FROM PREFERENCE TO RECIPROCITY: DECENTRALIZED MATCHING WITH EMPIRICALLY GROUNDED LLM-AGENT BASED MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Bipartite matching is a fundamental problem in game theory and market design. Classical approaches such as Gale–Shapley assume complete preferences and centralized computation, whereas many real-world matching processes are decentralized, asynchronous, and shaped by sequential interaction under limited information. We propose a dynamic bipartite matching framework that combines large language model (LLM) agents with contextual bandits. In a simulated Chinese marriage market, economically grounded LLM agents evaluate locally encountered candidates, while agent-specific Logistic-UCB models learn reciprocal acceptance from realized proposal outcomes. The mechanism therefore separates two decisions—*whom do I like?* and *who is likely to like me back?*—without requiring ex ante market-wide preference rankings. We first validate LLM-induced mate preferences against the empirical conditional-logit reference across multiple LLM backbones. In the 50×50 matching experiment, Bandit-UCB achieves the highest mean mutual welfare (56.01 versus 54.87 for Gale–Shapley), a smaller gender rank gap than the classical baselines, and the fewest blocking pairs among the LLM-ABM policies. Learned acceptance models show economically interpretable gender-differentiated associations, while counterfactual setups reveal no systematic unilateral advantage from prior search knowledge. Overall, these results support the advantages of decentralized matching with LLM-based behavioral modeling and online learning under incomplete information for economic simulation and computational social science research.

1 INTRODUCTION

Two-sided matching is a fundamental problem in game theory and market design, with applications ranging from labor markets and school choice to marriage markets (Gale & Shapley, 1962; Kamecke, 1992; Roth & Sotomayor, 1992). The classical Gale–Shapley deferred-acceptance algorithm (Gale & Shapley, 1962) guarantees a stable matching by taking agents' ordinal preference rankings as given. While this formulation has been highly successful as a normative allocation mechanism, it abstracts away from several features of naturally occurring matching processes (Eyupoglu et al., 2021; Axtell & Kimbrough, 2008; Zhang & Fang, 2024): agents typically observe only a limited set of alternatives, interactions occur sequentially and asynchronously, and preferences or feasible opportunities may be learned only through experience. Moreover, stability does not necessarily coincide with aggregate welfare. Both analytical and computational studies have shown that restricting attention to stable outcomes can impose a non-negligible welfare cost (Axtell & Kimbrough, 2008; Boudreau & Knoblauch, 2013; Chen et al., 2021; Ortega et al., 2024). These observations motivate a complementary view of matching: rather than asking only how to compute an allocation given a complete preference profile, we study how preference information, beliefs, and matches can jointly evolve through decentralized interaction.

Agent-based modeling (ABM) provides a natural bottom-up framework for such a perspective. By specifying heterogeneous agents, local information, and interaction rules, ABMs generate aggregate outcomes from repeated individual decisions (Epstein, 2012; Axtell & Farmer, 2025). This is particularly relevant to marriage markets, where individual partner choices depend on the current pool of available alternatives, while every new relationship changes that pool for subsequent deci-

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.

sions (Kalmijn, 1998; Chiappori & Weiss, 2006; Van Bavel & Grow, 2016; Grow, 2019). Axtell & Kimbrough (2008), for example, replace centralized deferred acceptance with decentralized search, proposal, and rematching, achieving higher average welfare without preserving guaranteed stability, which is not necessarily a strict desideratum in marriage markets. Yet decentralizing the *matching process* does not by itself solve the problem of *preference formation*: candidate rankings still need to be specified before the matching dynamics begin. This is a central difficulty in agent-based marriage-market modeling. As emphasized by Grow (2019), theories of mate choice are often *competing*, *multidimensional*, and *verbal* rather than directly mathematical, forcing modelers to translate behavioral assumptions into hand-designed utility functions and parameterizations.

Large language models (LLMs) provide a novel way to operationalize semantically rich behavioral assumptions. Existing studies show that LLM agent-based modeling (LLM-ABM) can condition decisions on heterogeneous personas and contextual information and can be embedded in agents that interact over extended trajectories (Argyle et al., 2023; Aher et al., 2023; Park et al., 2023). However, we do not treat an LLM as an unconstrained oracle of human preferences: expressive behavior is not necessarily economically valid behavior, motivating careful specification and validation when LLMs are used as simulated economic actors (Horton et al., 2023; Ludwig et al., 2024). Nor do we provide the agents with market-wide preference rankings in advance, or let them infer the results in one-shot. Instead, LLM valuation is triggered only when agents encounter one another. Yet knowing how much an agent values a candidate does not reveal whether that candidate will reciprocate. Because the proposer observes acceptance only after making a proposal, reciprocity becomes a sequential learning problem under partial feedback. This naturally motivates an agent-specific contextual bandit that learns, from realized interactions, which desirable candidates are likely to accept. Unlike matching-bandit approaches that learn latent rewards or unknown preference rankings (Liu et al., 2021; Dai & Jordan, 2021; Cen & Shah, 2022), our framework separates *behavioral evaluation* from *interaction learning*: the LLM determines how much an agent values an encountered candidate, while the bandit learns how likely that candidate is to accept.

