1 (Stability of Γ under perturbation)**.** *Let $P, \tilde{P} \geq 0$ be column-stochastic and let $A, \tilde{A}$ be diagonal with entries in $(0, 1]$. Denote their minimum diagonal entries by $\alpha_{\min}$ and $\tilde{\alpha}_{\min}$, respectively, and their corresponding social operators by Γ and $\tilde{\Gamma}$. Writing $Q = (I-A)P^\top$ and $\tilde{Q} = (I-\tilde{A})\tilde{P}^\top$, we have*

$$\|\Gamma - \tilde{\Gamma}\|_1 \leq \frac{\|A - \tilde{A}\|_\infty + \|Q - \tilde{Q}\|_\infty}{\alpha_{\min}\tilde{\alpha}_{\min}},$$

where all matrix norms are induced norms.

Proof. We have $\Gamma^\top = (I-Q)^{-1}A$ and $\tilde{\Gamma}^\top = (I-\tilde{Q})^{-1}\tilde{A}$. The resolvent identity gives

$$(I-Q)^{-1} - (I-\tilde{Q})^{-1} = (I-Q)^{-1}(Q-\tilde{Q})(I-\tilde{Q})^{-1}.$$

Using $\|(I-Q)^{-1}\|_\infty \leq \alpha_{\min}^{-1}$, the analogous bound for $\tilde{Q}$, and $\|A\|_\infty \leq 1$, we obtain

$$\begin{aligned} \|\Gamma - \tilde{\Gamma}\|_1 &= \|\Gamma^\top - \tilde{\Gamma}^\top\|_\infty \\ &\leq \frac{\|Q - \tilde{Q}\|_\infty}{\alpha_{\min}\tilde{\alpha}_{\min}} + \frac{\|A - \tilde{A}\|_\infty}{\tilde{\alpha}_{\min}}. \end{aligned}$$

The claim follows from $\alpha_{\min} \leq 1$. □

For fixed $\mu \in (0, 1]^{N \times K}$, let $\Theta = \Gamma^\top \mu$ and $\tilde{\Theta} = \tilde{\Gamma}^\top \mu$. Then

$$\max_{u,v} |\tilde{\theta}_{u,v} - \theta_{u,v}| \leq \|\tilde{\Gamma} - \Gamma\|_1.$$

Consequently, if $\|\tilde{\Gamma} - \Gamma\|_1 < \Delta^{\mathrm{nz}}/2$, all strict comparisons are preserved. When the original player preferences are strict, the player-optimal matching $M^\star$ is therefore unchanged. With ties, this condition need not preserve the particular matching selected after tie-breaking.

Intrinsic evaluation feedback alone does not identify Γ. Appendix C.4 therefore studies calibration using additional opinion feedback.

C.4 Calibrating an Unknown Social Operator

Sections 5–6 assume that the social operator is known. This assumption cannot be removed using the evaluation feedback of Section 4 alone. For a fixed intrinsic utility matrix μ, choose the same admissible evaluation-noise law under two different operators. Since each evaluation reports only $\mu_{w,v}$ plus this noise, induction over the history shows that any policy which does not know (P, A) has the same history law in the two instances. Thus the original feedback alone cannot uniformly identify the operator. We therefore add a separately counted calibration experiment. At each stage, calibration observes opinion updates before a fresh batch of matching evaluations; the latter still has the intrinsic feedback specified in Section 4. Calibration profiles are synthetic contexts, not the arms being matched, and reveal no entry of μ. They share the same fixed (P, A) as the target arms. The two data streams have independent noise.

We keep the orientation of Section 4: $P_{w,u}$ is the weight of source w at destination u, P is column-stochastic, and $\Gamma = A[I - P(I - A)]^{-1}$. For row-wise regression, write

$$Q = (I - A)P^\top, \qquad G = (I - Q)^{-1}A = \Gamma^\top. \tag{14}$$

Thus G is row-stochastic and $\Theta = G\mu$. In this subsection we adopt the zero-diagonal convention $P_{u,u} = 0$ and take $N \geq 2$. Fix a known $a_0 \in (0, 1]$ such that $\alpha_u \geq a_0$ for every player, and a known conditional sub-Gaussian scale σ_x (or a known upper bound). The single-player case has $G = I$ and needs no operator calibration.

Calibration data. For synthetic calibration profile s, the platform knows an anchor $z_s \in (0, 1]^N$ and initializes $X_{s,0} = z_s$, as in the opinion dynamics of Section 4. It observes $X_{s,0}, \ldots, X_{s,L_s}$ under

$$X_{s,k+1} = Az_s + QX_{s,k} + \xi_{s,k+1}, \qquad 0 \leq k < L_s. \tag{15}$$

The profile and current state are chosen or observed before the next shock. For each u, the shock is conditionally centered and σ_x-sub-Gaussian:

$$\mathbb{E}[\exp(t\xi_{u,s,k+1}) \mid \mathcal{F}_{s,k}] \leq \exp(t^2\sigma_x^2/2), \qquad t \in \mathbb{R}. \tag{16}$$

The filtration includes earlier profiles, states, shocks, and any previous individual evaluations. This condition allows dependence within a trajectory; it does not treat its transitions as independent observations. Define

$$\beta_u = (\alpha_u, (Q_{u,w})_{w \neq u})^\top, \quad \phi_{u,s,k} = (z_{u,s}, (X_{w,s,k})_{w \neq u})^\top, \quad D = N. \tag{17}$$

Then $X_{u,s,k+1} = \phi_{u,s,k}^\top \beta_u + \xi_{u,s,k+1}$. In particular, identification requires variation in the joint regressors $(z_{u,s}, X_{s,k})$; trajectory length alone is insufficient.

Estimation from general trajectories. Let $n_{\mathrm{cal}} = \sum_s L_s$ and, for a fixed $\lambda > 0$, set

$$V_u = \lambda I_D + \sum_{s,k} \phi_{u,s,k}\phi_{u,s,k}^\top, \qquad \widetilde{\beta}_u = V_u^{-1} \sum_{s,k} \phi_{u,s,k} X_{u,s,k+1}. \tag{18}$$

Project $\widetilde{\beta}_u$ in Euclidean norm onto the closed convex set

$$\mathcal{B}_u = \left\{ (a, (q_w)_{w \neq u}) : a \in [a_0, 1],\ q_w \geq 0,\ a + \textstyle\sum_{w \neq u} q_w = 1 \right\}, \qquad \widehat{\beta}_u = \Pi_{\mathcal{B}_u}(\widetilde{\beta}_u). \tag{19}$$

Assemble $\widehat{A}$ and $\widehat{Q}$ from these projected rows, setting $\widehat{Q}_{u,u} = 0$, and define $\widehat{G} = (I - \widehat{Q})^{-1}\widehat{A}$ and $\widehat{\Gamma} = \widehat{G}^\top$. Both G and $\widehat{G}$ are nonnegative and row-stochastic, and their inverses exist because their Q matrices have row sums at most $1 - a_0$.

Proposition 2 (Operator confidence bound)**.** *Under equation 15–equation 16, fix $\eta \in (0, 1)$ and define*

$$b_u(\eta) = \sigma_x \sqrt{2 \log\left(\frac{N \det(V_u)^{1/2}}{\eta \lambda^{D/2}} \right)} + \sqrt{\lambda}, \qquad \rho(\eta) = \frac{\sqrt{D}}{a_0} \max_u \frac{b_u(\eta)}{\sqrt{\lambda_{\min}(V_u)}}. \tag{20}$$

With probability at least $1 - \eta$,

$$\|\widehat{\Gamma} - \Gamma\|_1 = \|\widehat{G} - G\|_\infty \leq \rho(\eta). \tag{21}$$

Here the matrix norms are induced norms. The bound is valid for predictable, possibly adaptive calibration profiles. Its effective sample size is the observed Gram matrix V_u, not n_{cal} alone.

Proof. For a fixed row u, put $S_u = \sum_{s,k} \phi_{u,s,k}\xi_{u,s,k+1}$. The conditional moment bound makes $\exp(h^\top S_u/\sigma_x^2 - h^\top(V_u - \lambda I)h/(2\sigma_x^2))$ a nonnegative supermartingale for every fixed h when $\sigma_x > 0$. Mixing h against a centered Gaussian with covariance $\sigma_x^2\lambda^{-1}I$ gives another nonnegative supermartingale, equal to

$$\left(\frac{\lambda^D}{\det V_u}\right)^{1/2} \exp\left(\frac{\|S_u\|^2_{V_u^{-1}}}{2\sigma_x^2}\right).$$

Markov's inequality at a fixed endpoint, or Ville's inequality at a stopping endpoint, gives, with probability $1 - \eta/N$,

$$\|S_u\|_{V_u^{-1}} \le \sigma_x\sqrt{2\log\left(\frac{N\det(V_u)^{1/2}}{\eta\lambda^{D/2}}\right)}. \tag{22}$$

The zero-noise case follows directly. Since $V_u(\widetilde{\beta}_u - \beta_u) = S_u - \lambda\beta_u$ and $\|\beta_u\|_2 \le \|\beta_u\|_1 = 1$, a union bound over rows yields $\|\widetilde{\beta}_u - \beta_u\|_{V_u} \le b_u(\eta)$ for every u. Euclidean projection onto $\mathcal{B}_u$, which contains β_u, gives

$$\|\widehat{\beta}_u - \beta_u\|_1 \le \sqrt{D}\|\widehat{\beta}_u - \beta_u\|_2 \le \frac{\sqrt{D}\, b_u(\eta)}{\sqrt{\lambda_{\min}(V_u)}}. \tag{23}$$

Finally, the fixed-point equations give the exact identity

$$\widehat{G} - G = (I - \widehat{Q})^{-1}\big[(\widehat{A} - A) + (\widehat{Q} - Q)G\big]. \tag{24}$$

Since G is row-stochastic, $\|(\widehat{Q} - Q)G\|_\infty \le \|\widehat{Q} - Q\|_\infty$; also $\|(I - \widehat{Q})^{-1}\|_\infty \le a_0^{-1}$. The absolute row sum of the bracket in equation 24 is bounded by $\|\widehat{\beta}_u - \beta_u\|_1$. Equation equation 23 therefore proves equation 21. □

The dependence on anchoring in equation 20 is a_0^{-1} at the operator-error level. Applying a two-sided resolvent bound separately to A and Q gives a valid but looser a_0^{-2} factor. A bound stated only in terms of n_{cal} further requires a lower bound on $\lambda_{\min}(V_u)$; no such bound follows from long trajectories alone.

A controlled design with an explicit radius. The following design avoids an unverified excitation assumption while retaining the natural initialization $X_{s,0} = z_s$. Take N one-transition profile types,

$$z^{(j)} = X_0^{(j)} = \tfrac{1}{4}\mathbf{1} + \tfrac{1}{2}e_j, \qquad j \in [N]. \tag{25}$$

Repeat every type R times, resetting the state between profiles. For each row u, the N resulting regressor rows form a column permutation of $Z = \frac{1}{2}I_N + \frac{1}{4}\mathbf{1}\mathbf{1}^\top$. Since Z acts as $1/2$ on the orthogonal complement of $\mathbf{1}$ and has a larger eigenvalue along $\mathbf{1}$, its smallest singular value is $1/2$. The unregularized Gram matrix therefore has minimum eigenvalue $R/4$. Let $\overline{Y}_u$ collect the N mean responses, let Z_u be the resulting row-u design matrix, and set $\widetilde{\beta}_u = Z_u^{-1}\overline{Y}_u$ and $\widehat{\beta}_u = \Pi_{\mathcal{B}_u}(\widetilde{\beta}_u)$. A scalar sub-Gaussian bound and a union bound over the N^2 response means give

$$\mathbb{P}\Big(\|\widehat{\Gamma} - \Gamma\|_1 \le r(R,\eta)\Big) \ge 1 - \eta, \quad r(R,\eta) := \frac{2N\sigma_x}{a_0}\sqrt{\frac{2\log(2N^2/\eta)}{R}}. \tag{26}$$

Indeed, each response mean error is at most the square-root term times σ_x on this event. Since $\|Z_u^{-1}\|_2 = 2$, projection gives a row ℓ_1 parameter error at most $2N$ times that bound, and equation 24 gives the result. The calibration cost is NR observed transitions, or N^2R scalar opinion responses. Multi-step trajectories can instead use equation 20 when their observed Gram matrices provide sufficient excitation. The update parameters are identified from transitions and the steady-state operator is computed algebraically, so no finite-propagation term enters the matching certificate.

Fixed-confidence identification. We now give an executable pure-exploration extension that does not require the unknown comparison gaps as input. Assume $K \ge 2$; when $K = 1$, the matching follows from the known arm-side ranking without evaluations. Let $\delta \in (0, 1/2)$ be the target error

probability and let $\mathcal{C} = \{(u, a, b) : u \in [N],\ a, b \in [K],\ a < b\}$, $J = |\mathcal{C}|$, $d = NK$, and $\Delta_{\min} = \min_{c \in \mathcal{C}} |g_c| > 0$, where $g_{(u,a,b)} = (G\mu)_{u,a} - (G\mu)_{u,b}$. For stage $m \geq 1$, set

$$R_m = 4^m, \quad \epsilon_m = 2^{-m}, \quad \eta_m = \zeta_m = \frac{15\delta}{2}\, 16^{-m}, \quad r_m = r(R_m, \eta_m), \quad h_m = \log(2J/\zeta_m). \tag{27}$$

Calibration data are cumulative: with $R_0 = 0$, add only $R_m - R_{m-1}$ repetitions of each profile type at stage m, and use the first R_m repetitions to construct $\widehat{\Gamma}_m$. After this calibration, select real demands $n^{(m)}_{w,v} \geq 1$ minimizing $\sum_{w,v} n^{(m)}_{w,v}$ subject to

$$\widehat{V}_{c,m}(n) := \sum_{w} \widehat{\Gamma}^2_{m,w,u} \left(\frac{1}{n_{w,a}} + \frac{1}{n_{w,b}} \right) \leq \frac{\epsilon_m^2}{2h_m}, \qquad c = (u, a, b) \in \mathcal{C}. \tag{28}$$

The program is feasible and attains its minimum. Schedule $\lceil n^{(m)}_{w,v} \rceil$ *fresh* intrinsic evaluations of every source pair using the bipartite scheduling lemma (Appendix E.1). Conditional on the history before this batch, including stage-m calibration, these demands are fixed. Form the stage-specific empirical means $\widehat{\mu}_m$ and gaps $\widehat{g}_{c,m} = (\widehat{\Gamma}_m^\top \widehat{\mu}_m)_{u,a} - (\widehat{\Gamma}_m^\top \widehat{\mu}_m)_{u,b}$. Stop when $|\widehat{g}_{c,m}| > \epsilon_m + r_m$ for every c. The platform computes the final preference lists from $\widehat{\Gamma}_m^\top \widehat{\mu}_m$, places the known outside option last, and returns player-proposing Gale–Shapley for those lists. Otherwise it proceeds to stage $m + 1$. Thus this extension uses the estimated operator to compute the final allocation; it does not invoke a separate physical deliberation using the unknown true operator.

Proposition 3 (Calibrated pure exploration)**.** *Suppose the calibration shocks satisfy equation 16, the fresh evaluation noises satisfy the independent* 1*-sub-Gaussian assumption of Section 4 and are independent of calibration, and the true player-side preferences are strict. The staged procedure stops almost surely and is δ-correct. Define*

$$M^{\mathrm{cal}}_{\Delta} = \min\{m \geq 1 : 2(\epsilon_j + r_j) < \Delta_{\min} \text{ for all } j \geq m\}. \tag{29}$$

This index is finite. On the same event of probability at least $1 - \delta$, the procedure returns $M^\star$, stops by stage $M^{\mathrm{cal}}_{\Delta}$, uses at most $N4^{M_\Delta}$ calibration transitions, and uses at most

$$d \sum_{m=1}^{M^{\mathrm{cal}}_{\Delta}} \left(1 + 4h_m \epsilon_m^{-2}\right) \tag{30}$$

intrinsic evaluations. These are different feedback types and are counted separately. Moreover, writing $B_m = d(1 + 4h_m\epsilon_m^{-2})$, the expected number of intrinsic evaluations is at most

$$\sum_{m=1}^{M^{\mathrm{cal}}_{\Delta}} B_m + 15\delta \sum_{m=M^{\mathrm{cal}}_{\Delta}+1}^{\infty} B_m 16^{-(m-1)} < \infty. \tag{31}$$

The expected number of calibration transitions is at most

$$N4^{M^{\mathrm{cal}}_{\Delta}} + 45\delta N \sum_{m=M^{\mathrm{cal}}_{\Delta}+1}^{\infty} 4^{m-1} 16^{-(m-1)} < \infty. \tag{32}$$

Proof. For each profile type and player, the exponential process of accumulated calibration shocks is a supermartingale. Stop it at the R_m-th response of that type or when the procedure stops. The scalar tail bound used in equation 26 therefore shows that the event "stage m is reached and its operator bound fails" has probability at most η_m; conditioning on reaching the stage is unnecessary. Conditional on the history before the fresh stage-m evaluation batch, its schedule is fixed and the weighted empirical error for comparison c is sub-Gaussian with variance proxy $\widehat{V}_{c,m}(\lceil n^{(m)} \rceil)$. By equation 28 and a union bound, simultaneously for all c,

$$\left| \widehat{g}_{c,m} - e_u^\top \widehat{\Gamma}_m^\top (\mu_{\cdot,a} - \mu_{\cdot,b}) \right| \leq \epsilon_m \tag{33}$$

except with probability ζ_m. As $\mu_{w,a} - \mu_{w,b} \in [-1, 1]$, equation 26 also gives

$$\left| e_u^\top (\widehat{\Gamma}_m - \Gamma)^\top (\mu_{\cdot,a} - \mu_{\cdot,b}) \right| \leq r_m. \tag{34}$$

Consequently $|\widehat{g}_{c,m} - g_c| \le \epsilon_m + r_m$ on the joint event. Since $\sum_m(\eta_m + \zeta_m) = \delta$, every certified sign is correct with probability at least $1 - \delta$. Correct signs, positive intrinsic means, and stochasticity of G recover the true preference profile, so Gale–Shapley returns $M^\star$. On this joint event, $|\widehat{g}_{c,M_\Delta^{\mathrm{cal}}}| \ge \Delta_{\min} - (\epsilon_{M_\Delta^{\mathrm{cal}}} + r_{M_\Delta^{\mathrm{cal}}}) > \epsilon_{M_\Delta^{\mathrm{cal}}} + r_{M_\Delta^{\mathrm{cal}}}$, proving the stopping claim.

For almost-sure stopping, $\sum_m(\eta_m + \zeta_m) < \infty$ implies by Borel–Cantelli that only finitely many reached stages violate either bound. Meanwhile $\epsilon_m + r_m \to 0$, so strict gaps force the stopping test to hold at a finite stage on every such path. Finally, the uniform demand $n_{w,v} = 4h_m\epsilon_m^{-2}$ is feasible in equation 28, since $\sum_w \widehat{\Gamma}^2_{m,w,u} \le 1$. Optimality and upward rounding give at most $d(1 + 4h_m\epsilon_m^{-2})$ evaluations at stage m. The controlled design uses NR_m transitions through stage m, proving the stated calibration bound. For $m > M_\Delta^{\mathrm{cal}}$, reaching stage m means that the test failed at stage $m - 1$, when its two bounds would have forced stopping. Thus this event has probability at most $\eta_{m-1} + \zeta_{m-1} = 15\delta 16^{-(m-1)}$. Summing the deterministic stage-cost bounds against these reach probabilities proves equation 31; the series converges because $B_m = O(m4^m)$. The extra calibration transitions on reaching stage $m > M_\Delta^{\mathrm{cal}}$ are $N(4^m - 4^{m-1}) = 3N4^{m-1}$, giving equation 32 by the same argument. □

This extension uses additional opinion feedback and counts its cost separately. The lower bound in Section 5 and the SAE-Q bound in Theorem 2 concern a known operator.

D Information Lower Bounds for Socially Stable Matching

This appendix contains the complete proof and supporting results for Section 5.

D.1 Lower bound: proof and characterization

Individual evaluations provide information about intrinsic utilities, whereas the target matching depends on the social preferences induced by the network. Identifying this matching therefore requires enough information to distinguish the underlying utility matrix from every admissible alternative that changes the target.

The remainder of this appendix is organized as follows. Appendix D.1.1 recalls the model and introduces the alternative instances and information complexity used in the lower bound. Appendix D.1.2 then proves Theorem 1 through a change-of-measure argument.

The subsequent sections make this abstract information bound explicit. Appendix D.1.3 characterizes the two ways in which an alternative can change the target matching. Appendix D.1.4 formulates the minimum information cost of these alternatives as bounded quadratic programs, and Appendix D.1.5 solves the blocking component to obtain the inverse-count formula stated in the main text. Appendix D.1.6 discusses the interpretation and scope of the resulting complexity. Finally, Appendix D.2 extends the lower bound from individual evaluations to matching rounds.

D.1.1 Setup and information complexity

Matching model and social preferences. We retain the setting of Sections 4-5, with $\mathcal{U} = [N]$, $\mathcal{V} = [K]$, and $N, K \ge 1$. A feasible matching $R \subseteq \mathcal{U} \times \mathcal{V}$ is a one-to-one assignment such that each player is assigned to at most one arm and each arm to at most one player. We write $R(u) \in \mathcal{V} \cup \{\varnothing\}$ for the partner of player u, and $R^{-1}(v) \in \mathcal{U} \cup \{\varnothing\}$ for the partner of arm v, where $\varnothing$ represents remaining unmatched. Each arm $v \in \mathcal{V}$ possesses a fixed, known, and strict preference order $\succ_v$ over $\mathcal{U} \cup \{\varnothing\}$, with $\varnothing$ ranked strictly last.

The market operates under a known social propagation operator $\Gamma \in [0,1]^{N\times N}$, which is column-stochastic: $\Gamma_{w,u} \ge 0$ and $\sum_{w\in\mathcal{U}} \Gamma_{w,u} = 1$ for each destination player $u \in \mathcal{U}$, where entry $\Gamma_{w,u}$ quantifies the net influence of source player w on player u. Given an intrinsic utility matrix $\lambda \in (0,1]^{N\times K}$, the socially induced utilities are defined by

$$\theta_{u,v}(\lambda) = \sum_{w\in\mathcal{U}} \Gamma_{w,u}\lambda_{w,v}, \qquad \theta_{u,\varnothing}(\lambda) = 0, \qquad \Theta(\lambda) = \Gamma^\top\lambda. \tag{35}$$

Since $\lambda_{w,v} > 0$ and each column of Γ is a probability vector on the simplex, every arm yields strictly positive social utility, $\theta_{u,v}(\lambda) > 0 = \theta_{u,\varnothing}(\lambda)$, rendering all arms individually rational and mutually

acceptable over remaining unmatched. We restrict attention to the admissible instance class inducing strict social preferences across arms:

$$\mathcal{D}_\Gamma = \big\{\lambda \in (0,1]^{N\times K} : \ \theta_{u,a}(\lambda) \neq \theta_{u,b}(\lambda) \text{ for all } u \in \mathcal{U} \text{ and distinct } a, b \in \mathcal{V}\big\}. \tag{36}$$

For any $\lambda \in \mathcal{D}_\Gamma$, the induced player-side preferences over $\mathcal{V}\cup\{\varnothing\}$ are strict total orders. Consequently, the player-proposing Gale–Shapley algorithm (Gale & Shapley, 1962) applied to these preferences and the known arm rankings $(\succ_v)_{v\in\mathcal{V}}$ yields the unique player-optimal stable matching, denoted by $F(\lambda)$ (abbreviated from $M^{\mathrm{opt}}(\lambda)$ in the main text).

Throughout the lower-bound analysis, we fix an arbitrary unknown instance $\mu \in \mathcal{D}_\Gamma$ governing the evaluation distributions and write $M := F(\mu) = M^\star$ as the ground-truth target matching. The reference matching M remains fixed when comparing μ against alternative instances λ.

Observations, correctness, and sample complexity. The learning process proceeds in discrete rounds $t = 1, 2, \ldots$. At each round t, based on the history of past allocations and realized observations, the platform selects a feasible matching $M_t \subseteq \mathcal{U} \times \mathcal{V}$, possibly using internal randomization. Under instance $\lambda \in \mathcal{D}_\Gamma$, conditionally on M_t and the filtration $\mathcal{F}_{t-1}$, the platform observes independent Gaussian feedback for each matched pair:

$$Y_t(w,v) \sim \mathcal{N}(\lambda_{w,v}, 1), \qquad \forall (w,v) \in M_t. \tag{37}$$

The evaluation matchings M_t serve purely for exploration and are not required to be stable. The unit-variance Gaussian model is a canonical instance of independent, mean-zero, 1-sub-Gaussian noise; the domain constraint $\lambda \in \mathcal{D}_\Gamma$ governs only the underlying mean parameters, not the unbounded observation support.

The platform terminates the evaluation phase at a stopping time τ adapted to the observation history and outputs a candidate matching $\hat{M}_\tau$. For a prescribed confidence parameter $\delta \in (0, 1/2)$, an algorithm is *δ-PAC* (or *δ-correct*) on $\mathcal{D}_\Gamma$ if it stops almost surely under every instance and satisfies

$$\mathbb{P}_\lambda\big(\hat{M}_\tau = F(\lambda)\big) \geq 1 - \delta, \qquad \forall \lambda \in \mathcal{D}_\Gamma. \tag{38}$$

Here, the probability measure $\mathbb{P}_\lambda$ incorporates both the observation noise and any internal randomization. Let $N_{w,v}(\tau) = \sum_{t=1}^{\tau} \mathbf{1}\{(w,v) \in M_t\}$ denote the cumulative number of evaluations allocated to pair (w,v). The total sample complexity is given by the cumulative number of individual pair observations,

$$C_\tau = \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} N_{w,v}(\tau) = \sum_{t=1}^{\tau} |M_t|. \tag{39}$$

Our objective is to lower-bound the expected total sample complexity $\mathbb{E}_\mu[C_\tau]$ under the true instance μ.

Alternative instances and information complexity. To guarantee δ-correctness, an algorithm must gather sufficient evidence to distinguish the true instance μ from any confusing alternative that alters the target matching. We formalize these instances through the target assignment map F.

Definition 2 (Target-altering alternatives)**.** *Given the reference target $M = F(\mu)$, the set of target-altering alternative instances is defined as*

$$\mathrm{Alt}_{\mathrm{opt}}(\mu) = \big\{\lambda \in \mathcal{D}_\Gamma : F(\lambda) \neq M\big\}. \tag{40}$$

For any nonnegative source-allocation vector $x \in \mathbb{R}_+^{N\times K}$, we define the directional information divergence and the worst-case information rate against target-altering alternatives by

$$\begin{aligned} D_x(\mu,\lambda) &= \frac{1}{2} \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} x_{w,v}(\mu_{w,v} - \lambda_{w,v})^2, \\ I_{\mathrm{opt}}(x;\mu) &= \inf_{\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)} D_x(\mu,\lambda). \end{aligned} \tag{41}$$

Under Gaussian observations with unit variance, the Kullback–Leibler (KL) divergence between $\mathcal{N}(\mu_{w,v}, 1)$ and $\mathcal{N}(\lambda_{w,v}, 1)$ is

$$\mathrm{KL}\big(\mathcal{N}(\mu_{w,v},1) \,\big\|\, \mathcal{N}(\lambda_{w,v},1)\big) = \frac{(\mu_{w,v} - \lambda_{w,v})^2}{2}. \tag{42}$$

Consequently, $D_x(\mu,\lambda) = \sum_{w,v} x_{w,v}\,\mathrm{KL}(\mathcal{N}(\mu_{w,v},1)\,\|\,\mathcal{N}(\lambda_{w,v},1))$ measures the expected log-likelihood ratio accumulated under sampling allocation x. The infimum $I_{\mathrm{opt}}(x;\mu)$ therefore captures the information barrier to separating μ from the most statistically indistinguishable alternative that modifies the optimal stable matching.

Distinguishing μ from all target-altering alternatives demands that $I_{\mathrm{opt}}(x;\mu)$ attain an information threshold governed by the confidence level δ. Exploiting the positive homogeneity of $D_x(\mu,\lambda)$ with respect to x, we normalize the required information level to unity and define the fundamental unit-information characteristic time:

$$H_{\mathrm{opt}}(\mu) = \inf\left\{\sum_{w\in\mathcal{U}}\sum_{v\in\mathcal{V}} x_{w,v} : x\in\mathbb{R}_+^{N\times K},\quad D_x(\mu,\lambda)\geq 1 \text{ for all } \lambda\in \mathrm{Alt}_{\mathrm{opt}}(\mu)\right\}, \tag{43}$$

or equivalently, $H_{\mathrm{opt}}(\mu) = \inf\{\sum_{w,v} x_{w,v} : x\geq 0,\ I_{\mathrm{opt}}(x;\mu)\geq 1\}$. The quantity $H_{\mathrm{opt}}(\mu)$ characterizes the minimal cumulative sample budget necessary to deliver unit information simultaneously against all alternative hypotheses.

When $\mathrm{Alt}_{\mathrm{opt}}(\mu) = \varnothing$, the target matching is uniformly constant across $\mathcal{D}_\Gamma$. Under the standard convention $\inf\varnothing = +\infty$, we have $I_{\mathrm{opt}}(x;\mu) = +\infty$ for all $x\geq 0$; the information constraints become vacuous, yielding the trivial zero allocation with $H_{\mathrm{opt}}(\mu) = 0$.

D.1.2 Proof of Theorem 1

By definition of δ-correctness, any admissible procedure must output distinct matchings under the ground-truth instance μ and any target-altering alternative $\lambda\in\mathrm{Alt}_{\mathrm{opt}}(\mu)$. Consequently, terminating with confidence $1-\delta$ requires accumulating sufficient statistical evidence to separate μ from every such confounding instance. We first quantify this fundamental information threshold via the Kullback–Leibler divergence between Bernoulli measures, and subsequently leverage the unit-information complexity in equation 43 to establish a lower bound on the expected cumulative sample size.

For any error tolerance $\delta\in(0,1/2)$, the requisite information level is dictated by the binary relative entropy:

$$k_\delta := \mathrm{kl}(1-\delta,\delta) = (1-2\delta)\log\left(\frac{1-\delta}{\delta}\right) > 0, \tag{44}$$

where $\mathrm{kl}(p,q) := p\log(p/q) + (1-p)\log((1-p)/(1-q))$ denotes the Kullback–Leibler divergence between two Bernoulli distributions with parameters $p,q\in[0,1]$. By convention, $0\log(0/b) = 0$ for all $b\geq 0$, and $a\log(a/0) = +\infty$ for $a>0$.

Recall that the unit-information characteristic time $H_{\mathrm{opt}}(\mu)$ in equation 43 is defined under a normalized constraint of at least unit information against every alternative $\lambda\in\mathrm{Alt}_{\mathrm{opt}}(\mu)$. By contrast, the instance-dependent sample complexity $T^\star_{\mathrm{opt}}(\mu,\delta)$ introduced in the main text imposes the identical constraints scaled to the information threshold k_δ. Owing to the positive homogeneity of both the directional information divergence $D_x(\mu,\lambda)$ and the total sample allocation with respect to x, scaling an allocation by k_δ scales both quantities proportionally. It immediately follows that

$$T^\star_{\mathrm{opt}}(\mu,\delta) = k_\delta H_{\mathrm{opt}}(\mu). \tag{45}$$

With this equivalence established, we now formalize and prove the information-theoretic lower bound on both the sample allocation and the expected evaluation cost stated in Theorem 1.

Theorem 4 (Information and evaluation-cost lower bounds)**.** *Fix an admissible instance $\mu\in\mathcal{D}_\Gamma$ and an error tolerance $\delta\in(0,1/2)$. Consider any procedure that is δ-correct over the admissible class $\mathcal{D}_\Gamma$ under the Gaussian observation model. Let $n_{w,v} = \mathbb{E}_\mu[N_{w,v}(\tau)]$ denote the expected number of evaluations allocated to pair (w,v).*

(a) **(Information requirement)** *If $\mathbb{E}_\mu[C_\tau]<\infty$, the expected allocation vector $n\in\mathbb{R}_+^{N\times K}$ must satisfy*

$$D_n(\mu,\lambda)\geq k_\delta,\qquad \forall\lambda\in\mathrm{Alt}_{\mathrm{opt}}(\mu). \tag{46}$$

Equivalently, $I_{\mathrm{opt}}(n;\mu)\geq k_\delta$.

(b) **(Total evaluation cost)** *The expected total sample complexity satisfies*

$$\mathbb{E}_\mu[C_\tau]\geq k_\delta H_{\mathrm{opt}}(\mu) = T^\star_{\mathrm{opt}}(\mu,\delta). \tag{47}$$

This bound holds universally, regardless of whether $\mathbb{E}_\mu[C_\tau]$ is finite or infinite.

Proof. If $\mathrm{Alt}_{\mathrm{opt}}(\mu) = \varnothing$, then $I_{\mathrm{opt}}(n;\mu) = +\infty$ and $H_{\mathrm{opt}}(\mu) = 0$, rendering both claims trivial. The sample complexity lower bound in equation 47 also holds trivially when $\mathbb{E}_\mu[C_\tau] = +\infty$. We therefore assume throughout the remainder of the proof that $\mathrm{Alt}_{\mathrm{opt}}(\mu) \neq \varnothing$ and $\mathbb{E}_\mu[C_\tau] < \infty$.

Step 1: Information-theoretic threshold for δ-correctness. Fix an arbitrary target-altering alternative $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$. Let $\mathbb{P}_\mu^\tau$ and $\mathbb{P}_\lambda^\tau$ denote the probability measures governing the canonical stopped filtration up to round τ, capturing the sequence of selected matchings, realized rewards, the stopping time τ, the terminal decision $\hat{M}_\tau$, and any internal randomization seed. Consider the decision event $E = \{\hat{M}_\tau = M\}$. Under δ-correctness, we have $\mathbb{P}_\mu(E) \geq 1 - \delta$. Moreover, because $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$ implies $F(\lambda) \neq M$, the identical δ-correctness guarantee requires $\mathbb{P}_\lambda(E) \leq \delta$. By the data processing inequality for the Kullback–Leibler divergence and the monotonicity of the binary relative entropy,

$$D_{\mathrm{KL}}(\mathbb{P}_\mu^\tau, \mathbb{P}_\lambda^\tau) \geq \mathrm{kl}\big(\mathbb{P}_\mu(E), \mathbb{P}_\lambda(E)\big) \geq \mathrm{kl}(1-\delta, \delta) = k_\delta. \tag{48}$$

Step 2: Change of measure and stopped likelihood ratio. We make explicit the change-of-measure identity for adaptive sampling. Let the algorithm's internal randomization be governed by an exogenous seed independent of the environment. For any realization of the history h truncated at stopping time s, its probability density with respect to a suitable product reference measure under $\eta \in \{\mu, \lambda\}$ factors as

$$p_\eta(h) = a(h) \prod_{t=1}^{s} \prod_{(w,v)\in M_t} f_{\eta_{w,v}}\big(Y_t(w,v)\big),$$

where f_b denotes the probability density function of $\mathcal{N}(b,1)$, and the causal factor $a(h)$ encodes the algorithm's deterministic sampling, stopping, and recommendation kernels. Because $a(h)$ is instance-independent, it cancels out entirely in the likelihood ratio on the common support.

Setting $d_{w,v} = \mu_{w,v} - \lambda_{w,v}$, the stopped log-likelihood ratio simplifies to

$$\begin{aligned} L_\tau := \log \frac{d\mathbb{P}_\mu^\tau}{d\mathbb{P}_\lambda^\tau} &= \sum_{t=1}^{\tau} \sum_{(w,v)\in M_t} \log \frac{f_{\mu_{w,v}}(Y_t(w,v))}{f_{\lambda_{w,v}}(Y_t(w,v))} \\ &= \sum_{t=1}^{\tau} \sum_{(w,v)\in M_t} \left[d_{w,v}\big(Y_t(w,v) - \mu_{w,v}\big) + \frac{1}{2} d_{w,v}^2 \right]. \end{aligned}$$

Decompose $L_\tau = Z_\tau + A_\tau$, where $Z_\tau := \sum_{t=1}^{\tau} \sum_{(w,v)\in M_t} d_{w,v}(Y_t(w,v) - \mu_{w,v})$ represents the martingale fluctuation term and $A_\tau := \frac{1}{2}\sum_{t=1}^{\tau} \sum_{(w,v)\in M_t} d_{w,v}^2$ denotes the cumulative predictable drift.

Under $\mathbb{P}_\mu$, the summands of Z_τ have conditional mean zero given the filtration $\mathcal{F}_{t-1}$ and the action M_t. Hence, the stopped process $(Z_{\tau\wedge T})_{T\geq 1}$ is a centered martingale with respect to the underlying filtration. Its quadratic variation satisfies

$$\mathbb{E}_\mu\big[Z_{\tau\wedge T}^2\big] = \mathbb{E}_\mu\left[\sum_{t=1}^{\tau\wedge T} \sum_{(w,v)\in M_t} d_{w,v}^2\right] \leq \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} n_{w,v} d_{w,v}^2 \leq \mathbb{E}_\mu[C_\tau] < \infty,$$

where the final inequality follows from $|d_{w,v}| \leq 1$. Since $\sup_T \mathbb{E}_\mu[Z_{\tau\wedge T}^2] \leq \mathbb{E}_\mu[C_\tau] < \infty$, the L^2-bounded martingale convergence theorem yields convergence in L^2. As $\tau < \infty$ almost surely, the limit is Z_τ. Consequently, $\mathbb{E}_\mu[Z_\tau] = 0$. Concurrently, the sequence of nondecreasing nonnegative random variables $A_{\tau\wedge T}$ converges monotonically to A_τ with $A_\tau \leq C_\tau/2$, ensuring L^1-convergence by the monotone convergence theorem. Taking expectations under $\mathbb{P}_\mu$ yields Wald's identity for the stopped likelihood ratio:

$$D_{\mathrm{KL}}(\mathbb{P}_\mu^\tau, \mathbb{P}_\lambda^\tau) = \mathbb{E}_\mu[L_\tau] = \mathbb{E}_\mu[A_\tau] = \frac{1}{2} \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} n_{w,v} d_{w,v}^2 = D_n(\mu,\lambda). \tag{49}$$

Combining equation 49 with the information threshold from Step 1 establishes that $D_n(\mu,\lambda) \geq k_\delta$. Taking the infimum over all $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$ immediately yields $I_{\mathrm{opt}}(n;\mu) \geq k_\delta$, which proves part (a).

Step 3: Derivation of the sample complexity lower bound. By the positive homogeneity of the directional information divergence,

$$D_{n/k_\delta}(\mu,\lambda) = \frac{1}{k_\delta} D_n(\mu,\lambda) \geq 1, \qquad \forall \lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu).$$

Hence, the scaled allocation vector n/k_δ is feasible for the unit-information optimization problem defining $H_{\mathrm{opt}}(\mu)$ in equation 43. By definition of the infimum, the total allocation satisfies

$$\sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} \frac{n_{w,v}}{k_\delta} \geq H_{\mathrm{opt}}(\mu),$$

which, upon multiplying through by k_δ, leads to

$$\mathbb{E}_\mu[C_\tau] = \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} n_{w,v} \geq k_\delta H_{\mathrm{opt}}(\mu) = T^\star_{\mathrm{opt}}(\mu,\delta),$$

where the final equality invokes the relation in equation 45. This completes the proof. □

The following lemma establishes foundational geometric and analytical properties of the unit-information allocation problem in equation 43, confirming that the normalization and subsequent scale reductions are rigorously well-defined whenever target-altering alternatives exist.

Lemma 5 (Structural properties of the information complexity)**.** *For any instance $\mu \in \mathcal{D}_\Gamma$, the feasible region of equation 43 is closed and convex, and $H_{\mathrm{opt}}(\mu) < \infty$. Furthermore, $H_{\mathrm{opt}}(\mu) > 0$ if and only if $\mathrm{Alt}_{\mathrm{opt}}(\mu) \neq \varnothing$.*

Proof. Convexity and closedness. For every fixed alternative $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$, the directional information divergence $x \mapsto D_x(\mu,\lambda)$ is linear in x. The feasible region of equation 43 can be expressed as the intersection of closed half-spaces with the nonnegative orthant:

$$\mathcal{X}(\mu) = \mathbb{R}_+^{N\times K} \cap \bigcap_{\lambda\in\mathrm{Alt}_{\mathrm{opt}}(\mu)} \left\{x \in \mathbb{R}^{N\times K} : D_x(\mu,\lambda) \geq 1\right\}.$$

As an arbitrary intersection of closed convex sets, $\mathcal{X}(\mu)$ is convex and closed.

If $\mathrm{Alt}_{\mathrm{opt}}(\mu) = \varnothing$, the infimum in equation 43 is unconstrained over $\mathbb{R}_+^{N\times K}$, whence the null allocation $x = 0$ is feasible, yielding $H_{\mathrm{opt}}(\mu) = 0$. For the remainder of the proof, assume $\mathrm{Alt}_{\mathrm{opt}}(\mu) \neq \varnothing$.