Based on this separation, we develop a decentralized and asynchronous bipartite matching framework that combines behaviorally grounded LLM agents with agent-specific contextual bandits. In the marriage-market setting, agents encounter only locally available alternatives, evaluate them through LLM-based preferences, and learn reciprocal acceptance from realized outcomes, without requiring complete rankings in advance. Experiments on a simulated Chinese marriage market validate the behavioral grounding of the agents and show favorable welfare–stability outcomes relative to classical and decentralized baselines, while learned acceptance models and counterfactual experiments reveal structured reciprocal-choice patterns and the limits of unilateral information advantages.

Our contributions are threefold:

- **Empirically grounded LLM-agent based modeling.** We develop a general approach for constructing and validating LLM agents from domain-specific empirical evidence, and instantiate it in a marriage market using socioeconomic characteristics and gender-specific mate-preference evidence. We validate the induced preferences against an empirical reference across multiple LLM backbones and market scales.
- **Decoupled decentralized matching mechanism.** We formulate matching as an asynchronous local-interaction process that separates LLM-based partner valuation from agent-specific reciprocity learning, without requiring agents to access complete preference rankings. For a theoretically calibrated variant, we establish local proposal-regret bounds under realizable and dynamically misspecified acceptance models.
- **Economic evaluation and interpretable results.** We show that the proposed framework preserves the core two-sided matching problem while relaxing strong information assumptions, and achieves better overall welfare–stability outcomes. The learned reciprocal-acceptance models yield economically interpretable patterns, while warm-start counterfactuals reveal how prior market knowledge translates into individual and market-level matching outcomes.

Our code is available at https://anonymous.4open.science/r/Bi-match-simulation-FEB2/, and related work is reviewed in Appendix A.

2 EMPIRICALLY GROUNDED LLM-AGENT MODELING

2.1 SYNTHETIC POPULATION WITH EMPIRICAL HETEROGENEITY

Our modeling principle is to construct LLM agents from domain-specific empirical evidence rather than unconstrained persona generation. This requires identifying behaviorally relevant observable characteristics and preserving important population heterogeneity and dependence structures. The resulting structured profiles provide both the population representation and the empirical basis for downstream LLM-based behavioral evaluation.

We instantiate this approach in a synthetic Chinese marriage market. Each agent is endowed with attributes emphasized in empirical studies of mate choice (Zhou et al., 2023), including *age, income, education, family background, housing status, and physical appearance*. These dimensions capture the *gender heterogeneity* and *social-class differences* documented in the empirical reference.

Following the dependency-aware construction of Li et al. (2026), we sample demographic variables to approximate observed population heterogeneity while preserving key relationships, such as income–education dependence, rather than treating attributes independently. Demographic and economic variables are based on the 2015 National 1% Population Sample Survey of China (Population Census Office under the State Council & Department of Population and Employment Statistics, National Bureau of Statistics, 2016), consistent with the period studied by Zhou et al. (2023), while physical appearance is sampled from an approximately normal distribution based on empirical evidence (Hamermesh & Biddle, 1993). Full sampling details and realized population composition are reported in Appendix B.

2.2 LLM PERSONAS AND MATE PREFERENCE

Each structured profile is rendered as a concise public persona containing only observable attributes available to potential partners, ensuring that LLM evaluations use the same information encoded in the demographic profile. The rendering template is provided in Appendix C.

To model mate preferences, we construct gender-specific prompts grounded in the regression evidence of Zhou et al. (2023) and use chain-of-thought prompting (Wei et al., 2022) to structure evaluation. The full private preference prompts are given in Appendix C.2.

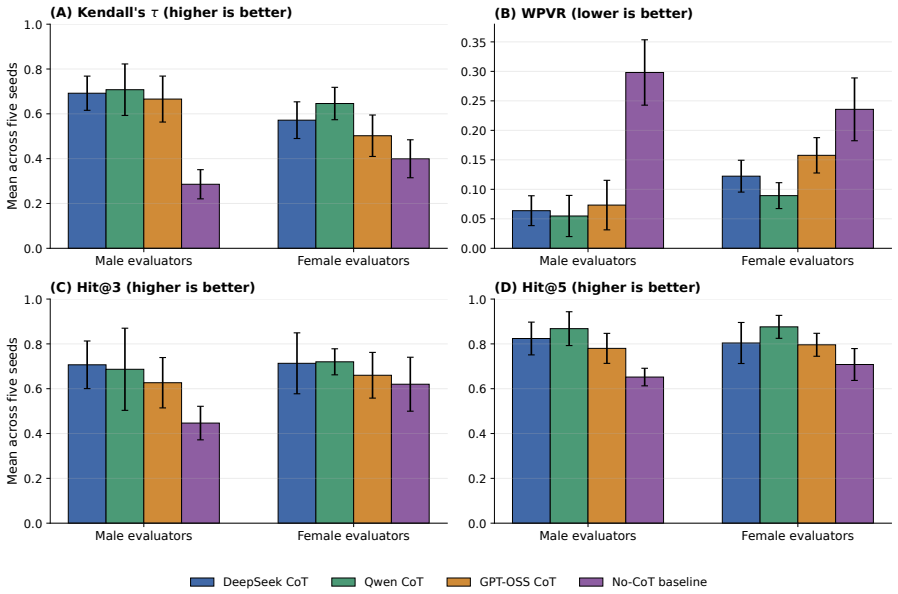


Figure 1: Empirical-reference alignment in the 10×10 validation task. Bars show means across five seeds and error bars show standard deviations.