Finiteness. Under the reference instance $\mu \in \mathcal{D}_\Gamma$, all player-side social preference comparisons between distinct arms are strictly separated. Because there are finitely many pairs of arms and each socially induced utility $\theta_{u,v}(\lambda) = \sum_w \Gamma_{w,u}\lambda_{w,v}$ is a continuous linear functional of λ, there exists a stability radius $\varepsilon > 0$ such that any instance $\lambda \in \mathcal{D}_\Gamma$ satisfying $\|\lambda - \mu\|_\infty < \varepsilon$ induces the exact same strict ordinal preferences over arms for every player. Since arm-side preferences are fixed, the Gale–Shapley mapping is invariant on this ℓ_∞-ball, implying $F(\lambda) = F(\mu) = M$. Consequently, any target-altering alternative $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$ must lie outside this ball and satisfies

$$\sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} (\mu_{w,v} - \lambda_{w,v})^2 \geq \|\lambda - \mu\|_\infty^2 \geq \varepsilon^2.$$

Consider the isotropic candidate allocation $x_{w,v} = 2/\varepsilon^2$ for all $(w,v) \in \mathcal{U}\times\mathcal{V}$. For every $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$,

$$D_x(\mu,\lambda) = \frac{1}{\varepsilon^2} \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} (\mu_{w,v} - \lambda_{w,v})^2 \geq 1.$$

Thus, x is feasible for equation 43, which immediately implies the upper bound

$$H_{\mathrm{opt}}(\mu) \leq \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} x_{w,v} = \frac{2NK}{\varepsilon^2} < \infty.$$

Strict positivity. Fix an arbitrary $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$. Since the intrinsic utility space is bounded in $(0,1]$, we have $|\mu_{w,v} - \lambda_{w,v}| < 1$ for all (w,v). Every feasible allocation $x \in \mathcal{X}(\mu)$ must satisfy

$$1 \leq D_x(\mu,\lambda) = \frac{1}{2} \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} x_{w,v}(\mu_{w,v} - \lambda_{w,v})^2 < \frac{1}{2} \sum_{w\in\mathcal{U}} \sum_{v\in\mathcal{V}} x_{w,v}.$$

Taking the infimum over all feasible x yields $H_{\mathrm{opt}}(\mu) \geq 2 > 0$, completing the proof. □

This establishes the validity of the universal information-theoretic lower bound in Theorem 4. To convert this abstract variational characterization into explicit source-level sample allocations, we next analyze the geometric structure of the confounding set $\mathrm{Alt}_{\mathrm{opt}}(\mu)$ through the dual lenses of stability violation and loss of player-optimality.

D.1.3 Stability-Loss and Optimality-Loss Alternatives

The information constraint in Theorem 4 applies to every instance with a different target. To make this constraint explicit, we must distinguish two ways in which the target can change: M can lose stability, or it can remain stable while losing player-optimality. We first express stability through social comparisons, then describe the alternatives of each type.

For $a, b \in \mathcal{V} \cup \{\varnothing\}$, write

$$L_{u;a,b}(\lambda) = \theta_{u,a}(\lambda) - \theta_{u,b}(\lambda). \tag{50}$$

For any feasible matching R, define

$$\mathcal{B}(R) = \{(u, v) \in \mathcal{U} \times \mathcal{V} : v \neq R(u), \quad u \succ_v R^{-1}(v)\}. \tag{51}$$

These are exactly the pairs for which the arm would prefer to deviate from R. In particular, $\mathcal{B}(M) = \mathcal{C}_{\mathrm{blk}}(\mu)$ is the blocking-critical family defined in the main text (Definition 1). The following lemma identifies the player-side comparisons that rule out these deviations.

Lemma 6 (Stability comparisons)**.** *For $\lambda \in \mathcal{D}_\Gamma$, a feasible matching R is socially stable if and only if $L_{u;R(u),v}(\lambda) > 0$ for every $(u, v) \in \mathcal{B}(R)$. Moreover, if $(u, b) \in \mathcal{B}(M)$, then $M(u) \in \mathcal{V}$ and $\Delta_{u,b} := L_{u;M(u),b}(\mu) > 0$.*

Proof. A pair (u, v) can block R only if $v \neq R(u)$ and $u \succ_v R^{-1}(v)$, that is, $(u, v) \in \mathcal{B}(R)$. For each such pair, the arm-side condition for blocking is already satisfied. Since player preferences are strict, the pair does not block R if and only if

$$\theta_{u,R(u)}(\lambda) > \theta_{u,v}(\lambda).$$

Moreover, all pairs are acceptable, so individual rationality holds automatically. Thus R is socially stable exactly when these inequalities hold for every $(u, v) \in \mathcal{B}(R)$.

In particular, since M is socially stable under μ, every $(u, b) \in \mathcal{B}(M)$ satisfies

$$\theta_{u,M(u)}(\mu) > \theta_{u,b}(\mu) > 0.$$

The outside option has utility zero, so $M(u) \neq \varnothing$. The same inequality also gives $\Delta_{u,b} = \theta_{u,M(u)}(\mu) - \theta_{u,b}(\mu) > 0$. □

Accordingly, define the disjoint classes

$$\mathcal{A}_{\mathrm{blk}}(M) = \{\lambda \in \mathcal{D}_\Gamma : M \text{ is not socially stable under } \lambda\}, \tag{52}$$

$$\mathcal{A}_{\mathrm{loss}}(M) = \{\lambda \in \mathcal{D}_\Gamma : M \text{ is socially stable under } \lambda,\ F(\lambda) \neq M\}. \tag{53}$$

Lemma 6 characterizes loss of stability through reversed player-side comparisons. The target can also change while M remains stable, as the following example illustrates.

Example 2 (Loss of player-optimality without loss of stability)**.** *Consider two players and two arms, with*

$$\Gamma = \begin{pmatrix} 1/2 & 0 \\ 1/2 & 1 \end{pmatrix}, \qquad \mu = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}, \qquad \lambda = \begin{pmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{pmatrix}.$$

Arm a prefers player 1 to 2, and arm b prefers player 2 to 1. The induced social utilities are

$$\Gamma^\top \mu = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}, \qquad \Gamma^\top \lambda = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}.$$

Both instances have strict social preferences. Under μ, the matching $M = \{(1, a), (2, b)\}$ gives each player their first choice and is therefore player-optimal. It also gives each arm its first choice, so $\mathcal{B}(M) = \varnothing$ and M remains stable under λ. Under λ, however, $R = \{(1, b), (2, a)\}$ gives each player their first choice and becomes the player-optimal matching. Thus $\lambda \in \mathcal{A}_{\mathrm{loss}}(M)$.

Since $\mathcal{B}(M) = \varnothing$, a lower bound based only on blocking alternatives is zero. The target-changing alternative λ nevertheless imposes the information constraint

$$0.18(n_{1,a} + n_{1,b}) + 0.02(n_{2,a} + n_{2,b}) \geq k_\delta \tag{54}$$

by Theorem 4. The left-hand side is at most $0.18\,\mathbb{E}_\mu[C_\tau]$, giving $\mathbb{E}_\mu[C_\tau] \geq k_\delta/0.18 > 0$. Therefore, certifying stability alone can require strictly less information than identifying the player-optimal stable matching.

To characterize $\mathcal{A}_{\text{loss}}(M)$, we next use another stable matching as a witness: under the alternative preferences, it must improve upon M for every player whose partner changes.

Definition 3 (Optimality witnesses)**.** *For a feasible matching $R \neq M$, let*

$$\mathcal{W}_R(M) = \left\{ \lambda \in \mathcal{D}_\Gamma : \begin{array}{ll} L_{u;M(u),v}(\lambda) > 0, & (u,v) \in \mathcal{B}(M), \\ L_{u;R(u),v}(\lambda) > 0, & (u,v) \in \mathcal{B}(R), \\ L_{u;R(u),M(u)}(\lambda) > 0, & u \text{ with } R(u) \neq M(u) \end{array} \right\}. \tag{55}$$

Let $\mathcal{R}(M) = \{R \neq M : \mathcal{W}_R(M) \neq \varnothing\}$, where R ranges over feasible matchings.

The three rows in equation 55 require, respectively, stability of M, stability of R, and strict improvement for every player whose partner changes. All conditions are evaluated under λ; R need not be stable under μ.

Proposition 4 (Exact alternative decomposition)**.** *For the target $M = F(\mu)$,*

$$\text{Alt}_{\text{opt}}(\mu) = \mathcal{A}_{\text{blk}}(M) \,\dot\cup\, \mathcal{A}_{\text{loss}}(M), \tag{56}$$

$$\mathcal{A}_{\text{blk}}(M) = \bigcup_{(u,b)\in\mathcal{B}(M)} \big\{\lambda \in \mathcal{D}_\Gamma : L_{u;M(u),b}(\lambda) < 0\big\}, \tag{57}$$

$$\mathcal{A}_{\text{loss}}(M) = \bigcup_{R\in\mathcal{R}(M)} \mathcal{W}_R(M). \tag{58}$$

Here $\dot\cup$ denotes a disjoint union.

Proof. Since $F(\lambda)$ is stable, any instance that makes M unstable changes the target. The remaining target-changing instances are exactly those in $\mathcal{A}_{\text{loss}}(M)$, proving equation 56. Lemma 6 gives equation 57, since instability is equivalent to at least one reversed comparison.

To prove equation 58, first suppose $\lambda \in \mathcal{W}_R(M)$. Then M and R are both stable under λ. Every player weakly prefers R to M, and at least one strictly prefers it because $R \neq M$. Thus M is not player-optimal, so $\lambda \in \mathcal{A}_{\text{loss}}(M)$.

Conversely, let $\lambda \in \mathcal{A}_{\text{loss}}(M)$ and choose $R = F(\lambda)$. Both matchings are stable, and $R \neq M$. Player-optimality implies that every player weakly prefers R to M; strict preferences make this preference strict whenever partners differ. Together with Lemma 6, these conditions give $\lambda \in \mathcal{W}_R(M)$. Hence $\mathcal{W}_R(M) \neq \varnothing$, so $R \in \mathcal{R}(M)$, establishing the reverse inclusion. The argument also covers partial matchings, with $\varnothing$ treated as the outside option. □

Within $\mathcal{D}_\Gamma$, this decomposition describes all target-changing alternatives through finitely many systems of strict linear inequalities. We next evaluate the minimum information cost over these regions while respecting the bounds on intrinsic means.

D.1.4 Exact characterization by bounded quadratic programs

Proposition 4 establishes that any target-altering alternative induces one of two mutually exclusive regimes: the reference matching M either violates social stability, or remains stable while forfeiting player-optimality. For a given source allocation $x \in \mathbb{R}_+^{N\times K}$, the worst-case information divergence $I_{\text{opt}}(x;\mu)$ is therefore governed by the minimum over the infimum information costs across these two disjoint sets of alternatives. We compute these variational bounds explicitly by formulating a bounded convex quadratic program for each stability-violating comparison and each player-optimality witness.

Blocking alternatives: reversing a single stability comparison. Fix an allocation vector $x \geq 0$ and a blocking-critical pair $(u, b) \in \mathcal{B}(M)$. Let $a := M(u)$ denote player u's partner in M, let $g_w := \Gamma_{w,u} \geq 0$ denote the social influence weight of source w on u, and denote the true social gap by $\Delta := \Delta_{u,b} = \sum_{w \in \mathcal{U}} g_w(\mu_{w,a} - \mu_{w,b}) > 0$. By definition of $\mathcal{B}(M)$, arm b already strictly prefers player u over its assigned match $M^{-1}(b)$. Consequently, an admissible instance $\lambda \in \mathcal{D}_\Gamma$ causes (u, b) to block M if and only if player u's social preference between a and b is reversed, that is, $L_{u;a,b}(\lambda) < 0$. Under allocation x, the minimal information required to separate μ from all such blocking instances is characterized by the variational problem

$$\inf_{\lambda \in \mathcal{D}_\Gamma:\, L_{u;a,b}(\lambda) < 0} D_x(\mu, \lambda).$$

To compute this infimum, we first consider its closed relaxation over the compact parameter cube $[0,1]^{N \times K}$, minimizing $D_x(\mu, z)$ subject to $L_{u;a,b}(z) \leq 0$. Because the constraint $L_{u;a,b}(z) = \sum_{w \in \mathcal{U}} g_w(z_{w,a} - z_{w,b}) \leq 0$ depends exclusively on the columns of z indexed by arms a and b, for every $v \notin \{a, b\}$, setting $z_{w,v} = \mu_{w,v}$ preserves feasibility and sets the corresponding divergence term to zero without increasing the objective. Since $g_w \geq 0$ for all w, replacing $z_{w,a}$ with $\min\{z_{w,a}, \mu_{w,a}\}$ and $z_{w,b}$ with $\max\{z_{w,b}, \mu_{w,b}\}$ preserves feasibility while weakly decreasing $(z_{w,a} - \mu_{w,a})^2$ and $(z_{w,b} - \mu_{w,b})^2$. It therefore entails no loss of generality to parametrize the candidate perturbations as coordinatewise downward shifts $r_w \geq 0$ on arm a and upward shifts $s_w \geq 0$ on arm b:

$$z_{w,a} = \mu_{w,a} - r_w, \qquad z_{w,b} = \mu_{w,b} + s_w, \qquad r_w, s_w \geq 0.$$

These perturbations shrink the social utility gap from Δ to $\Delta - \sum_{w \in \mathcal{U}} g_w(r_w + s_w)$. The perturbed comparison equals $\Delta - \sum_w g_w(r_w + s_w)$. Thus the closed reversal constraint $L_{u;a,b}(z) \leq 0$ is equivalent to $\sum_w g_w(r_w + s_w) \geq \Delta$; strict reversal requires a strict inequality.

Under unit-variance Gaussian observations, the single-observation Kullback–Leibler divergences incurred by these shifts are $r_w^2/2$ and $s_w^2/2$, respectively. Aggregating these contributions under the sampling weights x yields the following box-constrained convex quadratic program:

$$\begin{aligned} B_{u,b}(x; \mu) := \min_{r,s \geq 0} \quad & \frac{1}{2} \sum_{w \in \mathcal{U}} \left(x_{w,a} r_w^2 + x_{w,b} s_w^2 \right) \\ \text{subject to} \quad & r_w \leq \mu_{w,a}, \qquad s_w \leq 1 - \mu_{w,b}, \qquad \forall w \in \mathcal{U}, \\ & \sum_{w \in \mathcal{U}} g_w(r_w + s_w) \geq \Delta. \end{aligned} \tag{59}$$

The coordinate bounds guarantee that the perturbed instance z remains within the closure of the mean-parameter domain $[0,1]^{N \times K}$. The following lemma establishes that this closed relaxation is tight, in that its optimal value coincides with the infimum over all admissible instances in $\mathcal{D}_\Gamma$ that strictly reverse the comparison.

Lemma 7 (Exactness of the bounded blocking projection)**.** *For every blocking-critical pair $(u, b) \in \mathcal{B}(M)$ and every nonnegative allocation $x \geq 0$, with $a := M(u)$,*

$$B_{u,b}(x; \mu) = \inf_{\lambda \in \mathcal{D}_\Gamma:\, L_{u;a,b}(\lambda) < 0} D_x(\mu, \lambda).$$

Proof. By the coordinatewise perturbation reduction established prior to equation 59, $B_{u,b}(x; \mu)$ coincides with the minimum of the convex quadratic function $z \mapsto D_x(\mu, z)$ over the compact cube $[0,1]^{N \times K}$ subject to the closed linear half-space constraint $L_{u;a,b}(z) \leq 0$. Every admissible strict reversal $\lambda \in \mathcal{D}_\Gamma$ satisfying $L_{u;a,b}(\lambda) < 0$ belongs to $(0,1]^{N \times K} \subset [0,1]^{N \times K}$ and obeys $L_{u;a,b}(\lambda) \leq 0$. Hence, every such λ is feasible for this closed relaxation, which immediately implies

$$\inf_{\lambda \in \mathcal{D}_\Gamma:\, L_{u;a,b}(\lambda) < 0} D_x(\mu, \lambda) \geq B_{u,b}(x; \mu).$$

It remains to establish the reverse inequality. By compactness of the feasible region and continuity of $D_x(\mu, \cdot)$, the closed relaxation admits an optimal solution $z^\star \in [0,1]^{N \times K}$ such that $L_{u;a,b}(z^\star) \leq 0$ and $D_x(\mu, z^\star) = B_{u,b}(x; \mu)$. To show that $z^\star$ can be approximated arbitrarily closely from within the admissible strict-reversal region $\{\lambda \in \mathcal{D}_\Gamma : L_{u;a,b}(\lambda) < 0\}$, we construct a strictly admissible anchor instance λ^0. Let $\lambda^0 \in (0,1]^{N \times K}$ be obtained from μ by transposing columns a and b. Because

$\Theta(\lambda^0) = \Gamma^\top \lambda^0$, swapping columns a and b of the intrinsic utility matrix simply transposes columns a and b of the socially induced utility profile $\Theta(\mu)$. Since $\mu \in \mathcal{D}_\Gamma$, all player-side social utilities under μ are pairwise distinct across distinct arms; swapping two columns preserves this pairwise distinctness, whence $\lambda^0 \in \mathcal{D}_\Gamma$. Moreover,

$$L_{u;a,b}(\lambda^0) = \theta_{u,a}(\lambda^0) - \theta_{u,b}(\lambda^0) = \theta_{u,b}(\mu) - \theta_{u,a}(\mu) = -\Delta < 0.$$

For $t \in (0,1]$, define the line segment $z_t = (1-t)z^\star + t\lambda^0$. Because $z^\star \in [0,1]^{N\times K}$ and $\lambda^0 \in (0,1]^{N\times K}$, the convex combination satisfies $z_t \in (0,1]^{N\times K}$ for all $t \in (0,1]$. By linearity of the functional $L_{u;a,b}$,

$$L_{u;a,b}(z_t) = (1-t)L_{u;a,b}(z^\star) + tL_{u;a,b}(\lambda^0) \le -t\Delta < 0.$$

Furthermore, for any player $i \in \mathcal{U}$ and distinct arms $v \neq v'$, the difference map $t \mapsto \theta_{i,v}(z_t) - \theta_{i,v'}(z_t)$ is affine in t. At $t = 1$, this map evaluates to $\theta_{i,v}(\lambda^0) - \theta_{i,v'}(\lambda^0) \neq 0$ because $\lambda^0 \in \mathcal{D}_\Gamma$. A non-constant affine function on the real line vanishes at at most one point. Since the collection of player-arm triples $\{(i,v,v') : i \in \mathcal{U},\ v \neq v'\}$ is finite, the condition $\theta_{i,v}(z_t) = \theta_{i,v'}(z_t)$ can hold for at most finitely many values of $t \in (0,1]$. Consequently, there exists $t_0 \in (0,1]$ such that $z_t \in \mathcal{D}_\Gamma$ for all $t \in (0,t_0)$.

Every such z_t is therefore admissible and satisfies the strict reversal condition $L_{u;a,b}(z_t) < 0$. Taking the limit along this sequence and invoking the continuity of $D_x(\mu,\cdot)$,

$$\inf_{\lambda \in \mathcal{D}_\Gamma:\, L_{u;a,b}(\lambda) < 0} D_x(\mu,\lambda) \le \lim_{t \downarrow 0} D_x(\mu, z_t) = D_x(\mu, z^\star) = B_{u,b}(x;\mu).$$

Combining both directional inequalities concludes the proof. □

Optimality-loss alternatives: satisfying a witness matching. Fix an admissible candidate matching $R \in \mathcal{R}(M)$. To determine the minimal information cost over the witness region $\mathcal{W}_R(M)$, we consider its closed convex relaxation over the parameter cube $[0,1]^{N\times K}$, relaxing the strict stability and improvement inequalities to weak linear constraints:

$$\mathcal{P}_R(M) := \left\{ z \in [0,1]^{N\times K} : \begin{array}{rl} L_{u;M(u),v}(z) \ge 0, & \forall (u,v) \in \mathcal{B}(M) \\ L_{u;R(u),v}(z) \ge 0, & \forall (u,v) \in \mathcal{B}(R) \\ L_{u;R(u),M(u)}(z) \ge 0, & \forall u \text{ such that } R(u) \neq M(u) \end{array} \right\}.$$

The corresponding relaxation cost under allocation $x \ge 0$ is given by

$$J_R(x;\mu) := \min_{z \in \mathcal{P}_R(M)} D_x(\mu, z). \tag{60}$$

The following lemma establishes that this relaxation is exact.

Lemma 8 (Closure and exactness for witness regions)**.** *For every $R \in \mathcal{R}(M)$ and every allocation $x \ge 0$,*

$$\overline{\mathcal{W}_R(M)} = \mathcal{P}_R(M) \qquad \textit{and} \qquad \inf_{\lambda \in \mathcal{W}_R(M)} D_x(\mu,\lambda) = J_R(x;\mu). \tag{61}$$

Proof. Because $\mathcal{P}_R(M)$ is defined by weak inequalities on the compact set $[0,1]^{N\times K}$, it is closed. Since $\mathcal{W}_R(M) \subseteq \mathcal{P}_R(M)$, it follows that $\overline{\mathcal{W}_R(M)} \subseteq \mathcal{P}_R(M)$.

To establish the reverse inclusion, fix any $z \in \mathcal{P}_R(M)$. Since $R \in \mathcal{R}(M)$, the witness set $\mathcal{W}_R(M)$ is nonempty, so there exists an admissible instance $\lambda^0 \in \mathcal{W}_R(M) \subset (0,1]^{N\times K}$. For $t \in (0,1]$, define the line segment $z_t = (1-t)z + t\lambda^0$. Because $z \in [0,1]^{N\times K}$ and $\lambda^0 \in (0,1]^{N\times K}$, we have $z_t \in (0,1]^{N\times K}$. Moreover, for every defining comparison functional $L \in \{L_{u;M(u),v}, L_{u;R(u),v}, L_{u;R(u),M(u)}\}$ defining $\mathcal{W}_R(M)$, linearity yields

$$L(z_t) = (1-t)L(z) + tL(\lambda^0) \ge tL(\lambda^0) > 0.$$

For any other pair of arms $a \neq b$, the evaluation $t \mapsto \theta_{u,a}(z_t) - \theta_{u,b}(z_t)$ is affine in t and non-zero at $t = 1$ (as $\lambda^0 \in \mathcal{D}_\Gamma$). Because a non-constant univariate affine map vanishes at most once, and the set of player-arm triples is finite, ties can occur for at most finitely many values of t. Thus, there exists $t_0 \in (0,1]$ such that $z_t \in \mathcal{W}_R(M)$ for all $t \in (0,t_0)$. Taking $t \downarrow 0$ yields $z_t \to z$, which proves $\mathcal{P}_R(M) \subseteq \overline{\mathcal{W}_R(M)}$, establishing the set equality.

The cost equality follows immediately: by continuity of $z \mapsto D_x(\mu,z)$ on $\mathbb{R}^{N\times K}$, the infimum over any dense subset of a compact set equals the minimum over its closure. □

The requirement $R \in \mathcal{R}(M)$ is essential: nonemptiness of the closed relaxation $\mathcal{P}_R(M)$ does not by itself guarantee that the strict witness region $\mathcal{W}_R(M)$ is nonempty.

Unifying stability and optimality barriers. The decomposition established in Proposition 4 allows us to reduce the abstract information divergence $I_{\mathrm{opt}}(x;\mu)$ over the continuum of admissible alternatives $\mathrm{Alt}_{\mathrm{opt}}(\mu)$ to a finite collection of bounded convex quadratic programs.

Theorem 5 (Finite variational characterization of the information divergence)**.** *For every nonnegative allocation $x \in \mathbb{R}_+^{N\times K}$, the worst-case information divergence satisfies*

$$I_{\mathrm{opt}}(x;\mu) = \min\left\{ \min_{(u,b)\in\mathcal{B}(M)} B_{u,b}(x;\mu), \quad \min_{R\in\mathcal{R}(M)} J_R(x;\mu) \right\}, \tag{62}$$

under the convention that the minimum over an empty set is $+\infty$. Consequently, the unit-information constraints defining $H_{\mathrm{opt}}(\mu)$ in equation 43 are equivalently characterized by the finite system of convex constraints:

$$\begin{aligned} B_{u,b}(x;\mu) &\geq 1, \quad \forall (u,b) \in \mathcal{B}(M), \\ J_R(x;\mu) &\geq 1, \quad \forall R \in \mathcal{R}(M). \end{aligned} \tag{63}$$

Proof. By Proposition 4, the set of target-altering alternatives admits the finite union $\mathrm{Alt}_{\mathrm{opt}}(\mu) = \big(\bigcup_{(u,b)\in\mathcal{B}(M)}\{\lambda \in \mathcal{D}_\Gamma : L_{u;M(u),b}(\lambda) < 0\}\big) \cup \big(\bigcup_{R\in\mathcal{R}(M)} \mathcal{W}_R(M)\big)$. Because the infimum of a functional over a finite union of sets equals the minimum of its infima over each constituent set, identity equation 62 follows immediately from Lemma 7 and Lemma 8. Enforcing $I_{\mathrm{opt}}(x;\mu) \geq 1$ is therefore equivalent to requiring that each constituent infimum satisfies the lower bound, which yields equation 63.

Finally, for each fixed $z \in [0,1]^{N\times K}$, the mapping $x \mapsto D_x(\mu, z)$ is affine. As pointwise infima of affine functions, both $x \mapsto B_{u,b}(x;\mu)$ and $x \mapsto J_R(x;\mu)$ are concave on $\mathbb{R}_+^{N\times K}$. The constraint regions in equation 63 are thus closed convex upper-level sets of concave functions, confirming that the overall feasible set is convex. □

Deciding admissibility of optimality witnesses. To operationalize the characterization in Theorem 5, one must verify whether a candidate matching $R \neq M$ belongs to $\mathcal{R}(M)$. This is equivalent to determining whether there exists an admissible instance under which M and R are simultaneously socially stable, with all displaced players strictly preferring R over M.

Let $z \in [0,1]^{N\times K}$ denote a candidate intrinsic mean matrix, and introduce a scalar margin $\rho \geq 0$. We enforce that each required strict social comparison exhibits a clearance of at least ρ, while all coordinates of z remain bounded away from the boundary $\{0,1\}$ by ρ. This margin is maximized via the following linear program:

$$\begin{aligned} \max_{z,\rho} \quad & \rho \\ \text{subject to} \quad & L_{u;M(u),v}(z) \geq \rho, && \forall (u,v) \in \mathcal{B}(M), \\ & L_{u;R(u),v}(z) \geq \rho, && \forall (u,v) \in \mathcal{B}(R), \\ & L_{u;R(u),M(u)}(z) \geq \rho, && \forall u \text{ such that } R(u) \neq M(u), \\ & \rho \leq z_{w,v} \leq 1-\rho, && \forall (w,v) \in \mathcal{U}\times\mathcal{V}, \\ & 0 \leq \rho \leq \tfrac{1}{2}. \end{aligned} \tag{64}$$

The matching R is an admissible witness (i.e., $R \in \mathcal{R}(M)$) if and only if the optimal value of equation 64 is strictly positive.

To verify this equivalence, suppose first that $R \in \mathcal{R}(M)$, so that there exists $\lambda^0 \in \mathcal{W}_R(M) \subset (0,1)^{N\times K}$. Consider the interior convex combination $\lambda_\epsilon = (1-\epsilon)\lambda^0 + \epsilon\frac{1}{2}\mathbf{1}_{N\times K}$ for $\epsilon \in (0,1)$. Because λ^0 satisfies all defining inequalities of $\mathcal{W}_R(M)$ strictly, choosing $\epsilon > 0$ sufficiently small preserves strict positivity of these inequalities while contracting all coordinates strictly into the open hypercube $(0,1)^{N\times K}$. Consequently, λ_ϵ attains a strictly positive common slack, certifying that the optimal value of equation 64 is strictly positive.

Conversely, assume that equation 64 admits an optimal solution $(z^\star, \rho^\star)$ with $\rho^\star > 0$. Then $z^\star \in [\rho^\star, 1-\rho^\star]^{N\times K} \subset (0,1)^{N\times K}$, and all stability and player-improvement inequalities defining

$\mathcal{W}_R(M)$ hold with positive clearance at least $\rho^\star$. By continuity, these inequalities remain strictly satisfied within an open Euclidean ball $\mathcal{B}(z^\star, \epsilon) \subset (0,1)^{N\times K}$ for small $\epsilon > 0$. It remains only to ensure that the unconstrained player-arm comparisons avoid ties. For any player u and distinct arms $a \neq b$, the tie condition $\theta_{u,a}(z) = \theta_{u,b}(z)$ defines the hyperplane

$$\sum_{w\in\mathcal{U}} \Gamma_{w,u}(z_{w,a} - z_{w,b}) = 0.$$

Because $\sum_{w\in\mathcal{U}} \Gamma_{w,u} = 1$ with $\Gamma_{w,u} \geq 0$, the normal vector of this hyperplane is non-zero, making it a proper affine subspace of dimension $NK - 1$. The union of finitely many such tie hyperplanes has zero Lebesgue measure and cannot cover the open ball $\mathcal{B}(z^\star, \epsilon)$. Selecting any point in $\mathcal{B}(z^\star, \epsilon)$ that avoids this finite collection of hyperplanes yields an admissible instance in $\mathcal{W}_R(M)$, which proves $\mathcal{W}_R(M) \neq \varnothing$.

Although the number of candidate matchings R can be exponential in N, program equation 64 provides an exact, finite linear-programming oracle to certify whether any given matching induces target ambiguity. In the following subsection, we turn to the stability-violating program $B_{u,b}(x;\mu)$, for which we derive an explicit analytical solution and its corresponding inverse-count representation.

D.1.5 Explicit constraints for blocking comparisons

Throughout this subsection, fix $(u,b) \in \mathcal{B}(M)$ and write $a = M(u)$, $g_w = \Gamma_{w,u}$, and $\Delta = \Delta_{u,b} > 0$. We first solve equation 59 for allocations that are positive on all influential source pairs. When the optimal perturbations obtained without the coordinate upper bounds already satisfy those bounds, the optimal value reduces to an inverse-count expression. The case where some source pairs receive zero allocation is treated separately below.

Proposition 5 (Clipped optimal perturbations)**.** *Suppose $x_{w,a}, x_{w,b} > 0$ whenever $g_w > 0$. There exists a scalar $\eta > 0$ such that the perturbations*

$$r_w^\star = \min\left\{\mu_{w,a}, \frac{\eta g_w}{x_{w,a}}\right\}, \qquad s_w^\star = \min\left\{1-\mu_{w,b}, \frac{\eta g_w}{x_{w,b}}\right\} \quad (g_w > 0) \tag{65}$$

satisfy $\sum_w g_w(r_w^\star + s_w^\star) = \Delta$. With $r_w^\star = s_w^\star = 0$ when $g_w = 0$, these perturbations solve equation 59.

Proof. Let

$$Q(r,s) = \frac{1}{2}\sum_w \left(x_{w,a} r_w^2 + x_{w,b} s_w^2\right)$$

denote the original objective. For $\eta \geq 0$, define the auxiliary objective

$$\mathcal{L}_\eta(r,s) = Q(r,s) + \eta\left(\Delta - \sum_w g_w(r_w + s_w)\right).$$

Every feasible perturbation satisfies $\mathcal{L}_\eta(r,s) \leq Q(r,s)$. We seek an η for which a box-constrained minimizer of $\mathcal{L}_\eta$ achieves total gap reduction exactly Δ, making this lower bound exact.

Minimizing $\mathcal{L}_\eta$ over the box separates across coordinates. For $g_w > 0$, the scalar quadratics have unconstrained minimizers $\eta g_w/x_{w,a}$ and $\eta g_w/x_{w,b}$. Restricting them to their permitted intervals gives equation 65; coordinates with $g_w = 0$ can be set to zero.

For these minimizers, the total gap reduction $\sum_w g_w(r_w^\star + s_w^\star)$ is continuous and nondecreasing in η. It equals zero at $\eta = 0$ and, for sufficiently large η, equals

$$\sum_w g_w(\mu_{w,a} + 1 - \mu_{w,b}) = \Delta + 1 > \Delta,$$

where we used $\sum_w g_w = 1$. Hence some $\eta > 0$ makes the gap constraint tight.

For this choice, every feasible (r,s) satisfies

$$Q(r,s) \geq \mathcal{L}_\eta(r,s) \geq \mathcal{L}_\eta(r^\star, s^\star) = Q(r^\star, s^\star).$$

The middle inequality follows from minimization over the box, and the equality follows from the tight gap constraint. Thus $(r^\star, s^\star)$ is optimal for equation 59. □

When no clipping is needed, the multiplier η and the optimal value have explicit expressions.

Corollary 1 (Inverse-count formula). *Let $x \geq 0$ satisfy $x_{w,a}, x_{w,b} > 0$ whenever $g_w > 0$. Define*

$$q_{u,b}(x) = \sum_{w:g_w>0} g_w^2 \left(\frac{1}{x_{w,a}} + \frac{1}{x_{w,b}} \right) > 0, \tag{66}$$

and

$$r_w^0(x) = \frac{\Delta g_w}{q_{u,b}(x) x_{w,a}}, \qquad s_w^0(x) = \frac{\Delta g_w}{q_{u,b}(x) x_{w,b}} \quad (g_w > 0), \tag{67}$$

with both perturbations set to zero when $g_w = 0$. If $r_w^0(x) \leq \mu_{w,a}$ and $s_w^0(x) \leq 1 - \mu_{w,b}$ for every w with $g_w > 0$, then $(r^0(x), s^0(x))$ solves equation 59, and

$$B_{u,b}(x; \mu) = \frac{\Delta^2}{2 q_{u,b}(x)}. \tag{68}$$

For the expected counts n in Theorem 4, assume $\mathbb{E}_\mu[C_\tau] < \infty$. If the same positivity and coordinate-bound conditions hold with $x = n$, then

$$\sum_{w:\Gamma_{w,u}>0} \Gamma_{w,u}^2 \left(\frac{1}{n_{w,a}} + \frac{1}{n_{w,b}} \right) \leq \frac{\Delta_{u,b}^2}{2\mathrm{kl}(1-\delta, \delta)}. \tag{69}$$

Proof. Choose $\eta = \Delta / q_{u,b}(x)$. The candidate perturbations then satisfy

$$\sum_w g_w \big(r_w^0(x) + s_w^0(x) \big) = \eta q_{u,b}(x) = \Delta.$$

The assumed coordinate bounds make them equal to the perturbations in equation 65. Proposition 5 therefore establishes their optimality. Substituting them into the objective gives

$$\begin{aligned} B_{u,b}(x; \mu) &= \frac{1}{2} \sum_w \left[x_{w,a} \big(r_w^0(x) \big)^2 + x_{w,b} \big(s_w^0(x) \big)^2 \right] \\ &= \frac{\eta^2 q_{u,b}(x)}{2} = \frac{\Delta^2}{2 q_{u,b}(x)}. \end{aligned}$$

Every admissible strict reversal of this comparison changes the target. Theorem 4 and Lemma 7 thus give

$$B_{u,b}(n; \mu) \geq \mathrm{kl}(1-\delta, \delta).$$

Applying equation 68 at $x = n$ and rearranging yields equation 69. □

The coordinate bounds in Corollary 1 must be checked explicitly: the candidate perturbations in equation 67 need not satisfy them. If the candidate perturbations violate these bounds, $\Delta^2/(2q_{u,b}(x))$ is the value of a relaxed problem and may be smaller than $B_{u,b}(x; \mu)$. The information lower bound on $B_{u,b}(x; \mu)$ therefore does not imply equation 69; the bounded program equation 59 remains valid.

We now allow zero allocations. In the bounded program, r_w incurs no cost when $x_{w,a} = 0$, and s_w incurs no cost when $x_{w,b} = 0$. Their maximum contributions to reducing the gap are $g_w \mu_{w,a}$ and $g_w(1 - \mu_{w,b})$, respectively.

Proposition 6 (Zero allocations). *For any $x \geq 0$, define*

$$F_0(x) = \sum_{w:x_{w,a}=0} g_w \mu_{w,a} + \sum_{w:x_{w,b}=0} g_w (1 - \mu_{w,b}), \tag{70}$$

the maximum gap reduction achievable at zero cost. Then

$$B_{u,b}(x; \mu) = 0 \iff F_0(x) \geq \Delta.$$

If $F_0(x) < \Delta$, an optimal solution is obtained by setting $r_w = \mu_{w,a}$ whenever $x_{w,a} = 0$ and $s_w = 1 - \mu_{w,b}$ whenever $x_{w,b} = 0$, then applying the clipped construction to the remaining variables with positive allocations and residual gap $\Delta - F_0(x)$.

Proof. Setting all zero-allocation perturbations to their upper bounds and all other perturbations to zero achieves gap reduction $F_0(x)$ at zero cost. This is feasible when $F_0(x) \geq \Delta$. Conversely, the bounded program attains its minimum, and any zero-cost minimizer must have zero perturbation at every positive-allocation coordinate. Its gap reduction is therefore at most $F_0(x)$, proving the equivalence.

Now suppose $F_0(x) < \Delta$. Increasing a zero-allocation perturbation to its upper bound preserves feasibility without changing the cost. Thus an optimal solution can use all of this free capacity, leaving the positive-allocation variables to supply $\Delta - F_0(x)$. Their total available capacity is

$$\Delta + 1 - F_0(x) > \Delta - F_0(x).$$

The coordinatewise construction in the proof of Proposition 5 therefore applies to these remaining variables and yields the claimed residual solution. □

D.1.6 Blocking, Matching, and Ranking Complexities

Let $\mathcal{A}_{\mathrm{rank}}(\mu)$ denote the instances in $\mathcal{D}_\Gamma$ whose social rankings differ from those under μ. Define H_{blk} and H_{rank} by replacing $\mathrm{Alt}_{\mathrm{opt}}(\mu)$ in equation 43 with $\mathcal{A}_{\mathrm{blk}}(M)$ and $\mathcal{A}_{\mathrm{rank}}(\mu)$, respectively. Making M unstable changes the target, whereas preserving all social rankings preserves the target when arm preferences are fixed. Hence

$$\mathcal{A}_{\mathrm{blk}}(M) \subseteq \mathrm{Alt}_{\mathrm{opt}}(\mu) \subseteq \mathcal{A}_{\mathrm{rank}}(\mu).$$

Larger alternative sets impose more information constraints, so

$$H_{\mathrm{blk}}(\mu) \leq H_{\mathrm{opt}}(\mu) \leq H_{\mathrm{rank}}(\mu). \tag{71}$$

The first inequality can be strict, as Example 2 shows. The second explains why an upper bound based on recovering all social rankings need not match the lower bound for target identification.

The distinction between loss of stability and loss of player-optimality also arises without social influence. Pagare et al. (2025) study pure exploration with pairwise sampling, analyze unique stable matchings, and discuss optimality-changing alternatives that preserve stability. Athanasopoulos et al. (2026) derive a lower bound for optimal stable matching under two-sided uncertainty, using matching rounds as the sampling unit.

D.2 Extension to Matching-Round Complexity

An individual evaluation provides one source-pair observation, whereas a matching round may provide several observations simultaneously. To distinguish these resources, let p describe allocation proportions per observation and y describe sampling rates per round. Define

$$\begin{aligned} \mathcal{S} &= \left\{ p \geq 0 : \sum_{w,v} p_{w,v} = 1 \right\}, \\ \mathcal{Q} &= \left\{ y \geq 0 : \sum_{v} y_{w,v} \leq 1 \ (w \in \mathcal{U}), \quad \sum_{w} y_{w,v} \leq 1 \ (v \in \mathcal{V}) \right\}. \end{aligned} \tag{72}$$

The unit-sum constraint normalizes one observation. The row and column constraints reflect that each player and each arm can participate at most once per round.

Proposition 7 (Observation and round normalizations)**.** *Suppose* $\mathrm{Alt}_{\mathrm{opt}}(\mu) \neq \varnothing$. *Then*

$$H_{\mathrm{opt}}(\mu) = \left[\sup_{p \in \mathcal{S}} I_{\mathrm{opt}}(p; \mu) \right]^{-1}. \tag{73}$$

Moreover, any procedure that is δ*-correct over* $\mathcal{D}_\Gamma$ *and satisfies* $0 < \mathbb{E}_\mu[\tau] < \infty$ *obeys*

$$\mathbb{E}_\mu[\tau] \geq \frac{k_\delta}{\sup_{y \in \mathcal{Q}} I_{\mathrm{opt}}(y; \mu)}. \tag{74}$$

Proof. Every nonzero allocation can be written as $x = sp$, where $s = \sum_{w,v} x_{w,v}$ and $p = x/s \in \mathcal{S}$. Positive homogeneity gives $I_{\mathrm{opt}}(x;\mu) = sI_{\mathrm{opt}}(p;\mu)$. Thus, for a fixed p with $I_{\mathrm{opt}}(p;\mu) > 0$, the unit-information constraint holds exactly when $s \geq 1/I_{\mathrm{opt}}(p;\mu)$. Minimizing over such p proves equation 73.

For the round bound, let $n_{w,v} = \mathbb{E}_\mu[N_{w,v}(\tau)]$. Matching feasibility gives, pathwise, $\sum_v N_{w,v}(\tau) \leq \tau$ and $\sum_w N_{w,v}(\tau) \leq \tau$. Taking expectations therefore yields $y = n/\mathbb{E}_\mu[\tau] \in \mathcal{Q}$. Also, $C_\tau \leq \min\{N,K\}\tau$ ensures $\mathbb{E}_\mu[C_\tau] < \infty$. Theorem 4 and homogeneity now give

$$k_\delta \leq I_{\mathrm{opt}}(n;\mu) = \mathbb{E}_\mu[\tau] I_{\mathrm{opt}}(y;\mu) \leq \mathbb{E}_\mu[\tau] \sup_{y' \in \mathcal{Q}} I_{\mathrm{opt}}(y';\mu),$$

which yields equation 74.

Both suprema are positive: normalizing a feasible allocation from Lemma 5 gives positive information in $\mathcal{S} \subseteq \mathcal{Q}$. They are finite because $\mathrm{Alt}_{\mathrm{opt}}(\mu)$ is nonempty and the bounded means imply, for every $y \in \mathcal{Q}$, $I_{\mathrm{opt}}(y;\mu) \leq \frac{1}{2}\sum_{w,v} y_{w,v} \leq \frac{1}{2}\min\{N,K\}$. Thus all reciprocals above are well defined. □

The round bound is immediate when $\mathbb{E}_\mu[\tau] = \infty$. If $\mathrm{Alt}_{\mathrm{opt}}(\mu) = \varnothing$, the target is constant over $\mathcal{D}_\Gamma$ and zero observations suffice.

Dividing an observation lower bound by $\min\{N,K\}$ also gives a valid generic round bound. The normalization over $\mathcal{Q}$ accounts for the individual row and column capacity constraints directly.

E Analysis and Extensions of SAE-Q

This appendix proves the sample-complexity guarantee for Algorithm 1 (Theorem 2). We use the setting of Section 4 throughout: the platform observes intrinsic evaluations, social utilities are $\Theta = \Gamma^\top \mu$, and the final allocation uses the *cumulative* empirical utilities $\widehat{\Theta}_\tau = \Gamma^\top \widehat{\mu}(\tau)$. The cost is the total number of individual evaluations C_τ.