2.3 Validation of Modeling

As discussed in Appendix A, validating persona-conditioned decisions is important for establishing the economic interpretability of LLM-based social simulations. We therefore compare the mate-preference rankings induced by our LLM agents with an empirical reference model derived from Zhou et al. (2023). The reference model maps each candidate profile to a conditional-logit utility under the corresponding gender-specific specification, while the LLM assigns an overall desirability score from the evaluator's public persona and private mate-preference prompt. We evaluate ranking agreement using Kendall's τ, weighted pairwise violation rate (WPVR), and top-K overlap, with definitions provided in Appendix D.

Figure 1 reports the validation results for a 10×10 market across five random seeds. Across all three LLM backbones, preference-structured prompts achieve stronger alignment with the empirical reference than the no-CoT baseline, indicating that the result is not specific to a single backbone. A 50×50 scale-up validation is reported in Appendix F.

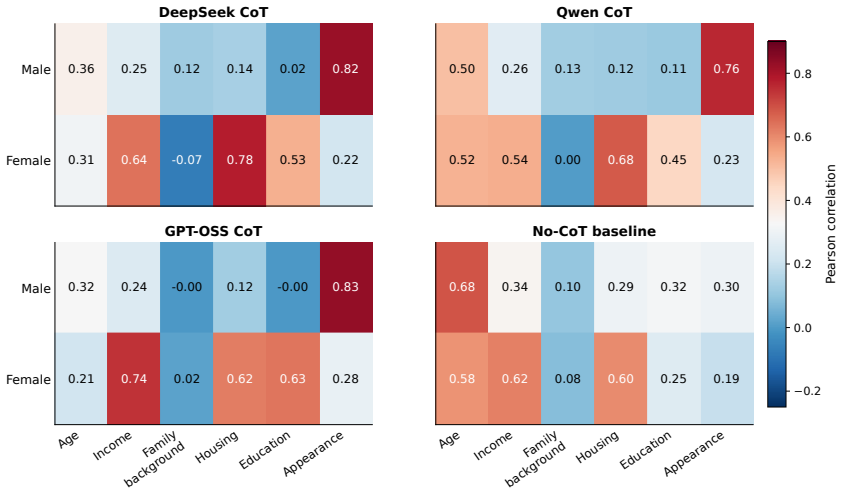


Figure 2: Dimension-consistency heatmap for agents in the 10×10 validation. Entries report Pearson correlations between overall desirability and dimension-level scores, averaged across five seeds.

Figure 2 provides a complementary internal-consistency diagnostic. Appearance is most strongly associated with overall desirability for male agents, whereas female evaluations are more closely associated with socioeconomic dimensions, particularly housing, income, and education. These patterns are consistent with the gender-specific preference structure documented by Zhou et al. (2023), while the no-CoT baseline fails to reproduce the same structure. Results for the 50×50 setting are reported in Appendix F.

3 Decentralized Matching System

3.1 A Decentralized and Asynchronous Matching Market

We consider a two-sided marriage market with male agents $\mathcal{M} = \{m_1, \ldots, m_{N_M}\}$, female agents $\mathcal{W} = \{w_1, \ldots, w_{N_W}\}$, and population $\mathcal{A} = \mathcal{M} \cup \mathcal{W}$. At period t, $\mu_t(i)$ denotes agent i's current partner, with $\mu_t(i) = \varnothing$ if i is single and $\mu_t(i) = j \Leftrightarrow \mu_t(j) = i$. A match is either *temporary* or *permanent*. Temporary relationships remain contestable and can be replaced by later accepted proposals, whereas permanently matched agents exit the active market.

The market evolves through decentralized and asynchronous activation. At the beginning of period t, let $\mathcal{A}_t^{\text{act}} \subseteq \mathcal{A}$ denote agents that have not entered permanent matches. A period consists of $|\mathcal{A}_t^{\text{act}}|$ sequential activation opportunities, each sampling one agent uniformly with replacement from this set. Thus, an agent may be activated multiple times within a period or not at all.

Activation does not give an agent access to the entire opposite side of the market. When agent i is activated, it observes only a locally available candidate set $\mathcal{C}_t(i) \subseteq \mathcal{P}_t(i)$, where $\mathcal{P}_t(i)$ denotes the set of currently feasible agents on the opposite side who have not entered permanent matches and the candidate set $\mathcal{C}_t(i)$ is randomly sampled from this feasible pool. Decisions therefore depend on both preferences and locally encountered opportunities. This differs from Gale–Shapley (Gale & Shapley, 1962), which uses complete rankings in a centralized procedure, and from Axtell & Kimbrough (2008), which decentralizes interaction but retains pre-specified preference lists.

3.2 Whom Do I Like? Local LLM Valuation

The first decision of an activated agent is a valuation problem: *whom do I like?* For evaluator i and locally observed candidate j, the simulated agent in Section 2 produces a normalized subjective utility represented by an overall preference score:

$$U_i(j) \in [0,1], \qquad U_i(j) \neq U_j(i) \text{ in general.} \tag{1}$$

The asymmetry is essential: i's valuation of j does not reveal whether j would reciprocate.