Notation and the confidence budget. Write $\mathcal{U} = [N]$ and $\mathcal{V} = [K]$. For $N \geq 1$ and $K \geq 2$, define

$$\begin{gathered} d := NK, \qquad \mathcal{C} := \{(u,a,b) : u \in \mathcal{U},\ a,b \in \mathcal{V},\ a < b\}, \qquad J := |\mathcal{C}| = N\binom{K}{2}, \\ g_c := \theta_{u,a} - \theta_{u,b}, \qquad \Delta_c := |g_c|, \qquad \Delta_{\min} := \min_{c \in \mathcal{C}} \Delta_c > 0, \end{gathered} \tag{75}$$

where $c = (u,a,b)$. Thus $\mathcal{C}$ contains all unordered arm comparisons for every player; it is distinct from the blocking-critical set $\mathcal{C}_{\mathrm{blk}}(\mu)$.

Write $e = (w,v)$ for a source pair and $x_e = x_{w,v}$. For $c = (u,a,b)$, define

$$\begin{gathered} q_{c,(w,v)} := \Gamma_{w,u}\big(\mathbf{1}\{v = a\} - \mathbf{1}\{v = b\}\big), \\ S_c := \{(w,v) : \Gamma_{w,u} > 0,\ v \in \{a,b\}\}, \qquad s := \max_{c \in \mathcal{C}} |S_c|, \\ V_c(x) = \sum_{e \in S_c} \frac{q_{c,e}^2}{x_e} = \sum_{w \in \mathcal{U}} \Gamma_{w,u}^2 \left(\frac{1}{x_{w,a}} + \frac{1}{x_{w,b}}\right). \end{gathered} \tag{76}$$

Terms with zero coefficients are omitted, and a positive coefficient divided by zero is interpreted as $+\infty$. These coefficients express each gap as $g_c = \sum_{(w,v) \in S_c} q_{c,(w,v)} \mu_{w,v}$. They are used only in the proof; the allocation problem remains the one in Algorithm 1.

All logarithms below are natural unless a base is displayed. An explicit choice of the constant B_0 used in Algorithm 1 and Theorem 2 is

$$\begin{gathered} A_0 := \frac{64J}{\delta}, \qquad b_0 := \log A_0, \qquad \ell_0 := \log\big(64d(b_0 + s)\big), \\ B_0 := 4b_0 + 16s\ell_0, \quad \beta_m := B_0(m+1), \quad \epsilon_m := 2^{-m}. \end{gathered} \tag{77}$$

This confidence budget and precision schedule are fixed before sampling using only the known dimensions, Γ, and δ; they do not depend on unknown gaps or realized counts.

Positive anchoring and the social operator in Section 4 give

$$\Gamma_{w,u} \geq 0, \qquad \sum_{w \in \mathcal{U}} \Gamma_{w,u} = 1, \qquad \Gamma_{u,u} \geq \alpha_u > 0. \tag{78}$$

Indeed, the Neumann expansion of $\Gamma = A[I - P(I - A)]^{-1}$ is nonnegative and has zeroth term A; also $[I - (I - A)P^\top]\mathbf{1} = A\mathbf{1}$ implies $\Gamma^\top \mathbf{1} = \mathbf{1}$. Consequently, $\sum_w \Gamma_{w,u}^2 \leq 1$, $\theta_{u,v} \in (0, 1]$, $0 < \Delta_c \leq 1$, and $2 \leq s \leq 2N$.

When forming the final preference lists, the known outside option is placed last for every player, even if an empirical utility is negative, and the known arm preferences are retained. No restriction $N \leq K$ is needed.

When $K = 1$, there are no arm comparisons and $\Delta_{\min}$ is undefined. The known arm preference, together with every player's preference for the only arm over being unmatched, determines $M^\star$ without evaluations. This case is handled before running Algorithm 1; the statement involving M_Δ concerns $K \geq 2$.

E.1 Bipartite scheduling

Algorithm 2 SCHEDULE(m): scheduling oracle for source demands

Require: Integer demand matrix $m \in \mathbb{Z}_+^{|\mathcal{U}| \times |\mathcal{V}|}$

1: Let
$$D(m) = \max\left\{ \max_{w \in \mathcal{U}} \sum_{v \in \mathcal{V}} m_{wv},\ \max_{v \in \mathcal{V}} \sum_{w \in \mathcal{U}} m_{wv} \right\}.$$
2: Construct the bipartite multigraph G_m with m_{wv} parallel edges between each $w \in \mathcal{U}$ and $v \in \mathcal{V}$.
3: Use a $D(m)$-edge-coloring of G_m to partition its edges into color classes $E_1, \ldots, E_{D(m)}$.
4: **for** $\ell = 1, \ldots, D(m)$ **do**
5: Set
$$M_\ell = \{(w, v) : \text{some edge between } w \text{ and } v \text{ belongs to } E_\ell\}.$$
6: **end for**
7: **return** partial matchings $M_1, \ldots, M_{D(m)}$.

Algorithm 2 implements the scheduling subroutine used by Algorithm 1. The following lemma shows that any integer source demand can be realized by feasible evaluation matchings without increasing the total number of evaluations.

Lemma 9 (König-based scheduling)**.** *Let $\mathbf{m} \in \mathbb{Z}_{\geq 0}^{N \times K}$ be an integer demand matrix and set*

$$D(\mathbf{m}) := \max\left\{ \max_{w \in \mathcal{U}} \sum_{v \in \mathcal{V}} m_{w,v},\ \max_{v \in \mathcal{V}} \sum_{w \in \mathcal{U}} m_{w,v} \right\}.$$

There exist $D(\mathbf{m})$ partial matchings such that each source pair (w, v) occurs in exactly $m_{w,v}$ of them. Executing these matchings uses exactly $\sum_{w,v} m_{w,v}$ individual evaluations.

Proof. Construct a bipartite multigraph with vertex classes $\mathcal{U}$ and $\mathcal{V}$ and $m_{w,v}$ parallel edges between w and v. Its maximum degree is $D(\mathbf{m})$. König's edge-coloring theorem for bipartite multigraphs gives a proper edge coloring with $D(\mathbf{m})$ colors. Each color class is a partial matching, and every edge belongs to exactly one class. Thus the resulting schedule realizes every demanded evaluation exactly once. For zero demand, the empty schedule suffices. □

We fix the scheduling tie-breaking in advance. In particular, initialization costs d evaluations, and rounding an epoch's real demand n^m upward adds at most d evaluations:

$$\sum_{w,v} \lceil n_{w,v}^m \rceil \leq \sum_{w,v} n_{w,v}^m + d. \tag{79}$$

The scheduling constraint therefore causes no additional evaluation cost beyond the integer demand. Matching rounds and individual evaluations remain distinct quantities.

E.2 Existence of the allocation optimizers

Lemma 10 (Well-defined allocation problems)**.** *At every reached epoch, Algorithm 1's allocation problem is feasible and attains its minimum. Its optimizer can be chosen as a deterministic function of the history. Furthermore, the minimum defining $H_2(\mu)$ is finite, positive, and attained at a vector $x^\star \in (0,\infty)^d$.*

Proof. Initialization ensures that $N_{w,v}^{m-1} \geq 1$. For every comparison, $n \mapsto V_c(N^{m-1}+n)$ is continuous, convex, and coordinatewise nonincreasing on $n \geq 0$. All targets ρ_c^m are positive. Sending every coordinate of n to infinity makes all variances tend to zero, so the feasible set is nonempty and closed. Intersecting it with an objective sublevel set determined by any feasible point gives a nonempty compact set. The continuous linear objective attains its minimum there. The inputs to this problem are the integer counts, the epoch, the active set, and the recorded elimination epochs. They range over a countable state space; we fix one optimizer for each state in advance. This makes the selected demand a measurable function of the past. Coordinatewise upward rounding preserves every constraint.

For $H_2(\mu)$, the uniform vector $x_{w,v} = 2/\Delta_{\min}^2$ is feasible because

$$V_c(x) = \Delta_{\min}^2 \sum_w \Gamma_{w,u}^2 \leq \Delta_{\min}^2 \leq \Delta_c^2.$$

To establish attainment despite the open domain $x > 0$, fix any source (w,v) and any arm $b \neq v$. For the comparison $c = (w, \min\{v,b\}, \max\{v,b\})$, feasibility implies

$$x_{w,v} \geq \frac{\Gamma_{w,w}^2}{\Delta_c^2} > 0.$$

Thus each coordinate is bounded away from zero over the entire feasible set. Its intersection with a finite objective sublevel set is compact, and the constraints are continuous on that set. A minimizer $x^\star$ therefore exists. These positive coordinate lower bounds also give $H_2(\mu) > 0$. □

E.3 Uniform concentration under adaptive evaluation

We first establish a concentration inequality that remains valid when cumulative counts depend on earlier rewards. Conditioning on the final counts and treating historical source means as independent would not justify the allocation constraints in Algorithm 1.

Predictable evaluations. For this probability argument only, index individual evaluations by $r = 1, 2, \ldots$, using a fixed order within each matching. This index is different from the matching-round index t in Section 4. Let e_r be the source evaluated at step r, let $\xi_r = Y_r - \mu_{e_r}$, and let $\mathcal{F}_{r-1}$ include all preceding observations and the randomization used to select e_r. Selection is predictable. The independent, mean-zero, 1-sub-Gaussian noise assumption of Section 4 implies

$$\mathbb{E}_\mu[\exp(\lambda \xi_r) \mid \mathcal{F}_{r-1}] \leq \exp(\lambda^2/2), \qquad \lambda \in \mathbb{R}. \tag{80}$$

This also holds when the primitive noise distributions vary with the calendar round; identical distributions across rounds are not needed. Define

$$T_e(r) := \sum_{i=1}^r \mathbf{1}\{e_i = e\}, \qquad S_e(r) := \sum_{i=1}^r \mathbf{1}\{e_i = e\}\xi_i.$$

At an epoch endpoint, $T_e(r)$ is exactly N_e^m; at the terminal endpoint it is $N_e(\tau)$. Whenever all counts on S_c are positive,

$$\widehat{g}_c(r) - g_c = \sum_{e \in S_c} q_{c,e} \frac{S_e(r)}{T_e(r)}. \tag{81}$$

For a fixed deterministic vector $h \in \mathbb{R}^d$, the process

$$L_r(h) := \exp\left\{ \sum_e h_e S_e(r) - \frac{1}{2} \sum_e h_e^2 T_e(r) \right\} \tag{82}$$

is a nonnegative supermartingale with $L_0(h) = 1$. Indeed, its multiplicative increment is $\exp(h_{e_r}\xi_r - h_{e_r}^2/2)$, whose conditional expectation is at most one by equation 80. Extend the processes constantly after stopping. For every $a > 0$,

$$\Pr_{\mu}\left\{\sup_{r\geq 0} L_r(h) \geq e^a\right\} \leq e^{-a}. \tag{83}$$

For completeness, stop at the first crossing or at a deterministic index R. Optional stopping and nonnegativity give $1 \geq \mathbb{E}_\mu L_{\sigma\wedge R}(h) \geq e^a \Pr_\mu(\sigma \leq R)$; let $R \to \infty$.

Lemma 11 (Count-vector confidence and its tail)**.** *For a positive integer vector $n = (n_e)_{e\in S_c}$, define*

$$b_c(n) := b_0 + \sum_{e\in S_c} \log\big(n_e(n_e+1)\big), \qquad R_c(n) := \sqrt{2b_c(n)V_c(n)}. \tag{84}$$

Here V_c depends only on the coordinates in S_c. Let

$$Z := \sup_{c,r} \frac{|\widehat{g}_c(r) - g_c|}{R_c(T(r))},$$

where the supremum ranges over comparisons and indices for which all counts in S_c are positive. Then, for every real $z \geq 1$,

$$\Pr_{\mu}(Z > z) \leq 2JA_0^{-z^2}. \tag{85}$$

In particular, $Z < \infty$ almost surely and $\Pr_\mu(Z \leq 1) \geq 1 - \delta/32$.

Proof. Fix c, a deterministic positive count vector n on S_c, and $u_0 > 0$. Put $v_0 = V_c(n) > 0$ and choose $h_e = (u_0/v_0)q_{c,e}/n_e$ on S_c, and $h_e = 0$ elsewhere. At any index r with $T_{S_c}(r) = n$ and $\widehat{g}_c(r) - g_c \geq u_0$, the exponent in equation 82 is at least

$$\frac{u_0}{v_0}\,u_0 - \frac{1}{2}\left(\frac{u_0}{v_0}\right)^2 v_0 = \frac{u_0^2}{2v_0}.$$

Applying equation 83 to h and $-h$ yields

$$\Pr_{\mu}\{\exists r : T_{S_c}(r) = n,\ |\widehat{g}_c(r) - g_c| > u_0\} \leq 2\exp\left(-\frac{u_0^2}{2V_c(n)}\right). \tag{86}$$

This bounds all visits to a fixed vector without conditioning on the event that the algorithm selects that vector.

Substitute $u_0 = zR_c(n)$ and take a union bound over all comparisons and all positive integer support-count vectors. Since $z \geq 1$,

$$\begin{aligned}\Pr_{\mu}(Z > z) &\leq 2e^{-b_0 z^2} \sum_{c\in\mathcal{C}} \sum_{n\in\mathbb{N}_{\geq 1}^{S_c}} \prod_{e\in S_c} [n_e(n_e+1)]^{-z^2} \\ &\leq 2e^{-b_0 z^2} \sum_{c\in\mathcal{C}} \prod_{e\in S_c} \left(\sum_{k=1}^{\infty} \frac{1}{k(k+1)}\right) = 2JA_0^{-z^2}.\end{aligned}$$

The scalar series equals one by telescoping. Taking $z = 1$ gives $2J/A_0 = \delta/32$, and letting $z \to \infty$ gives almost-sure finiteness. □

E.4 Deterministic count and confidence-budget bounds

Lemma 12 (Budget domination at every epoch)**.** *At every reached epoch $m \geq 1$, on every sample path,*

$$C_m := \sum_{w,v} N_{w,v}^m \leq L_m := 8d\beta_m 4^m, \qquad b_c(N^m) \leq \beta_m \quad (c \in \mathcal{C}). \tag{87}$$

Consequently, on $\{Z \leq z\}$, for every reached epoch and every comparison,

$$|\widehat{g}_c^m - g_c| \leq z\sqrt{2\beta_m V_c(N^m)} \leq z\rho_c^m. \tag{88}$$

Here C_m is the cost through epoch m and is distinct from C_τ, which uses the matching-round stopping time.

Proof. For every comparison, $\rho_c^m \geq \epsilon_m$: an eliminated comparison has its precision frozen at an earlier, larger value. Consider the uniform additional demand $n_{w,v} = 4\beta_m 4^m$. It satisfies

$$V_c(N^{m-1} + n) \leq \frac{2\sum_w \Gamma_{w,u}^2}{4\beta_m 4^m} \leq \frac{\epsilon_m^2}{2\beta_m} \leq \frac{(\rho_c^m)^2}{2\beta_m}.$$

Optimality and equation 79 bound the cost of epoch m by $4d\beta_m 4^m + d$. Including initialization, and using $\beta_j \leq \beta_m$, $m + 1 \leq 4^m$, and $\sum_{j=1}^m 4^j \leq (4/3)4^m$, we obtain

$$\begin{aligned} C_m &\leq d(m+1) + 4d\sum_{j=1}^m \beta_j 4^j \\ &\leq d4^m + \frac{16}{3} d\beta_m 4^m \leq 8d\beta_m 4^m = L_m, \end{aligned}$$

where $\beta_m \geq 1$. This count bound uses neither a confidence event nor the unknown preference gaps.

It remains to check the constants in equation 77. Since $d \geq 2$, $s \geq 2$, $J \geq 1$, and $\delta < 1/2$, we have $A_0 > 128$ and $\ell_0 \geq 4$. In particular,

$$B_0 \leq 16(b_0 + s)\ell_0, \qquad \log(16dB_0) \leq \log(4e^{\ell_0}\ell_0) \leq 2\ell_0.$$

The last inequality follows from $4\ell_0 \leq e^{\ell_0}$ for $\ell_0 \geq 4$. Since $L_m \geq 1$ and $m + 1 \leq 2^m$,

$$\begin{aligned} \log(L_m + 1) &\leq \log(16dB_0) + \log(m+1) + m\log 4 \\ &\leq 2\ell_0 + m\log 8. \end{aligned}$$

Every support count is at most $C_m \leq L_m$, and $|S_c| \leq s$. Hence

$$\begin{aligned} b_c(N^m) &\leq b_0 + 2s\log(L_m + 1) \\ &\leq b_0 + 4s\ell_0 + 2sm\log 8 \leq B_0(m+1) = \beta_m, \end{aligned}$$

because $B_0 \geq b_0 + 4s\ell_0$ and $B_0 \geq 2s\log 8$. Lemma 11 now gives the first inequality in equation 88. The second follows from the allocation constraint and the fact that rounding increases counts. □

E.5 Persistence of Eliminated Comparisons

Lemma 13 (Correct current rankings after elimination)**.** *On $\{Z \leq 1\}$, if comparison c is eliminated at epoch j_c, then its cumulative empirical gap has the correct sign at epoch j_c and at every later reached epoch. If Algorithm 1 terminates on this event, its prescribed final output satisfies $\widehat{M}_\tau = M^\star$.*

Proof. At the elimination epoch $j = j_c$, the comparison is active and $\rho_c^j = \epsilon_j$. The elimination rule and equation 88 imply

$$\Delta_c \geq |\widehat{g}_c^j| - |\widehat{g}_c^j - g_c| > 2\epsilon_j - \epsilon_j = \epsilon_j.$$

At that epoch the error is therefore strictly smaller than $|g_c|$, so the sign is correct. At every later epoch m, the allocation still includes c with $\rho_c^m = \epsilon_j$. Thus

$$|\widehat{g}_c^m - g_c| \leq \epsilon_j < \Delta_c, \qquad \operatorname{sign}(\widehat{g}_c^m) = \operatorname{sign}(g_c).$$

At termination every comparison has been eliminated. All pairwise rankings among actual arms induced by the current cumulative matrix $\widehat{\Theta}_\tau$ consequently agree with those induced by Θ. Since $\theta_{u,v} > 0 = \theta_{u,\emptyset}$, placing the known outside option last also reproduces the true comparison with being unmatched. Together with the unchanged known arm preferences, these are exactly the preference lists defining $M^\star$. Player-proposing Gale–Shapley therefore returns $M^\star$, including when some players are unmatched. □

The maintained constraints are essential to this conclusion: the final output uses the current cumulative means, not a separate list of stored comparison signs. The count-vector radius R_c need not decrease when an individual source count increases, so permanent sign correctness requires the epochwise control just proved.

E.6 Termination and evaluation cost on bounded-error paths

For each integer $z \geq 1$, define the deterministic quantities

$$M(z) := \left\lceil \log_2 \frac{2(z+2)}{\Delta_{\min}} \right\rceil, \qquad F(z) := d\big(M(z)+1\big) + 4B_0(z+2)^2 H_2(\mu) M(z)\big(M(z)+3\big). \tag{89}$$

In the notation of Theorem 2, $M(1) = M_\Delta$ and $F(1) = d(M_\Delta+1) + 36B_0H_2(\mu)M_\Delta(M_\Delta+3) = F_Q$.

Lemma 14 (Pathwise stopping and cost)**.** *On $\{Z \leq z\}$, Algorithm 1 stops by epoch $M(z)$ and*

$$C_\tau \leq F(z) \leq z^4 F_Q. \tag{90}$$

Proof. If $c \in \mathcal{A}_{m+1}$ survives epoch m, then it was active at that epoch. The survival rule and equation 88 give

$$\Delta_c \leq |\widehat{g}_c^m| + |\widehat{g}_c^m - g_c| \leq (z+2)\epsilon_m. \tag{91}$$

If epoch $M(z)$ is reached, its right-hand side is at most $\Delta_{\min}/2$, contradicting $\Delta_c \geq \Delta_{\min}$. Thus no comparison can survive that epoch, and the algorithm stops by $M(z)$.

We next show that every precision target at every reached epoch satisfies

$$\rho_c^m \geq \frac{\Delta_c}{2(z+2)}. \tag{92}$$

If c is active at epoch $m \geq 2$, it survived epoch $m-1$, so equation 91 gives $\Delta_c \leq (z+2)\epsilon_{m-1} = 2(z+2)\epsilon_m$. If it was eliminated at epoch $j_c \geq 2$, it survived epoch $j_c - 1$, and the same argument gives $\Delta_c \leq 2(z+2)\epsilon_{j_c}$, its frozen target. The remaining cases have target $\epsilon_1 = 1/2$; then equation 92 follows from $\Delta_c \leq 1$ and $z \geq 1$. This argument does not assume that the eliminated sign was correct.

Let $x^\star$ attain $H_2(\mu)$, as in Lemma 10. At epoch m, consider the comparison demand $n = 8\beta_m(z+2)^2 x^\star$. By coordinatewise monotonicity and the homogeneity $V_c(tx) = V_c(x)/t$,

$$V_c(N^{m-1}+n) \leq \frac{V_c(x^\star)}{8\beta_m(z+2)^2} \leq \frac{\Delta_c^2}{8\beta_m(z+2)^2} \leq \frac{(\rho_c^m)^2}{2\beta_m}.$$

It is therefore feasible for the actual allocation problem, including the constraints for previously eliminated comparisons. Optimality and rounding imply that epoch m costs at most $8\beta_m(z+2)^2 H_2(\mu) + d$. Including the d initialization samples and summing through epoch $M(z)$ yields

$$\begin{aligned} C_\tau &\leq d\big(M(z)+1\big) + 8(z+2)^2 H_2(\mu) \sum_{m=1}^{M(z)} \beta_m \\ &= d\big(M(z)+1\big) + 4B_0(z+2)^2 H_2(\mu) M(z)\big(M(z)+3\big) = F(z). \end{aligned}$$

Here the comparison allocation is only a proof device: the algorithm does not need $x^\star$ or the unknown gaps.

Finally, for integer $z \geq 1$, $z+2 \leq 3z$ and $\log_2 z \leq z-1$ give

$$M(z) \leq M_\Delta + z - 1 \leq zM_\Delta, \qquad M(z)+3 \leq z(M_\Delta+3).$$

Consequently the second term in $F(z)$ is at most $36B_0z^4H_2(\mu)M_\Delta(M_\Delta+3)$, while its first term is at most $z\,d(M_\Delta+1) \leq z^4 d(M_\Delta+1)$. This proves $F(z) \leq z^4 F_Q$. □

E.7 Proof of Theorem 2

Proof of Theorem 2. By Lemma 11, $Z < \infty$ almost surely. Every such path belongs to $\{Z \leq z\}$ for some finite positive integer z, so Lemma 14 gives a finite stopping epoch. Every reached epoch

has finite integer demand and hence a finite schedule by Lemmas 10 and 9. It follows that the matching-round stopping time τ is finite almost surely.

On $\{Z \le 1\}$, Lemma 14 gives $C_\tau \le F_Q$, and Lemma 13 gives $\widehat{M}_\tau = M^\star$. Hence we obtain the slightly stronger bound

$$\Pr_\mu\{\widehat{M}_\tau = M^\star,\ C_\tau \le F_Q\} \ge 1 - \frac{\delta}{32} \ge 1 - \delta. \tag{93}$$

Together with almost-sure termination, this proves δ-correctness in the sense of Section 4 and the high-probability assertion of Theorem 2.

For the expectation, it is necessary to include paths outside $\{Z \le 1\}$. Partition the almost-surely finite values of Z into $\{Z \le 1\}$ and $\{j < Z \le j+1\}$, $j = 1, 2, \ldots$. Using nonnegativity, Lemma 14, and equation 85, we obtain

$$\begin{aligned}\mathbb{E}_\mu[C_\tau] &\le F_Q \left(1 + \sum_{j=1}^{\infty} (j+1)^4 \Pr_\mu(Z > j)\right) \\ &\le F_Q \left(1 + 2J \sum_{j=1}^{\infty} (j+1)^4 A_0^{-j^2}\right).\end{aligned}$$

For integer $j \ge 1$, $j + 1 \le 2^j$ and $j^2 \ge j$. Since $A_0 > 128$,

$$\begin{aligned}2J \sum_{j=1}^{\infty} (j+1)^4 A_0^{-j^2} &\le 2J \sum_{j=1}^{\infty} \left(\frac{16}{A_0}\right)^j \\ &= \frac{32J/A_0}{1 - 16/A_0} \le \frac{\delta/2}{7/8} = \frac{4\delta}{7}.\end{aligned}$$

Therefore

$$\mathbb{E}_\mu[C_\tau] \le \left(1 + \frac{4\delta}{7}\right) F_Q < 2F_Q, \tag{94}$$

which implies the stated expectation bound. □

Interpretation of the bound. The explicit choice equation 77 gives

$$F_Q = O\left(dM_\Delta + H_2(\mu) M_\Delta^2 \left[\log \frac{J}{\delta} + s \log\left(d\left(s + \log \frac{J}{\delta}\right)\right)\right]\right).$$

The same bound, up to a universal factor, holds for expected evaluation cost. The inverse-count geometry agrees with the source-level variance structure in Section 5.2. Nevertheless, $H_2(\mu)$ resolves every comparison in $\mathcal{C}$, and this proof also incurs an explicit adaptive confidence factor and epochwise allocation overhead. The result establishes the guarantee in Theorem 2, without identifying the matching-specific characteristic time from the lower bound.

E.8 Extension to Two-Sided Unknown Preferences

Our main framework extends naturally to markets where both sides' preferences are formed through social influence and learned from feedback. In this setting, the arm side also undergoes Friedkin–Johnsen deliberation governed by a known, column-stochastic operator Γ^V and unknown intrinsic utilities $T \in (0,1]^{K \times N}$ with $T_{v,u} = t_v(u)$, inducing social utilities $\Psi = (\Gamma^V)^\top T$ where $\psi_{v,u} = \sum_{a \in V} \Gamma^V_{a,v} t_a(u)$.

Paired Feedback and Joint Variance Constraints. Evaluating a matched pair (u, v) simultaneously yields noisy intrinsic observations for both parties:

$$Y^U(u,v) = \mu_{u,v} + \eta^U(u,v), \qquad Y^V(u,v) = t_v(u) + \eta^V(u,v), \tag{95}$$

where the noise vector $(\eta^U(u,v), \eta^V(u,v))$ is conditionally centered and entrywise 1-sub-Gaussian. Although noise coordinates within the same edge may be dependent, evaluations are independent across edges and rounds.

Both empirical means $\widehat{\mu}_{u,v}$ and $\widehat{t}_v(u)$ share the same cumulative count $N_{u,v}$. Thus, a single sample simultaneously reduces propagated uncertainty for both the player comparison $c = (u, a, b) \in \mathcal{C}_U$ and the arm comparison $c' = (v, u, u') \in \mathcal{C}_V$, whose variance proxies are respectively

$$V_c^U(x) = \sum_{w \in U} (\Gamma_{w,u}^U)^2 \left(\frac{1}{x_{w,a}} + \frac{1}{x_{w,b}} \right), \qquad V_{c'}^V(x) = \sum_{a \in V} (\Gamma_{a,v}^V)^2 \left(\frac{1}{x_{u,a}} + \frac{1}{x_{u',a}} \right). \tag{96}$$

Let $\mathcal{C}_{2s} = \mathcal{C}_U \sqcup \mathcal{C}_V$ collect all $J = N\binom{K}{2} + K\binom{N}{2}$ pairwise comparisons, and let $\Delta_{\min} = \min_{c \in \mathcal{C}_{2s}} \Delta_c > 0$. The joint allocation complexity is defined by

$$H_2^{2s}(\mu, T) = \min_{x > 0} \left\{ \sum_{u \in U, v \in V} x_{u,v} : V_c(x) \le \Delta_c^2, \ \forall c \in \mathcal{C}_{2s} \right\}. \tag{97}$$

Define $H_2^U(\mu)$ and $H_2^V(T)$ as the infima of $\sum_{u,v} x_{u,v}$ over $x > 0$ subject to the respective one-sided comparison constraints; an empty comparison family has value zero. Dropping either family of constraints gives $\max\{H_2^U(\mu), H_2^V(T)\} \le H_2^{2s}(\mu, T)$. Conversely, for any $\varepsilon > 0$, choose feasible allocations $x^U, x^V > 0$ with costs at most $H_2^U(\mu) + \varepsilon/2$ and $H_2^V(T) + \varepsilon/2$, respectively. Since every variance proxy is coordinatewise nonincreasing, $x^U + x^V$ satisfies both families of constraints. Hence $H_2^{2s}(\mu, T) \le H_2^U(\mu) + H_2^V(T) + \varepsilon$. Letting $\varepsilon \downarrow 0$ proves the upper bound.

Joint SAE-Q Guarantee. Joint SAE-Q executes the epoch structure of Algorithm 1, replacing the player-side comparisons $\mathcal{C}$ with the joint set $\mathcal{C}_{2s}$ and updating threshold constants with the updated support size $s \le 2\max\{N, K\}$ and comparison count J.

Proposition 8 (Two-Sided Sample Complexity Guarantee)**.** *Assume* $\max\{N, K\} \ge 2$. *Under strict social preferences on both sides, Joint SAE-Q terminates almost surely and identifies the player-optimal socially stable matching* $M_{2s}^\star$ *with probability at least* $1 - \delta$. *Its total evaluation cost* $C_\tau = \sum_{u,v} N_{u,v}(\tau)$ *satisfies*

$$\Pr\left\{ \widehat{M}_\tau = M_{2s}^\star, \ C_\tau \le F_{2s} \right\} \ge 1 - \delta, \qquad \mathbb{E}[C_\tau] \le 2F_{2s}, \tag{98}$$

where $F_{2s} = d(M_\Delta + 1) + 36B_0 H_2^{2s}(\mu, T) M_\Delta (M_\Delta + 3)$, *with* $M_\Delta = \lceil \log_2(6/\Delta_{\min}) \rceil$ *and* B_0 *defined as in Appendix E with the updated* J *and* s.

Proof Sketch. The proof mirrors Theorem 2:

1. **Decoupled Noise Concentration:** Although $\eta^U(u, v)$ and $\eta^V(u, v)$ may be correlated, no comparison $c \in \mathcal{C}_{2s}$ mixes observations from both sides. Consequently, the exponential supermartingale argument (Appendix E.3) holds independently for each coordinate. A single union bound over all J comparisons establishes that the uniform concentration bound $|\widehat{g}_c^m - g_c| \le z\rho_c^m$ holds on $\{Z \le z\}$ with $\Pr(Z > z) \le 2JA_0^{-z^2}$.

2. **Epoch Termination and Cost:** Feasibility and attainment follow identically from the positive diagonals $\Gamma_{u,u}^U > 0$ and $\Gamma_{v,v}^V > 0$. On the good event $\{Z \le 1\}$, every surviving comparison satisfies $\Delta_c \le 3\epsilon_m$, ensuring termination by epoch M_Δ. Scaling the joint optimizer $x^\star$ of equation 97 by $8\beta_m(z + 2)^2$ yields a valid epochwise upper bound, which integrates against the tail bound of Z to guarantee $\mathbb{E}[C_\tau] \le 2F_{2s}$.

□

F MATCHING-CRITICAL EXPLORATION AND RANKING OVERHEAD

F.1 OVERVIEW AND NOTATION

Overview and motivation. SAE-Q accounts for the way source observations propagate through the social network, but its stopping certificate resolves every player's complete arm ranking. Matching identification can require much less information: a nearly tied comparison may have no effect on the final assignment.
This appendix asks whether network-aware allocation can be combined with a smaller, matching-specific certificate. Under a common known arm priority, we answer this question with SAE-Q-SD. The algorithm discovers the serial-allocation certificate, preserves it under subsequent sampling, and returns Gale–Shapley computed from the current cumulative empirical utilities. Its expected-cost guarantee depends on matching-relevant gaps, rather than the smallest gap in a complete ranking.

The algorithm in one pass. Process players in priority order. For the current player, eliminate arms until one remains; record the comparisons that justified these eliminations. Compress the resulting tree into comparisons between the winner and each loser, and maintain their precision while processing later players. A doubling evaluation budget restarts provisional decisions when necessary, while retaining all observations. Algorithm 3 makes these steps explicit; Figure 3 explains the compression.

Reading guide. The main argument is *separation* (Section F.2), *certificate* (Section F.3), *algorithm* (Section F.4), and *guarantee and return to the example* (Section F.5). Section F.7 gives the full proof. Sections F.8 and F.9 explain information alignment and the remaining limitations.

Scope. We retain the original model, known social operator Γ, intrinsic mean domain $(0,1]^{N\times K}$, and strict social-preference domain $\mathcal{D}_\Gamma$. Write $\Theta = \Gamma^\top \mu$. All actual partners are acceptable, with the outside option last. The algorithmic results hold for independent, mean-zero, 1-sub-Gaussian observation noise; information lower bounds concern the unit-variance Gaussian subclass. Throughout, $K \geq 2$ and $0 < \delta < 1/2$. Common arm priority is required for the certificate and algorithm and holds throughout their correctness class, including all admissible alternatives.

Four complexities, two normalizations. Let $\mathcal{C} = \{(u,a,b) : u \in [N],\ a < b\}$. For an oriented comparison $c = (u,a,b)$, define

$$g_c = \theta_{u,a} - \theta_{u,b}, \qquad \Delta_c = |g_c|,$$
$$V_c(x) = \sum_{w:\Gamma_{w,u}>0} \Gamma_{w,u}^2 \left(\frac{1}{x_{w,a}} + \frac{1}{x_{w,b}}\right),$$
$$D_x(\mu,\lambda) = \frac{1}{2}\sum_{w,v} x_{w,v}(\mu_{w,v} - \lambda_{w,v})^2.$$

The variance proxy V_c is unchanged by reversing the comparison. Positive coefficients divided by zero are $+\infty$; zero coefficients are omitted. Write $V_{u,a,a} = 0$ and $\Delta_{\min} = \min_{c\in\mathcal{C}} \Delta_c$. The ranking-changing set is

$$\mathcal{A}_{\text{rank}}(\mu) = \{\lambda \in \mathcal{D}_\Gamma : g_c(\lambda) g_c(\mu) < 0 \text{ for some } c \in \mathcal{C}\}.$$

Each quantity in Table 3 minimizes $\sum_{w,v} x_{w,v}$. For H_2 we retain the positive-allocation convention of the main text; the other programs allow zero on unused coordinates. The family $\mathcal{C}_{\text{SD}}$ is defined in Section F.3.

In particular, with $k_\delta = \text{kl}(1-\delta,\delta)$, the Gaussian information lower-bound value is $T^*_{\text{opt}}(\mu,\delta) = k_\delta H_{\text{opt}}(\mu)$. The ratio H_2/H_{opt} compares allocation complexities only. It does not include the confidence and epoch factors in SAE-Q's Theorem 2 bound $2F_Q$, whose ratio to this lower bound is $2F_Q/(k_\delta H_{\text{opt}})$.

Proposition 9 (Universal comparison)**.** *For every strict instance,*

$$H_{\text{opt}} \leq H_{\text{rank}} \leq 2H_2.$$

Table 3: Complexity dictionary. The factor 2 converts a variance/gap constraint into a unit-information constraint when the unconstrained Gaussian reversal is feasible in the mean box.

Quantity	Constraints defining the allocation	Normalization
H_{opt}	$D_x \geq 1$ for all matching-changing alternatives $\lambda \in \mathrm{Alt}_{\mathrm{opt}}(\mu)$	Unit information
H_{rank}	$D_x \geq 1$ for all $\lambda \in \mathcal{A}_{\mathrm{rank}}(\mu)$	Unit information
H_2	$V_c(x) \leq \Delta_c^2$ for all $c \in \mathcal{C}$	Variance/gap
H_{SD}	$V_c(x) \leq \Delta_c^2$ for all $c \in \mathcal{C}_{\mathrm{SD}}$	Variance/gap

For $\Gamma = I$, $H_{\mathrm{rank}} = 2H_2$. This equality is not asserted for an arbitrary social operator on the bounded mean domain.

Proof. Preserving all player rankings preserves the target matching when arm priorities are fixed, so $\mathrm{Alt}_{\mathrm{opt}} \subseteq \mathcal{A}_{\mathrm{rank}}$. For any reversal of c, weighted Cauchy–Schwarz gives $\Delta_c^2 \leq 2D_x(\mu,\lambda)V_c(x)$. If z is feasible for H_2, then $V_c(2z) \leq \Delta_c^2/2$; thus $2z$ supplies unit information against every ranking reversal. For $\Gamma = I$, the least costly tie sets the two means to their allocation-weighted average, which remains in their original range. The box therefore does not change the infimum reversal cost, and the information constraints are exactly $V_c(x) \leq \Delta_c^2/2$. A zero count on a compared coordinate gives zero infimum reversal cost. Rescaling proves the identity. □

The distinction is consequently about *which comparisons are required*, as well as how their precision is normalized. The next example isolates the first issue on a fixed connected network.

F.2 Why complete rankings can be arbitrarily wasteful

Proposition 10 (Unbounded separation with two players and two arms)**.** *Let $N = K = 2$, let both arms rank player 1 above player 2, and take*

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad A = \frac{2}{3} I_2, \qquad \Gamma = A[I - P(I-A)]^{-1} = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}.$$

For $0 < \varepsilon \leq 1/8$, set

$$\mu^{(\varepsilon)} = \begin{pmatrix} 11/16 - \varepsilon/4 & 5/16 + \varepsilon/4 \\ 7/16 + 3\varepsilon/4 & 9/16 - 3\varepsilon/4 \end{pmatrix}.$$

Then all intrinsic means belong to $[5/16, 11/16]$, and

$$\Gamma^\top \mu^{(\varepsilon)} = \begin{pmatrix} 5/8 & 3/8 \\ 1/2 + \varepsilon/2 & 1/2 - \varepsilon/2 \end{pmatrix}.$$

Moreover,

$$H_{\mathrm{opt}}(\mu^{(\varepsilon)}) = 128, \qquad H_2(\mu^{(\varepsilon)}) = \frac{4}{\varepsilon^2}, \qquad \frac{H_2}{H_{\mathrm{opt}}} = \frac{1}{32\varepsilon^2}. \tag{99}$$

In particular, no finite upper bound on H_2/H_{opt} depending only on N and K exists, even on this fixed connected social network.

Proof. Both arms prefer player 1, so any stable matching must give player 1 her preferred arm; otherwise she blocks with that arm. Player 2 receives the remaining arm. Thus the unique stable matching, and hence the player-optimal one, depends only on player 1's social comparison. At the displayed instance it is $\{(1,a),(2,b)\}$, and

$$\mathrm{Alt}_{\mathrm{opt}}(\mu^{(\varepsilon)}) = \{\lambda \in \mathcal{D}_\Gamma : \theta_{1,a}(\lambda) < \theta_{1,b}(\lambda)\}.$$

For the upper bound on H_{opt}, use

$$x^{\mathrm{opt}} = \begin{pmatrix} 48 & 48 \\ 16 & 16 \end{pmatrix}, \qquad V_1(x^{\mathrm{opt}}) = \frac{1}{32} = \frac{(1/4)^2}{2}.$$

Weighted Cauchy–Schwarz implies $D_{x^{\mathrm{opt}}}(\mu, \lambda) \geq 1$ for every matching-changing alternative, so $H_{\mathrm{opt}} \leq 128$. Conversely, for $t > 0$ sufficiently small, set

$$\lambda^{(t)}_{w,a} = \mu_{w,a} - 1/8 - t, \qquad \lambda^{(t)}_{w,b} = \mu_{w,b} + 1/8 + t \quad (w = 1, 2).$$

These alternatives lie strictly inside the mean box. Their social gaps are $-2t$ and $\varepsilon - 1/4 - 2t$, so both are strict and the target changes. For any feasible x, letting $t \downarrow 0$ gives

$$1 \leq \lim_{t \downarrow 0} D_x(\mu, \lambda^{(t)}) = \frac{1}{128} \sum_{w,v} x_{w,v}.$$

Thus $H_{\mathrm{opt}} \geq 128$.

Player 2's gap is ε, so weighted Cauchy–Schwarz gives $(\sum_{w,v} x_{w,v}) V_2(x) \geq 4$, hence $H_2 \geq 4/\varepsilon^2$. This bound is attained by

$$x^{\mathrm{rank}} = \frac{1}{\varepsilon^2} \begin{pmatrix} 1/2 & 1/2 \\ 3/2 & 3/2 \end{pmatrix}.$$

Indeed,

$$V_2(x^{\mathrm{rank}}) = \varepsilon^2, \qquad V_1(x^{\mathrm{rank}}) = \frac{7}{3}\varepsilon^2 \leq \frac{7}{192} < \frac{1}{16}.$$

There is only one arm comparison per player. Hence this allocation is feasible and has total cost $4/\varepsilon^2$. □

The separation is not caused by an ill-conditioned social operator or by means approaching the boundary: Γ is fixed with eigenvalues 1 and $1/2$, and all means remain in a fixed interior box. Its cause is the need to resolve player 2's irrelevant comparison.

The separation is intrinsic to recovering all rankings. On the same family,

$$H_{\mathrm{rank}}(\mu^{(\varepsilon)}) = \frac{8}{\varepsilon^2} = 2H_2(\mu^{(\varepsilon)}). \tag{100}$$

For the lower bound, shift every source mean for a downward and every source mean for b upward by $\varepsilon/2 + t$. As $t \downarrow 0$, these interior alternatives reverse player 2's ranking while preserving player 1's ranking and the matching. Their information cost approaches $(\varepsilon^2/8) \sum_{w,v} x_{w,v}$. The upper bound follows from $H_{\mathrm{rank}} \leq 2H_2$. Thus a procedure that recovers the complete social rankings with confidence $1 - \delta$ must pay at least $(8/\varepsilon^2)\,\mathrm{kl}(1 - \delta, \delta)$ evaluations in expectation. This applies to the ranking certificate produced by SAE-Q, using its simultaneous sign-correctness guarantee.

Only player 1's choice determines the matching; player 2 receives the remaining arm. The challenge is to exploit this kind of irrelevance without being given the true matching or its decisive comparisons. Section F.5 returns to this same family after presenting the general common-priority algorithm.