Let r_i denote agent i's reservation utility, the minimum acceptable partner utility. Its current benchmark is

$$b_{i,t} = \begin{cases} r_i, & \mu_t(i) = \varnothing, \\ \max\{r_i, U_i(\mu_t(i))\}, & \mu_t(i) \neq \varnothing. \end{cases} \tag{2}$$

For each $j \in \mathcal{C}_t(i)$, define the utility improvement and eligible set by

$$\Delta U_{ij,t} = U_i(j) - b_{i,t}, \qquad \mathcal{E}_t(i) = \{j \in \mathcal{C}_t(i) : \Delta U_{ij,t} > \epsilon_U, \epsilon_U \geq 0\}. \tag{3}$$

This eligibility condition separates *preference* from *strategic search*: only candidates preferred to the current benchmark are considered, so reciprocal-acceptance learning cannot make an otherwise undesirable candidate attractive. The LLM determines whom the agent wants; the remaining uncertainty is which of these desirable candidates are likely to reciprocate.

3.3 Who Is Likely to Like Me Back? Local Reciprocal-Acceptance Learning

The second decision concerns reciprocity: *who is likely to like me back?* If i proposes to j, the receiver independently evaluates i against its current benchmark,

$$b^R_{j,t} = \begin{cases} r_j, & \mu_t(j) = \varnothing, \\ \max\{r_j, U_j(\mu_t(j))\}, & \mu_t(j) \neq \varnothing. \end{cases} \tag{4}$$

The realized response is

$$Y_{i \to j,t} = \mathbf{1}\left\{U_j(i) > b^R_{j,t}\right\} \in \{0,1\}. \tag{5}$$

Crucially, proposer i does not observe $U_j(i)$, the receiver's incumbent valuation, or its eventual decision before proposing. Acceptance is revealed only after a proposal is made, so feedback is observed only for candidates actually approached. Because proposal opportunities are limited, the proposer must balance exploiting candidates believed likely to reciprocate with exploring uncertain candidates whose responses are informative. This sequential decision problem with contextual, partial feedback naturally motivates an agent-specific contextual bandit.

For each exposed candidate j, proposer i observes a dyadic context $x^{(i)}_{j,t} = \phi(z_i, z_j) \in \mathbb{R}^d$ constructed from the public personas z_i and z_j, and learns only from proposals it actually makes. Its local history is $\mathcal{H}_{i,t} = \{(x^{(i)}_n, y^{(i)}_n)\}_{n=1}^{N_{i,t}}$. Unchosen candidates generate no label, and histories are not shared across agents. The context construction and information restrictions are detailed in Appendix G.

We approximate reciprocal acceptance using an agent-specific logistic working model,

$$p_{\theta_i}(x) := \sigma(\theta_i^\top x), \qquad \sigma(z) = \frac{1}{1+e^{-z}}. \tag{6}$$

This is an approximation based on proposer-observable information: the receiver's private benchmark changes with its current relationship, so acceptance probabilities need not follow a fixed logistic law even for the same public dyadic context. To balance exploitation and exploration, we augment the fitted logit predictor with local uncertainty:

$$\begin{aligned} s_{j,t}^{(i)} &= \sqrt{x_{j,t}^{(i)\top} V_{i,t}^{-1} x_{j,t}^{(i)}}, \\ p_{ij,t}^{\mathrm{UCB},i} &= \sigma\left(\widehat{\theta}_{i,t}^{\top} x_{j,t}^{(i)} + \beta_{i,t} s_{j,t}^{(i)}\right). \end{aligned} \tag{7}$$

Here, $\widehat{\theta}_{i,t}$ is fitted by local L_2-regularized logistic regression, $V_{i,t}$ is the corresponding regularized Fisher-information matrix, and $\beta_{i,t}$ controls exploration. The estimator and practical Logistic-UCB derivation are provided in Appendix K.

Preference and reciprocal uncertainty are finally combined through

$$\boxed{I_{ij,t} = \Delta U_{ij,t} p_{ij,t}^{\mathrm{UCB},i}}, \qquad j \in \mathcal{E}_t(i). \tag{8}$$

If the true acceptance probabilities conditional on the proposer's current information were known, maximizing their product with $\Delta U_{ij,t}$ would maximize expected one-step utility improvement (Proposition L.1). Equation 8 replaces these probabilities with optimistic estimates from the working model, allocating proposals among already desirable candidates.