F.3 The comparisons that determine the matching

Assume that every arm has the same known strict priority

$$u_1 \succ_v u_2 \succ_v \cdots \succ_v u_N, \qquad v \in V.$$

This is a restriction on arm priorities, not on the social operator: Γ may be dense and non-diagonal. It holds for every instance in the correctness class. Under this restriction, serial allocation identifies the unique stable matching, so each player only needs to identify its best *remaining* arm.

The exact matching certificate. Let $L = \min\{N, K\}$, set $R_1 = V$, and recursively define

$$a_r = \arg\max_{v \in R_r} \theta_{u_r,v}, \qquad R_{r+1} = R_r \setminus \{a_r\}, \quad 1 \leq r \leq L.$$

The players after u_L, if any, are unmatched. Define the oriented family

$$\mathcal{C}_{\mathrm{SD}}(\mu) = \{(u_r, a_r, b) : 1 \leq r \leq L,\ b \in R_r \setminus \{a_r\}\}, \qquad \Delta_{r,b} = \theta_{u_r,a_r} - \theta_{u_r,b} > 0.$$

Its size is $LK - L(L+1)/2$, at most NK, compared with $N\binom{K}{2}$ full-ranking comparisons. A comparison involving an arm already assigned to a higher-priority player is absent, as is any comparison between two remaining losing arms.

Proposition 11 (Exact certificate and alternatives)**.** *Under the common-priority assumption the serial allocation above is the unique stable matching. Moreover, writing $M = M_{\mathrm{opt}}(\mu)$, for every $\lambda \in \mathcal{D}_\Gamma$,*

$$M_{\mathrm{opt}}(\lambda) = M \quad \Longleftrightarrow \quad \theta_{u_r,a_r}(\lambda) > \theta_{u_r,b}(\lambda) \quad \textit{for every } (u_r, a_r, b) \in \mathcal{C}_{\mathrm{SD}}(\mu).$$

Consequently,

$$\mathrm{Alt}_{\mathrm{opt}}(\mu) = \bigcup_{(u_r,a_r,b)\in\mathcal{C}_{\mathrm{SD}}(\mu)} \{\lambda \in \mathcal{D}_\Gamma : \theta_{u_r,b}(\lambda) > \theta_{u_r,a_r}(\lambda)\}. \tag{101}$$

In particular, $\mathcal{C}_{\mathrm{SD}}$ is exactly the blocking-critical family, with its assigned arm made explicit; there are no additional alternatives that preserve the stability of M while changing player-optimality.

Proof. In every stable matching u_1 must receive its favorite arm: otherwise it blocks with that arm, which prefers u_1 to every possible assignee or to being unmatched. Remove this player and arm and repeat. This proves uniqueness and the serial rule, including unbalanced markets. If all displayed inequalities hold under λ, induction gives the same choices $a_1, \dots, a_L$. Conversely, if one is reversed, arm b is either unassigned in M or assigned to a lower-priority player. Then (u_r, b) blocks M under λ, so the target changes. Strict preferences exclude equality in the admissible domain. □

Matching-specific allocation complexity. For $K \geq 2$, define

$$H_{\mathrm{SD}}(\mu) = \min_{x\geq 0} \left\{ \sum_{w,v} x_{w,v} : V_{u_r,a_r,b}(x) \leq \Delta_{r,b}^2,\ (u_r, a_r, b) \in \mathcal{C}_{\mathrm{SD}}(\mu) \right\}, \tag{102}$$

where positive coefficients divided by zero are $+\infty$ and zero coefficients are omitted. Coordinates unused by the certificate may be zero. This minimum is finite and attained: the feasible set is closed under the extended-value convention, it contains a sufficiently large uniform vector, and its objective sublevel sets are compact. It is positive when $K \geq 2$. Clearly $H_{\mathrm{SD}} \leq H_2$. Set

$$\Delta_{\mathrm{SD}} = \min_{(u_r,a_r,b)\in\mathcal{C}_{\mathrm{SD}}} \Delta_{r,b} > 0.$$

Unlike the minimum full-ranking gap, this quantity ignores irrelevant ties between losing arms. Strict preferences are nevertheless retained as a standing assumption here; an extension of the domain to ties is not needed.

The alternative decomposition and Proposition 9's Cauchy–Schwarz argument imply

$$H_{\mathrm{opt}} \leq 2H_{\mathrm{SD}} \leq 2H_2. \tag{103}$$

Equality $H_{\mathrm{opt}} = 2H_{\mathrm{SD}}$ requires additional analysis of the bounded domain; Section F.8 supplies it. The certificate above is an instance-dependent object used to describe complexity. The algorithm must discover its arms and comparisons from data.

F.4 SAE-Q-SD: DISCOVERY, COMPRESSION, AND PERSISTENCE

Three time scales serve different purposes: a *trial* sets a total evaluation budget; a *stage* assigns one player; and a local *epoch* reduces the precision target for that stage's surviving candidates. Local epoch numbers restart at each stage. Trial restarts discard provisional decisions but keep every observation, including observations from aborted trials.

Counts, budgets, and precision. Write $d = NK$, $J = N\binom{K}{2}$, and $L = \min\{N, K\}$. To distinguish comparison coefficients from the trial index q, use

$$a_{c,(w,v)} = \Gamma_{w,u}(\mathbf{1}\{v = a\} - \mathbf{1}\{v = b\}), \quad c = (u, a, b),$$
$$S_c = \{e : a_{c,e} \neq 0\}, \qquad s = \max_{c\in\mathcal{C}} |S_c| \leq 2N.$$

Thus $V_c(x) = \sum_{e\in S_c} a_{c,e}^2/x_e$. We write $\mathbf{N}$ for the cumulative count matrix, distinguishing it from the number of players N. For trial $q \geq 1$, set

$$B_q = 2^q, \qquad A_0 = 64J/\delta, \qquad \beta_q = \log A_0 + 2s\log(d + 2^{q+1} + 1). \tag{104}$$

Algorithm 3 SAE-Q-SD: adaptive matching-critical exploration

Require: Known Γ, common priority $u_1 \succ \cdots \succ u_N$, confidence δ, evaluation access, and SCHEDULE.
Ensure: A matching computed from current cumulative empirical social utilities.

if $K = 1$ **then**
  Return the known priority allocation without observations.
end if
Evaluate each source pair once; initialize $\mathbf{N}$ and all cumulative means.
for $q = 1, 2, \ldots$ **do**
  Set B_q, β_q by equation 104; $b \leftarrow B_q$; $\mathcal{P} \leftarrow \varnothing$; $R \leftarrow V$; `abort` $\leftarrow$ false.
  for $r = 1, \ldots, L$ **do**
    $R_r^{\mathrm{trial}} \leftarrow R$; $S \leftarrow R$; $\mathcal{T} \leftarrow \varnothing$; $m \leftarrow 1$.
    while $|S| > 1$ **do**
      $\varepsilon_m \leftarrow 2^{-m}$; choose a deterministic pivot $p \in S$.
      $\mathcal{Q} \leftarrow \mathcal{P} \cup \mathcal{T} \cup \{((u_r, p, v), \varepsilon_m) : v \in S \setminus \{p\}\}$.
      Solve equation 105; $k_e \leftarrow \lceil n_e \rceil$ for every source pair e.
      if $\sum_e k_e > b$ **then**
        `abort` $\leftarrow$ true; break the **while** loop without executing this demand.
      end if
      Execute k with SCHEDULE; update $\mathbf{N}$ and cumulative means; $b \leftarrow b - \sum_e k_e$.
      Recompute $\widehat{\Theta} = \Gamma^\top \widehat{\mu}$; choose $\ell \in \arg\max_{v \in S} \widehat{\theta}_{u_r, v}$ with fixed tie-breaking.
      $D \leftarrow \{v \in S \setminus \{\ell\} : \widehat{\theta}_{u_r,\ell} - \widehat{\theta}_{u_r,v} > 4q\varepsilon_m\}$.
      $\mathcal{T} \leftarrow \mathcal{T} \cup \{((u_r, \ell, v), 2\varepsilon_m) : v \in D\}$; interpret these as edges $\ell \to v$.
      $S \leftarrow S \setminus D$ (all deletions are simultaneous).
      $m \leftarrow m + 1$.
    end while
    if `abort` **then**
      Break the stage **for** loop.
    end if
    Let a be the unique member of S; provisionally assign a to u_r.
    For $v \in R_r^{\mathrm{trial}} \setminus \{a\}$, set $\rho_{r,v} \leftarrow \sum_{f \in \mathrm{path}_{\mathcal{T}}(a,v)} \rho_f$.
    $\mathcal{P} \leftarrow \mathcal{P} \cup \{((u_r, a, v), \rho_{r,v}) : v \in R_r^{\mathrm{trial}} \setminus \{a\}\}$; discard $\mathcal{T}$.
    $R \leftarrow R \setminus \{a\}$.
  end for
  if not `abort` **then**
    Return player-proposing Gale–Shapley on current $\widehat{\Theta}$, with fixed strict tie-breaking and the outside option last.
  end if
  Keep $\mathbf{N}$ and all observations; restart provisional decisions at trial $q + 1$.
end for

The budget B_q is measured in individual evaluations. The dimensionless quantity β_q determines confidence radii. For a comparison of target radius ρ_c, the allocation call is

$$\min_{n \geq 0} \sum_e n_e \quad \text{subject to} \quad V_c(\mathbf{N} + n) \leq \frac{\rho_c^2}{2\beta_q}, \qquad (c, \rho_c) \in \mathcal{Q}. \tag{105}$$

Fix a deterministic optimizer selection and round each demand upward. The same SCHEDULE oracle as in SAE-Q implements the integer demands by edge-coloring their bipartite multigraph. Its individual-evaluation cost is exactly the sum of these demands.

What is stored, and why? In Algorithm 3, $\mathcal{P}$ stores the compressed stars of completed stages, while $\mathcal{T}$ stores the current stage's elimination forest. Both store pairs (c, ρ_c), with frozen target radii. The active pivot star only compares surviving candidates at the current precision. If two candidates are each estimated relative to a common pivot with radius ε_m, their pairwise radius is at most $2\varepsilon_m$. This is why a deleted arm receives radius $2\varepsilon_m$, although the allocation call used radius ε_m.

When only one candidate a remains, the forest has become a tree rooted at a. Its path to each eliminated arm b records a chain of certified dominance comparisons. Compression replaces this chain by (u_r, a, b) with

$$\rho_{r,b} = \sum_{f \in \mathrm{path}(a,b)} \rho_f. \tag{106}$$

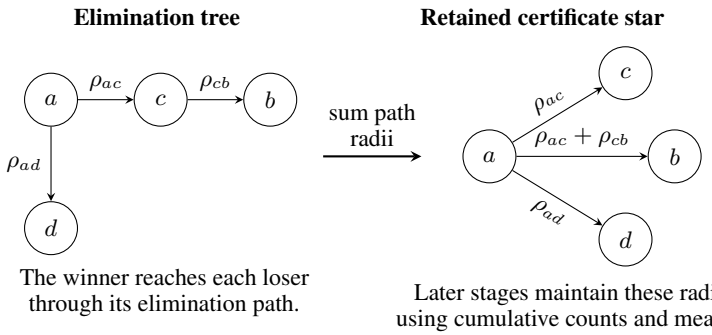


Figure 3: Star-tree compression within a single player stage. Arrows point from a defeating arm to an eliminated arm; a is the eventual winner. For the path $a \to c \to b$, the stored radius becomes $\rho_{ab} = \rho_{ac} + \rho_{cb}$. The additive quantities are radii (standard-deviation bounds), not variances. Compression changes the comparisons retained for future sampling; it does not claim to reduce the number of edges of the completed tree.

If a stage starts with a singleton, there are no comparisons or paths to store and no sampling is performed. After L stages, later players, if any, are unmatched.

The geometry behind compression. At the current cumulative counts, $\sqrt{V_c(\mathbf{N})}$ is the Euclidean norm of the coefficient vector $(a_{c,e}/\sqrt{N_e})_e$. Coefficient vectors telescope along a tree path, so the triangle inequality gives

$$\sqrt{V_{u_r,a,b}(\mathbf{N})} \le \sum_{f \in \mathrm{path}(a,b)} \sqrt{V_f(\mathbf{N})} \le \frac{\rho_{r,b}}{\sqrt{2\beta_q}}. \tag{107}$$

Thus the new star constraints already hold at the compression endpoint. On the confidence event used in the proof, true gaps along these edges also add and satisfy $\Delta_{r,b} > q\rho_{r,b}$. Maintaining the star therefore keeps the *current* empirical winner above every loser throughout subsequent allocation endpoints. Storing an old sign alone would not establish this property for the final cumulative estimator.

Size of the allocation program. If the current stage began with t arms and now has k survivors, its forest has $t - k$ edges and its active star has $k - 1$ edges. Hence this stage contributes $t - 1$ constraints, irrespective of its elimination history. Adding the completed-stage stars gives

$$|\mathcal{Q}| \le \sum_{r=1}^{L} (K - r) = LK - \frac{L(L+1)}{2} \le NK.$$

Exploratory comparisons between contenders are permitted: the true certificate is unknown. They need not all be resolved, and loser–loser comparisons are not retained after a stage is compressed.

F.5 SAE-Q-SD Guarantee and the Separation Example

The next result separates quantities the algorithm uses from quantities that only describe its performance. In particular, it never receives H_{SD}, Δ_{SD}, or the true certificate as input.

Table 4: Trial notation: operational quantities versus proof quantities.

Symbol	Meaning	Used by algorithm?
$B_q = 2^q$	Additional evaluations allowed in trial q	Yes
β_q	Confidence multiplier converting radii into variance constraints	Yes
M_q	Upper bound on local epochs per stage when the trial is accurate	No
F_q	Upper bound on all rounded evaluation demands in such a trial	No
q_0	First $q \ge 8$ whose budget covers F_q	No

Theorem 6 (Matching-critical SAE-Q under common priority)**.** *Assume $K \geq 2$ and the model above. For integers $q \geq 1$, define*

$$M_q = \left\lceil \log_2 \frac{12q}{\Delta_{\mathrm{SD}}} \right\rceil, \tag{108}$$

$$F_q = LM_q \left[d + 576q^2 \beta_q H_{\mathrm{SD}}(\mu) \right]. \tag{109}$$

Let q_0 be the smallest integer $q \geq 8$ such that $2^q \geq F_q$. Then $q_0 < \infty$, the algorithm stops almost surely, and

$$\Pr_{\mu}\{\widehat{M}_\tau = M_{\mathrm{opt}}(\mu),\ C_\tau \leq d + 2^{q_0+1}\} \geq 1 - \delta/32, \qquad \mathbb{E}_\mu C_\tau \leq d + 2^{q_0+2}.$$

The quantities q_0, F_q, M_q are for analysis only. In particular, the instance dependence involves H_{SD} and Δ_{SD}, not gaps between losing arms. The stage factor L and support-dependent confidence term remain explicit in F_q; this is not a characteristic-time-optimality statement. For $K = 1$, the matching is known without observations.

What the guarantee removes. The relevant gap Δ_{SD} only compares each assigned arm to its competitors at its serial stage. Neither a comparison between losing arms nor a comparison to an arm already assigned to a higher-priority player enters this minimum. The following corollary makes the distinction concrete.

Corollary 2 (The same instance, with and without full-ranking certification)**.** *On the family in Proposition 10,*

$$\mathcal{C}_{\mathrm{SD}} = \{(1, a, b)\}, \qquad \Delta_{\mathrm{SD}} = \frac{1}{4}, \qquad H_{\mathrm{SD}} = 64,$$

and

$$H_{\mathrm{opt}} = 2H_{\mathrm{SD}} = 128, \qquad H_{\mathrm{rank}} = 2H_2 = \frac{8}{\varepsilon^2}. \tag{110}$$

For fixed δ, Theorem 6's expected-cost bound is uniform over $0 < \varepsilon \leq 1/8$. In contrast, every uniformly correct full-ranking identification procedure requires at least $(8/\varepsilon^2)\,\mathrm{kl}(1-\delta, \delta)$ evaluations in expectation on this family.

Proof. The second serial stage has a single available arm, so only player 1's comparison belongs to the certificate. Cauchy–Schwarz gives $H_{\mathrm{SD}} \geq 4/(1/4)^2 = 64$, attained by

$$x^* = \begin{pmatrix} 24 & 24 \\ 8 & 8 \end{pmatrix}, \qquad V_1(x^*) = 1/16.$$

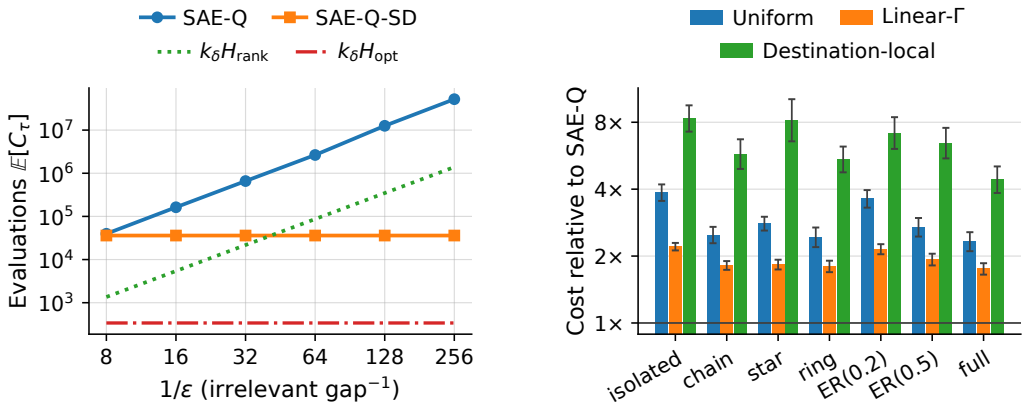
The other identities follow from the direct box-feasible alternatives in Proposition 10 and equation 100. Every parameter in F_q and q_0 is independent of ε: $d = 4$, $J = 2$, $L = 2$, $s = 4$, $H_{\mathrm{SD}} = 64$, and $\Delta_{\mathrm{SD}} = 1/4$. The last claim is the Gaussian full-ranking change-of-measure lower bound. □

This closes the main argument: the same fixed connected market admits a full-ranking information cost diverging as ε^{-2} and a matching-specific algorithmic bound that remains uniformly bounded. The statement concerns the upper bound; it does not assert that the algorithm's exact stopping-time distribution is independent of ε.

F.6 Numerical Illustrations

In this section, we provide numerical experiments to corroborate the theoretical insights established in the preceding analysis. Rather than serving as an exhaustive benchmarking suite, our empirical inquiry is designed to validate two foundational predictions:

1. the decoupling of matching-critical exploration from irrelevant preference gaps (Theorem 6 vs. Theorem 2), and
2. the empirical efficiency gained by network-aware quadratic sample allocation over heuristic and destination-local exploration rules across diverse interaction topologies (Section 6).



(a) Separation under irrelevant gaps (Proposition 10).

(b) Cost of allocation rules relative to SAE-Q across topologies.

Figure 4: Numerical illustrations of the theoretical mechanisms. **(a)** Empirical sample complexity on the two-player separation family (Proposition 10) across varying gap scales $\varepsilon \in [1/256, 1/8]$ (100 independent trials per point; log-log scale). Circles denote SAE-Q (full ranking), while squares denote SAE-Q-SD (matching-critical). Reference lines display the theoretical information barriers: $k_\delta H_{\text{rank}} = 8k_\delta/\varepsilon^2$ (dotted green) and $k_\delta H_{\text{opt}} = 128k_\delta$ (dash-dotted red), where $k_\delta = \text{kl}(1-\delta, \delta)$. **(b)** Sample complexity of alternative allocation strategies relative to SAE-Q across seven network topologies ($N = K = 5$, 420 random instances spanning $\alpha \in \{0.2, 0.5, 0.8\}$, 5 trials per instance). Bars depict the geometric mean of relative evaluation counts; error bars correspond to 95% confidence intervals under a log-normal approximation.

Experimental Environment and General Setup. Throughout all experiments, evaluation feedback is subject to standard independent Gaussian noise, $Y_t(w, v) \sim \mathcal{N}(\mu_{w,v}, 1)$, with the confidence parameter fixed at $\delta = 0.05$. For practical implementation, we deploy a standard non-asymptotic confidence radius for SAE-Q, replacing the conservative theoretical threshold $B_0(m+1)$ with $\beta_m = \log(2Jm(m+1)/\delta)$, where $J = N\binom{K}{2}$ represents the total number of unordered comparisons. For SAE-Q-SD, we instantiate a single-trial variant equipped with star compression and the matching-specific allocation program equation 105, omitting the budget-doubling schedule and q-dependent scaling factors while maintaining empirical correctness. In all trials, the algorithms terminated with an empirical success rate of 100%, correctly identifying the target matching $M^\star$.

Experiment 1: Separation of Matching Identification from Irrelevant Gaps. We first examine the two-player, two-arm parametric family established in Proposition 10, where arms possess a common priority over players. In this market, player 1's favorite arm dictates the unique stable assignment, inducing a constant matching-critical gap $\Delta_{1,a,b} = 1/4$, whereas player 2's preference gap $\Delta_{2,a,b} = \varepsilon$ is completely uninformative to the target matching. We sweep $\varepsilon \in \{2^{-3}, 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}, 2^{-8}\}$ and evaluate 100 independent realizations for each parameter setting.