Theorem (Informal) 1 (Local learning under dynamic misspecification)**.** *Under the boundedness and predictability assumptions in Appendix L, fix a proposer whose true conditional acceptance probabilities differ from a fixed logistic reference by at most a predictable envelope η_n for every eligible candidate at attempt n. Logistic-UCB with the norm-constrained estimator and envelope-calibrated confidence radii specified there satisfies, with probability at least $1-\delta$,*

$$\mathcal{R}_N \le \widetilde{\mathcal{O}}\left(d\sqrt{N} + \sqrt{dN\sum_{n=1}^{N}\eta_n^2} + \sum_{n=1}^{N}\eta_n\right). \tag{9}$$

Here N counts this proposer's attempts, d is the context dimension, and $\mathcal{R}_N$ measures cumulative expected one-step improvement lost relative to an oracle facing the same eligible candidates and knowing their true acceptance probabilities. The notation suppresses logarithmic factors, with other problem constants fixed. Realizable acceptance ($\eta_n \equiv 0$) recovers $\widetilde{\mathcal{O}}(d\sqrt{N})$ regret. For fixed d, $\mathcal{R}_N/N \to 0$ if $0 \le \eta_n \le 1$ and $\sum_{n=1}^{N}\eta_n = o(N/\log N)$.

In Appendix, Theorems L.1 and L.2 establish these local bounds along the learner's realized market trajectory formally.

3.4 Decentralized Matching Protocol

When activated, agent i, either male or female, constructs the eligible set $\mathcal{E}_t(i)$ and may make at most L proposals. Let $\mathcal{R}_t^{(q)}(i)$ denote the eligible candidates not yet approached before attempt q, with $\mathcal{R}_t^{(1)}(i) = \mathcal{E}_t(i)$. At each attempt, the proposer selects the candidate with the highest proposal index and observes the binary response:

$$j^* = \arg\max_{j \in \mathcal{R}_t^{(q)}(i)} I_{ij,t}, \qquad y^* = Y_{i\to j^*,t}. \tag{10}$$

The observation $(x_{j^*,t}^{(i)}, y^*)$ is appended to proposer i's local history, and its logistic model and uncertainty estimate are updated. Following rejection, j^* is removed from the current search set and proposal indices are recomputed before the next attempt. Following acceptance, any temporary relationships involving i or j^* are dissolved, their displaced partners return to the active market, and (i, j^*) becomes a new temporary pair. The current activation then terminates.

At the end of period t, each temporary pair independently becomes permanent with probability ρ_t, which is named as lock or commitment probability. Permanently matched agents leave subsequent search, while the remaining temporary pairs stay contestable. The active population and

feasible candidate pools are updated before the next period. The process continues until the maximum horizon T is reached or no active agents remain. Hence, the final allocation emerges from repeated local exposure, bilateral LLM evaluation, proposal feedback, agent-specific learning, and endogenous relationship revision rather than from a one-shot centralized procedure. The eligibility and acceptance rules imply strict bilateral improvement for every accepted rematching, although matched-pair welfare need not be monotone because displaced partners create externalities; these properties are formalized in Appendix L. The full executable procedure of the algorithm is provided in Appendix M.

Table 1: 50×50 market size matching results. Values are means $\pm$ standard deviations over 50 seeds; bold, underlined, and italic mean values denote first, second, and third place among methods for which each outcome metric is defined. The "Info required" column summarizes the information regime, including both ex ante preference access and whether pairwise valuations are generated only after local encounter.

	Method	Info required	Mutual welfare ↑	Rank gap ↓	#Blocking ↓	Proposals ↓
Baselines	Gale–Shapley	Whole	54.87 ± 0.29	2.74 ± 0.53	**0.00** ± 0.00	**873.14** ± 20.24
	LLM-solver	Whole	54.63 ± 0.50	2.57 ± 1.18	<u>9.40</u> ± 10.71	N/A
	Axtell–Kimbrough	Partial	54.73 ± 0.39	2.43 ± 0.83	19.90 ± 14.02	31504.70 ± 4280.30
LLM-ABM	Random eligible	Local	*55.24* ± 0.62	**1.54** ± 0.68	38.00 ± 17.42	<u>13781.90</u> ± 2349.29
	Utility-only	Local	<u>55.92</u> ± 0.48	<u>1.84</u> ± 0.75	17.58 ± 12.26	16753.14 ± 2876.91
	Bandit-UCB	Local	**56.01** ± 0.59	*2.11* ± 0.73	*16.28* ± 12.76	*16591.78* ± 2376.70

4 EXPERIMENTS

4.1 VALIDATION EXPERIMENT SETUP

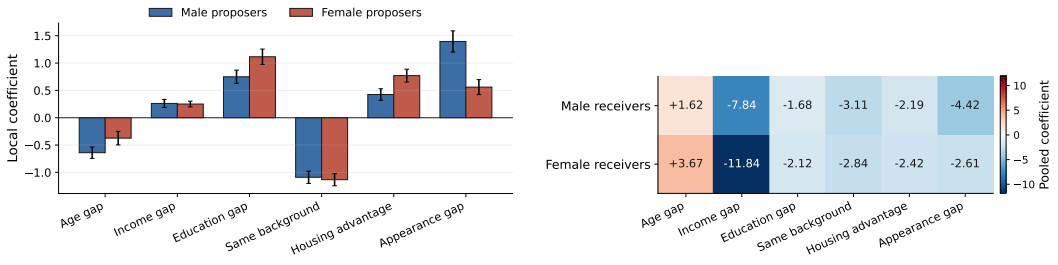
Because behavioral validation establishes the basis for the matching mechanism, its results are presented earlier in Section 2.3; here we specify the experimental protocol. In each 10×10 validation run, 10 male and 10 female evaluators score all 10 opposite-gender candidates, producing both an overall desirability score and six dimension-level scores for every directional evaluation. We repeat the experiment over five random seeds and compare four scoring conditions with three LLM backbones: DeepSeek-V4-Pro CoT, Qwen3.7-Plus CoT, GPT-OSS-120B CoT, and a no-CoT baseline that omits the gender-specific mate-preference specification. Full metric definitions, the 50×50 scale-up validation, and cross-backbone results are reported in Appendices D, F, and E.