As shown in Figure 4a, the empirical sample complexity of SAE-Q scales inversely with ε^2, rising sharply from 4.0×10^4 to 5.2×10^7 evaluations. This trajectory strictly mirrors the asymptotic rate dictated by the full-ranking lower bound, $k_\delta H_{\text{rank}} \propto \varepsilon^{-2}$. Conversely, SAE-Q-SD concentrates sampling purely on the matching-critical comparison $(1, a, b)$ via tree-to-star compression, maintaining a flat sample cost of approximately 3.6×10^4 evaluations throughout the entire parameter spectrum. This striking empirical contrast corroborates Corollary 2: global preference certification can incur unbounded sample inflation in the presence of uninformative near-ties, whereas adaptive matching-critical exploration successfully decouples sample complexity from irrelevant social comparisons.

Experiment 2: Network Topologies and Allocation Rule Ablations. To evaluate the efficacy of the quadratic allocation scheme in Algorithm 1, we conduct an ablation study across diverse

social graph structures. We fix a balanced market with $N = K = 5$. Synthetic market instances are generated by drawing intrinsic mean matrices $\mu \sim \mathrm{Unif}[0.1, 0.9]^{5\times 5}$ alongside independent, strict arm-side preferences. To avoid pathological stalling caused by exact ties, we filter instances by requiring all intrinsic and social preference gaps to satisfy $\min_c \Delta_c \geq 0.02 \cdot 50/J$. The social influence matrix is set to column-normalized adjacency $P = WD^{-1}$ with $D = \mathrm{diag}(W^\top \mathbf{1})$, so that $\sum_w P_{w,u} = 1$ ($P_{w,u}$: influence of source w on destination u); a node without neighbours keeps a self-loop, $P_{u,u} = 1$, and the anchoring matrix is set to $A = \alpha I$ with $\alpha \in \{0.2, 0.5, 0.8\}$.

We evaluate across seven canonical network topologies, each embodying distinct structural sparsity and propagation radii:

- **Isolated**: Empty graph ($P = I$), reducing the model to classical uncoupled multi-agent bandit feedback ($\Gamma = I$).
- **Chain**: Undirected 1D line graph (path), inducing slow, diameter-dependent variance propagation.
- **Star**: A single central hub connected to four periphery nodes, creating strong asymmetric influence concentration.
- **Ring**: Symmetric 1D cycle where each agent interacts with two immediate peers, providing homogeneous, non-hierarchical influence.
- **Erdős–Rényi (ER) Graphs**: Random graphs $G(N, p)$ with $p = 0.2$ and $p = 0.5$, drawn independently for each instance (not conditioned on connectivity; isolated nodes keep a self-loop), representing heterogeneous random social clusters.
- **Fully Connected**: Uniform complete graph ($P = (\mathbf{1}\mathbf{1}^\top - I)/(N-1)$), where social influence is maximally diffused across all participants.

For each topology and anchoring level α, we generate 20 independent market instances (420 instances in total), running 5 random noise trials per configuration. We compare SAE-Q against three allocation baselines that adhere to the exact same elimination and stopping criteria, differing only in the directional source-demand formulation in Step 5:

- UNIFORM: Samples all NK source-arm pairs at an identical rate ($n_{w,v} \propto 1$), rescaled by the minimal scalar required to certify active comparisons.
- LINEAR-Γ: Weights source pairs proportionally to their aggregate first-order influence on active comparisons, setting $n_{w,v} \propto \sum_{c\in\mathcal{A}_m : v\in\{a,b\}} \Gamma_{w,u}/\rho_c$.
- DESTINATION-LOCAL: Samples exclusively from the destination agent's direct pairs (u, a) and (u, b) by assigning $n_{w,v} \propto \sum_{c=(u,a,b):v\in\{a,b\}} 1/\rho_c$ when $w = u$, supplemented by a 1% sampling floor on external source pairs to guarantee eventual termination.

Ablation Results and Discussion. Figure 4b summarizes the evaluation costs normalized against SAE-Q. Across all 420 instances, SAE-Q strictly outperforms the alternatives, achieving substantial sample reductions across every topology. Averaged over all markets:

- UNIFORM incurs $2.8\times$ the sample complexity of SAE-Q. Even under the Isolated network ($\Gamma = I$), UNIFORM requires nearly $4\times$ more samples, demonstrating that tailoring sample budgets to heterogeneous gap scales remains critical even without network coupling.
- LINEAR-Γ requires $1.9\times$ more evaluations on average. While it correctly restricts exploration to the network support, its linear weighting fails to capture the quadratic curvature of the propagated variance $V_c(n) = \sum_w \Gamma^2_{w,u}(n^{-1}_{w,a} + n^{-1}_{w,b})$, underscoring the exactness of the quadratic program.
- DESTINATION-LOCAL performs the worst, requiring on average $6.4\times$ (and up to $8\times$ in hub-and-spoke and isolated structures) more samples. Because destination-local sampling fails to allocate observations to influential peers $\Gamma_{w,u} > 0$, it is forced to rely on the minimal sampling floor to resolve propagated uncertainty.

These empirical findings directly substantiate the core premise of our work: when preference formation is mediated by social networks, learning efficiency fundamentally hinges on source-destination alignment through quadratic information allocation.

F.7 PROOF OF THE MAIN GUARANTEE

The proof has two distinct jobs. The elimination rule must discover the correct serial choices, and the retained radii must keep those choices correct for the cumulative empirical utilities used at termination. The final step controls rare inaccurate trials to obtain an expectation bound.

Proof of Theorem 6. We prove four claims: a time-uniform confidence bound, persistence of the serial certificate, a per-trial cost bound, and termination with finite expectation.

Step 1: confidence under adaptive cumulative sampling. For comparison c with coefficient vector a_c and support S_c, put

$$b_c(n) = \log A_0 + \sum_{e \in S_c} \log(n_e(n_e+1)), \qquad R_c(n) = \sqrt{2b_c(n)V_c(n)}.$$

Let $S_e(t)$ be the sum of centered observations from source pair e. For fixed h, predictable evaluations and conditional sub-Gaussianity make

$$\exp\left\{\sum_e h_e S_e(t) - \frac{1}{2}\sum_e h_e^2 N_e(t)\right\}$$

a nonnegative supermartingale. Fix c, a positive count vector n on S_c, and $z \geq 1$. Apply Ville's inequality with

$$h_e = \pm z\sqrt{\frac{2b_c(n)}{V_c(n)}}\frac{a_{c,e}}{n_e} \quad (e \in S_c), \qquad h_e = 0 \quad (e \notin S_c).$$

At a time with $\mathbf{N}_{S_c}(t) = n$, this yields

$$\Pr\left\{\exists t : \mathbf{N}_{S_c}(t) = n,\ |\widehat{g}_c(t) - g_c| > z\sqrt{2b_c(n)V_c(n)}\right\} \leq 2e^{-z^2 b_c(n)}.$$

Since $\sum_{n=1}^{\infty}[n(n+1)]^{-z^2} \leq 1$, summing over count vectors and the J unoriented comparisons gives

$$Z := \sup_{c,t}\frac{|\widehat{g}_c(t) - g_c|}{R_c(\mathbf{N}(t))}, \qquad \Pr(Z > z) \leq 2JA_0^{-z^2}, \quad z \geq 1. \tag{111}$$

The supremum uses positive support counts, as ensured by initialization. In particular $Z < \infty$ almost surely and $\Pr(Z \leq 1) \geq 1 - \delta/32$. There is a statistical union over all comparisons; the allocation program need not store or enforce all those comparisons.

Every accepted endpoint in trial q has total count at most $d + \sum_{j \leq q} 2^j < d + 2^{q+1}$. Consequently $b_c(\mathbf{N}) \leq \beta_q$. At an allocated comparison of target ρ, therefore,

$$|\widehat{g}_c - g_c| \leq Z\rho. \tag{112}$$

Upward rounding preserves the variance constraints. Existence and measurability of an allocation optimizer follow by the compact sublevel-set argument used for SAE-Q; the relevant combinatorial and radius states are countable.

Every trial finishes or aborts in finitely many steps, even when $Z > q$. Its accepted endpoints have fewer than $d + 2^{q+1}$ cumulative evaluations. For every nontrivial comparison, Cauchy–Schwarz gives $V_c(\mathbf{N}) \geq 4/\sum_e N_e$. If at least two candidates survived infinitely many local epochs, the required variance $\varepsilon_m^2/(2\beta_q)$ would tend to zero while this lower bound remained positive for every affordable endpoint. Eventually the requested demand would exceed the remaining budget. There are only finitely many stages and candidate deletions.

Step 2: discovery and persistence when $Z \leq q$. Consider any such trial, including one preceded by arbitrarily erroneous trials. At its first stage the available set is the true R_1. Within a stage having the true available set, the active pivot star gives

$$|(\widehat{\theta}_{u_r,a} - \widehat{\theta}_{u_r,b}) - (\theta_{u_r,a} - \theta_{u_r,b})| \leq 2q\varepsilon_m, \quad a, b \in S.$$

The true best arm a_r is never deleted. If b survives, the empirical leader dominates the empirical value of a_r, so

$$\Delta_{r,b} \le 6q\varepsilon_m. \tag{113}$$

Every recorded elimination edge has true directed gap strictly larger than $2q\varepsilon_m = q\rho_e$. Since later allocations maintain target ρ_e, its cumulative empirical sign remains correct by equation 112. At the instant the edge is recorded its variance target is also satisfied: the standard-deviation triangle inequality through the active pivot gives

$$\sqrt{V_{u_r,\ell,b}(\mathbf{N})} \le \sqrt{V_{u_r,\ell,p}(\mathbf{N})} + \sqrt{V_{u_r,p,b}(\mathbf{N})} \le \frac{2\varepsilon_m}{\sqrt{2\beta_q}}.$$

Use $V_{u,v,v} = 0$ for coinciding endpoints.

The tree root at stage completion is a_r. Along its path to b, true gaps add, and each edge gap exceeds $q\rho_e$. Thus

$$\Delta_{r,b} > q\rho_{r,b}.$$

The same standard-deviation triangle inequality shows that compressed star constraints already hold at the compression endpoint. Maintaining these constraints later preserves every empirical inequality $\widehat{\theta}_{u_r,a_r} > \widehat{\theta}_{u_r,b}$. Induction gives the true available sets and choices at all later stages. At termination the current empirical serial choices therefore equal the true choices. Common arm priority makes current empirical Gale–Shapley produce exactly those choices, even if irrelevant empirical comparisons tie.

Step 3: cost of an accurate trial. At the beginning of a reached epoch m, every active arm v satisfies

$$\Delta_{r,v} \le 12q\varepsilon_m, \qquad \Delta_{r,a_r} = 0.$$

For $m \ge 2$ this follows from equation 113; for $m = 1$ use $\Delta_{r,v} \le 1$ and $q \ge 1$. Equation equation 113 implies completion by epoch M_q whenever the trial is not budget-aborted.

Let x^* attain H_{SD}. For any two arms v, w from the same true available set,

$$V_{u_r,v,w}(x^*) \le V_{u_r,a_r,v}(x^*) + V_{u_r,a_r,w}(x^*) \le \Delta_{r,v}^2 + \Delta_{r,w}^2. \tag{114}$$

The first inequality follows directly by writing each arm's contribution as $\sum_z \Gamma_{z,u_r}^2/x_{z,v}$; it also holds when an endpoint is a_r. An active comparison of target ε_m consequently has variance at most $288q^2\varepsilon_m^2$ under x^*.

A retained elimination edge created at epoch j has both endpoint gaps at most $12q\varepsilon_j$, and radius $2\varepsilon_j$, so the same bound is more than sufficient for it. For a completed-stage star comparison, consider the epoch j when b was deleted. Its eventual root-to-b path contains that deletion edge, hence

$$\rho_{r,b} \ge 2\varepsilon_j \ge \Delta_{r,b}/(6q).$$

Its variance under x^* is therefore at most $36q^2\rho_{r,b}^2$. For all types of retained or active constraint, the additional demand

$$n = 576q^2\beta_q x^*$$

is feasible in equation 105, by monotonicity and $V_c(tx) = V_c(x)/t$. This is a proof comparator, not an input to the algorithm. Optimality and rounding bound each epoch's cost by $d + 576q^2\beta_q H_{\mathrm{SD}}$. There are at most LM_q epochs. If $2^q \ge F_q$, the trial cannot abort: the remaining bound covers every remaining epoch, including its rounded demand. It finishes correctly.

Step 4: the doubling budget and the expectation bound. The function F_q grows only polynomially in q, up to logarithms, so q_0 exists. For $q \ge 8$, $M_q \ge 7$ and $M_{q+1} \le M_q + 1$, while

$$\frac{\beta_{q+1}}{\beta_q} \le \frac{q+2}{q+1}.$$

The latter follows from increasing $\log(d+2^{q+1}+1)$ by at most $\log 2$, and its lower bound $(q+1)\log 2$. Therefore

$$\frac{F_{q+1}}{F_q} \le \frac{8}{7}\left(\frac{9}{8}\right)^2\frac{10}{9} < 2.$$

It follows that $2^q \geq F_q$ for every $q \geq q_0$. On every path with finite Z, a reached trial $q \geq \max\{q_0, \lceil Z \rceil\}$ must finish, proving almost-sure termination. On $Z \leq 1$, every completed trial is correct and termination occurs by q_0, for total cost less than $d + 2^{q_0+1}$.

For $q \geq q_0$, reaching trial $q+1$ implies $Z > q$. Thus

$$\begin{aligned}\mathbb{E}C_\tau &\leq d + \sum_{j=1}^{q_0} 2^j + \sum_{q=q_0}^{\infty} 2^{q+1} \Pr(Z > q) \\ &\leq d + 2^{q_0+1} + \sum_{q=q_0}^{\infty} 2^{q+1}\, 2JA_0^{-q^2} \leq d + 2^{q_0+2}.\end{aligned}$$

For the last bound one may use $A_0 > 128$, $q^2 \geq q$, and the resulting geometric series. If $q_0 > 8$, minimality also gives $2^{q_0} < 2F_{q_0}$ and hence $\mathbb{E}C_\tau < d + 8F_{q_0}$. The case $q_0 = 8$ adds only a universal constant. The stage and support factors remain visible in this expansion. □

F.8 When the certificate matches the information lower bound

The algorithmic conclusion above does not require an equality between H_{opt} and $2H_{\mathrm{SD}}$. This section addresses that separate question. The bounded intrinsic mean domain matters: an unconstrained Gaussian projection need not remain in $[0,1]^{N\times K}$.

The exact box-constrained benchmark. For $c = (u_r, a_r, b) \in \mathcal{C}_{\mathrm{SD}}$, define

$$B_c(x;\mu) = \inf_{\substack{\lambda\in[0,1]^{N\times K} \\ \theta_{u_r,a_r}(\lambda)-\theta_{u_r,b}(\lambda)\leq 0}} \frac{1}{2}\sum_{w,v} x_{w,v}(\mu_{w,v} - \lambda_{w,v})^2.$$

Here B_c is an information projection value, distinct from the trial budget B_q. The exact alternative decomposition equation 101 gives

$$T^*_{\mathrm{opt}}(\mu,\delta) = \inf_{x\geq 0}\left\{\sum_{w,v} x_{w,v} : B_c(x;\mu) \geq k_\delta,\ c \in \mathcal{C}_{\mathrm{SD}}\right\}. \tag{115}$$

Passing from strict alternatives to the closed reversal halfspace does not change the infimum. Positive anchoring makes Γ invertible, so strict social profiles are dense. An interior point on the strict reversal side exists by making the alternative arm's intrinsic column uniformly larger than the assigned arm's column. Mixing any point in the closed reversal region with such an interior point, followed by an arbitrarily small perturbation removing unrelated ties, proves the closure claim.

For positive counts on S_c, the unconstrained reversal value is $\Delta_c^2/(2V_c(x))$. It equals $B_c(x;\mu)$ if its minimizing perturbation stays in the box. If this holds for every certificate comparison at the finite expected count vector of a correct Gaussian algorithm, the lower-bound constraints imply

$$\mathbb{E}_\mu C_\tau \geq 2k_\delta H_{\mathrm{SD}}(\mu).$$

For $\Gamma = I$, the two compared means meet at their allocation-weighted average, so the projection always stays in the box. Zero counts on a certificate support have zero infimum reversal cost. Consequently,

$$H_{\mathrm{opt}} = 2H_{\mathrm{SD}}, \qquad T^*_{\mathrm{opt}} = 2k_\delta H_{\mathrm{SD}} \quad (\Gamma = I).$$

For a general operator, equation 115 remains the exact benchmark. The next condition is sufficient, not necessary, for the same identities.

Proposition 12 (Interior reversal condition)**.** *Suppose that for every $c = (u,a,b) \in \mathcal{C}_{\mathrm{SD}}(\mu)$ and every w with $\Gamma_{w,u} > 0$,*

$$\Delta_c \leq \Gamma_{w,u}\min\{\mu_{w,a}, 1-\mu_{w,b}\}. \tag{116}$$

Then for every allocation positive on S_c, $B_c(x;\mu) = \Delta_c^2/(2V_c(x))$. Every allocation feasible for a positive information threshold is positive on all certificate supports. Therefore

$$H_{\mathrm{opt}}(\mu) = 2H_{\mathrm{SD}}(\mu), \qquad T^*_{\mathrm{opt}}(\mu,\delta) = 2k_\delta H_{\mathrm{SD}}(\mu). \tag{117}$$

Proof. Write $g_w = \Gamma_{w,u}$ and $V = V_c(x)$. The unconstrained optimizer lowers $\mu_{w,a}$ by r_w^0 and raises $\mu_{w,b}$ by s_w^0, where

$$r_w^0 = \frac{\Delta_c g_w}{V x_{w,a}}, \qquad s_w^0 = \frac{\Delta_c g_w}{V x_{w,b}}.$$

Since $V \geq g_w^2/x_{w,a}$ and $V \geq g_w^2/x_{w,b}$,

$$r_w^0 \leq \frac{\Delta_c}{g_w} \leq \mu_{w,a}, \qquad s_w^0 \leq \frac{\Delta_c}{g_w} \leq 1 - \mu_{w,b}.$$

Thus the optimizer belongs to the closed box and its value equals B_c. If $x_{w,a} = 0$ on a positive-weight coordinate, lowering only that mean by Δ_c/g_w reaches the tie at zero cost and is feasible by equation 116. The analogous upward shift applies when $x_{w,b} = 0$. The closure argument above gives zero infimum strict-reversal cost, even when the tie itself is not admissible. Positive information requirements therefore exclude such zero counts. The unit-information program becomes exactly $V_c(x) \leq \Delta_c^2/2$ for all certificate comparisons. Rescaling by two proves both identities. □

A simpler sufficient test, including non-diagonal operators. Let

$$\gamma_* = \min_{\substack{(u,a,b)\in\mathcal{C}_{\mathrm{SD}} \\ w:\Gamma_{w,u}>0}} \Gamma_{w,u}, \qquad \Delta_{\mathrm{SD}}^{\max} = \max_{c\in\mathcal{C}_{\mathrm{SD}}} \Delta_c.$$

If the relevant intrinsic means lie in $[\eta, 1-\eta]$ and $\Delta_{\mathrm{SD}}^{\max} \leq \eta\gamma_*$, then equation 116 holds. The *maximum* certificate gap is needed: a bound on Δ_{SD} alone would not control every reversal.

This condition admits instances for any fixed social operator in the model. Choose a fixed matrix $\Xi \in [-1,1]^{N\times K}$ such that $\Gamma^\top \Xi$ has strict row rankings, and set $\mu(t) = \frac{1}{2}\mathbf{1} + t\Xi$. For $t > 0$, all rankings and the certificate are fixed because $\Gamma^\top \mathbf{1} = \mathbf{1}$. For $0 < t \leq \gamma_*/8$, the means lie in $[1/4, 3/4]$ and all social gaps are at most $2t \leq \gamma_*/4$; take $\eta = 1/4$. No limit $\Gamma \to I$ is involved.

The two-arm family of Corollary 2 need not satisfy the stronger sufficient condition equation 116: its information identity follows directly from its box-feasible alternatives. Neither route removes the algorithm's stage, confidence, or epoch overhead.

F.9 Scope, Overhead, and General Priorities

Algorithmic Overhead and Cumulative Persistence. While SAE-Q-SD successfully tightens the instance complexity to H_{SD} and Δ_{SD}, its non-asymptotic bound in Theorem 6 still incurs polynomial overhead: the stagewise factor $L = \min\{N, K\}$, the support size s, and logarithmic epoch terms $M_{q_0} q_0^2$. Whether this stagewise accumulation is fundamentally unavoidable remains open. Crucially, the retained-radius mechanism ensures that signs remain empirically consistent across *all cumulative* means at termination, guaranteeing correctness without discarding past observations via wasteful sample splitting.

Role of Common Priorities vs. General Structures. A pointwise unique stable matching under the true instance does not prevent alternative instances from admitting competing stable matchings (as shown in Example 2). Certifying player-optimality therefore requires either checking exponentially many witness matchings or imposing structural invariants. The common-priority assumption adopts the latter, guaranteeing serial dictator equivalence across all admissible alternatives. For general arm priorities, finding a polynomial-sized certificate that rules out optimality-loss alternatives without exhaustive enumeration remains an open problem; hence, SAE-Q remains our primary algorithm for arbitrary market structures.