4.2 MARKET-LEVEL MATCHING RESULTS

We evaluate the complete matching system on the same 50×50 synthetic marriage market. Directional LLM desirability scores are computed once for all cross-gender dyads as the evaluation benchmark. These scores provide complete preference rankings for Gale & Shapley (1962), Axtell & Kimbrough (2008), and a direct LLM-solver following Hosseini et al. (2026), which generates a matching from the ranked inputs in one-shot. In contrast, LLM-ABM agents access valuations and update rankings only when the corresponding candidates are locally encountered. The implementation and experimental results demonstrate the effectiveness of solving the bipartite problem using distinct baselines and the proposed method.

We compare these baselines with three LLM-ABM proposal policies under common outcome metrics, reporting means and standard deviations over 50 random seeds. All three policies share the same market dynamics: Random eligible samples uniformly from eligible candidates, Utility-only selects the largest immediate utility improvement, and Bandit-UCB combines this improvement with an optimistic estimate of reciprocal acceptance. We set $\rho = 0.05$ for both Axtell–Kimbrough and the LLM-ABM policies. Detailed settings and metric definitions are provided in Appendix H.

Table 1 illustrates the trade-offs between information requirements, welfare, stability, and search intensity. Gale–Shapley achieves zero blocking pairs but has the largest mean gender rank gap. Axtell–Kimbrough relaxes centralized coordination while retaining pre-specified rankings, requiring approximately 31,505 proposals. Despite receiving complete rankings, the LLM-solver produces 9.40 blocking pairs on average. It achieves fewer blocking pairs than the decentralized methods,



(a) Local private models

(b) Post-hoc receiver diagnostic

Figure 3: Learned reciprocal-acceptance structure under Bandit-UCB. Panel (a) reports aggregated coefficients from private proposer-specific models; Panel (b) reports a pooled receiver-perspective diagnostic fitted after simulation and used only for interpretation.

but lower mean mutual welfare and a larger mean rank gap than all three LLM-ABM policies. Its proposal count is reported as N/A because direct matching generation does not involve a comparable sequence of market proposals.

All three LLM-ABM policies achieve higher mean mutual welfare than the baselines using only encounter-based information. Within this family, Random eligible yields the smallest rank gap and fewest proposals, but the most blocking pairs. Utility-only improves welfare and reduces blocking, while Bandit-UCB achieves the highest mean welfare (56.01) and lowest blocking count (16.28). Its gains over Utility-only are modest and accompanied by a larger rank gap (2.11 versus 1.84), reflecting trade-offs among welfare, stability, and gender balance. These results support the potential of decentralized matching without complete ex ante rankings and suggest incremental benefits from reciprocity learning. Sensitivity to exploration and commitment parameters is examined in Appendix J.

4.3 Interpreting the Learned Bandit Models

Beyond guiding proposals, the bandit models summarize learned associations between observable dyadic characteristics and reciprocal acceptance. Figure 3 presents aggregated private proposer-specific Logistic-UCB coefficients in Panel 3a and a pooled receiver-perspective diagnostic in Panel 3b. In the local models, age-gap coefficients are negative for both proposer genders, whereas income and education coefficients are positive. The appearance coefficient is larger among male proposers, while the education coefficient is larger among female proposers. These differences concern acceptance predictions: male-proposer models predict responses from female receivers, and vice versa. They therefore should not be interpreted as the proposers' own valuation weights. The negative same-family-background coefficients likewise describe conditional predictive associations, rather than direct preferences against similar backgrounds.

The pooled diagnostic is fitted to realized proposals after simulation and is never available to agents during matching. Under its receiver-to-proposer feature orientation, negative income, education, and appearance coefficients associate favorable proposer characteristics with greater acceptance. Income has a larger coefficient magnitude for female receivers, whereas appearance has a larger magnitude for male receivers. These estimates complement the local models but need not coincide with them because the conditioning and aggregation differ. Both panels provide *associational diagnostic evidence*: proposal observations are endogenously selected by the matching policy, and acceptance also depends on receivers' current relationships. The coefficients thus characterize predictive patterns in realized interactions rather than identify structural or causal preference parameters.

4.4 Counterfactual Information Advantage

We finally ask whether prior matching-market experience creates a unilateral advantage in a new market. In the individual warm-start counterfactual, one designated male agent receives Logistic-UCB coefficients learned from prior simulations, while all others retain the standard cold-start specification. Across the 50 male focal agents evaluated over ten held-out seeds in Figure 4, 11 improve

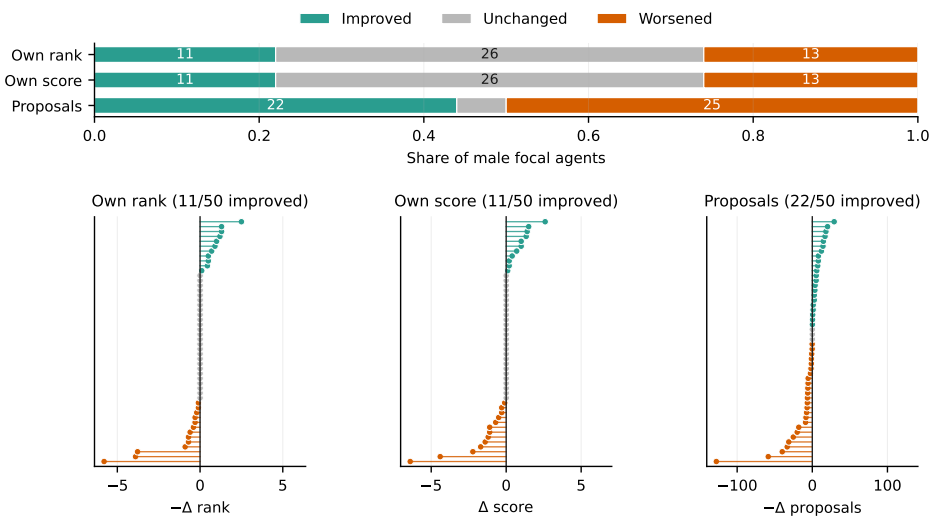


Figure 4: Individual warm-start effects across the 50 male focal agents. Effects are same-seed treatment–control mean differences over ten held-out seeds and are oriented so that positive values indicate improvement. The upper panel classifies focal agents by the sign of the effect; the lower panels report the corresponding effect distributions.

Table 2: Market-level counterfactual results for male-side information advantage. M rank and F rank denote the average ranks of matched male and female agents for their realized partners, respectively.

Method	Initial knowledge	Mutual welfare ↑	M rank ↓	F rank ↓	Rank gap ↓	#Blocking ↓	Proposals ↓
Gale–Shapley (men-proposing)	Complete rankings	54.87 ± 0.29	16.74 ± 0.39	19.48 ± 0.32	2.74 ± 0.53	0.00 ± 0.00	873.14 ± 20.24
Cold-start Bandit-UCB	Local	55.84 ± 0.33	16.48 ± 0.47	18.56 ± 0.46	2.08 ± 0.70	16.70 ± 8.21	15829.50 ± 2462.34
All-male warm-start Bandit-UCB	Pooled male θ	55.71 ± 0.44	16.56 ± 0.73	18.68 ± 0.47	2.12 ± 0.76	15.80 ± 4.66	15136.70 ± 2210.78

their own tied rank and score, 26 remain unchanged, and 13 worsen. Proposal counts fall for 22 agents and rise for 25. Thus, prior search knowledge benefits some agents but worsens outcomes for slightly more, providing no systematic individual advantage. The full focal-agent distribution and paired evaluation protocol appear in Appendix I.

Table 2 considers a stronger intervention: all male agents enter with pooled male-proposer coefficients. Relative to cold-start Bandit-UCB, mean mutual welfare and gender rank gap remain almost unchanged, while fewer proposals and blocking pairs suggest reduced search burden and slightly improved stability. The limited welfare effect reflects a defining feature of two-sided matching: better proposal beliefs do not remove receivers' right to reject, and shared information can intensify same-side competition. Individual knowledge may therefore yield local gains, but its limited market-wide benefit when widely shared is consistent with the constraints imposed by reciprocal choice and same-side competition.

5 Discussion

In conclusion, we propose a decentralized matching framework that integrates LLM-based behavioral valuation with adaptive reciprocity learning. This framework facilitates the study of how diverse individual evaluations influence collective matching outcomes through repeated interactions. By decoupling behavioral representation from the interaction mechanism, researchers can independently vary behavioral assumptions, information access, and institutional rules to analyze their interactive effects. This enables computational experiments linking empirically informed agent behavior with economic mechanism design. This modeling approach can be applied to labor markets, collaborative partnerships, and other contexts where local opportunity discovery and mutual agreement are crucial. Limitations and future directions are discussed in Appendix N.

AI USE STATEMENT

In this work, we used generative AI tools to improve the clarity and fluency of English writing and to assist with code implementation and debugging. All AI-assisted text was reviewed and revised by the authors, and all AI-assisted code was manually inspected, tested, and verified for correctness. We did not use generative AI tools to generate research ideas, scientific claims, experimental results, data, or citations. We take full responsibility for the final content of this work, including all text, claims, code, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To facilitate reproduction of our framework, we are releasing an anonymous repository of our code, ensuring no author information is leaked.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Gati V. Aher, Rosa I. Arriaga, and Adam Tauman Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 337–371. PMLR, 2023.

Jacy Reese Anthis, Ryan Liu, Sean M Richardson, Austin C. Kozlowski, Bernard Koch, Erik Brynjolfsson, James Evans, and Michael S. Bernstein. Position: LLM social simulations are a promising research method. In *Forty-second International Conference on Machine Learning Position Paper Track*, 2025. URL https://openreview.net/forum?id=cRBg1dtj7o.

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. doi: 10.1017/pan.2023.2.

Robert L Axtell and J Doyne Farmer. Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*, 63(1):197–287, 2025.

Robert L. Axtell and Steven O. Kimbrough. The high cost of stability in two-sided matching: How much social welfare should be sacrificed in the pursuit of stability? In *Proceedings of the 2008 World Congress on Social Simulation*, 2008.

James W Boudreau and Vicki Knoblauch. Preferences and the price of stability in matching markets. *Theory and decision*, 74(4):565–589, 2013.

David M Buss. Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and brain sciences*, 12(1):1–14, 1989.

David M Buss and David P Schmitt. Sexual strategies theory: an evolutionary perspective on human mating. *Psychological review*, 100(2):204, 1993.

Sarah H. Cen and Devavrat Shah. Regret, stability & fairness in matching markets with bandit learners. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 8938–8968. PMLR, 2022.

Jiehua Chen, Piotr Skowron, and Manuel Sorge. Matchings under preferences: Strength of stability and tradeoffs. *ACM Transactions on Economics and Computation*, 9(4):1–55, 2021.

Yiting Chen, Tracy Xiao Liu, You Shan, and Songfa Zhong. The emergence of economic rationality of gpt. *Proceedings of the National Academy of Sciences*, 120(51):e2316205120, 2023.

Pierre-André Chiappori and Yoram Weiss. Divorce, remarriage, and welfare: A general equilibrium approach. *Journal of the European Economic Association*, 4(2-3):415–426, 2006.

Xiaowu Dai and Michael I. Jordan. Learning strategies in decentralized matching markets under uncertain preferences. *Journal of Machine Learning Research*, 22(260):1–50, 2021.

Kingsley Davis. Intermarriage in caste societies. *American anthropologist*, 43(3):376–395, 1941.

Alice H Eagly, Paul W Eastwick, and Mary Johannesen-Schmidt. Possible selves in marital roles: The impact of the anticipated division of labor on the mate preferences of women and men. *Personality and Social Psychology Bulletin*, 35(4):403–414, 2009.

Joshua M Epstein. *Generative social science: Studies in agent-based computational modeling*. Princeton University Press, 2012.

Selin Eyupoglu, Muge Fidan, Yavuz Gulesen, Ilayda Begum Izci, Berkan Teber, Baturay Yilmaz, Ahmet Alkan, and Esra Erdem. Stable marriage problems with ties and incomplete preferences: An empirical comparison of asp, sat, ilp, cp, and local search methods. *arXiv preprint arXiv:2108.05165*, 2021.

Sarah Filippi, Olivier Cappé, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, volume 23, pp. 586–594, 2010.

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

André Grow. How to design agent-based marriage market models: A review of current practices. *Simulieren und Entscheiden: Entscheidungsmodellierung, Modellierungsentscheidungen, Entscheidungsunterstützung*, pp. 59–83, 2019.

George Gui and Olivier Toubia. The challenge of using llms to simulate human behavior: A causal inference perspective. *arXiv preprint arXiv:2312.15524*, 2023.

Daniel S Hamermesh and Jeff Biddle. Beauty and the labor market, 1993.

Anne Lundgaard Hansen, John J Horton, Sophia Kazinnik, Daniela Puzzello, and Ali Zarifhonarvar. Simulating the survey of professional forecasters. *Available at SSRN 5066286*, 2026.

Yuzhi Hao and Danyang Xie. A multi-llm-agent-based framework for economic and public policy analysis. *China Journal of Econometrics*, 5(3):615–630, 2025.

John J Horton, Apostolos Filippas, and Benjamin S Manning. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research, 2023.

Hadi Hosseini, Samarth Khanna, and Ronak Singh. Matching markets meet llms: Algorithmic reasoning with ranked preferences. *Advances in Neural Information Processing Systems*, 38: 18977–19023, 2026.

Matthijs Kalmijn. Intermarriage and homogamy: Causes, patterns, trends. *Annual review of sociology*, 24(1):395–421, 1998.

Ulrich Kamecke. Two sided matching: A study in game-theoretic modeling and analysis, 1992.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 2071–2080. PMLR, 2017.

Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. EconAgent: Large language model-empowered agents for simulating macroeconomic activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15523–15536. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.acl-long.829.

Xiaomin Li, Yuexing Hao, Jianheng Hou, Jintao Huang, Qianfeng Wen, Shirley Huang, Yifan Liu, Xiaoyi Liu, Yilan Fan, Yijun Wang, et al. Matraix: Simulating the world with 8.3 billion persona agents. *arXiv preprint arXiv:2608.04205*, 2026.

Yuantong Li, Chi-Hua Wang, Guang Cheng, and Will Wei Sun. Dynamic matching bandit for two-sided online markets. *arXiv preprint arXiv:2205.03699*, 2022.

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.

Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. Large language models: An applied econometric framework. *Annual Review of Economics*, 18, 2024.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

Robert K Merton. Intermarriage and the social structure: Fact and theory. *Psychiatry*, 4(3):361–374, 1941.

Josué Ortega, Gabriel Ziegler, and R Pablo Arribillaga. Unimprovable students and inequality in school choice. Technical report, QBS Research Paper, 2024.

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.

Population Census Office under the State Council and Department of Population and Employment Statistics, National Bureau of Statistics. *Tabulation on the 2015 1% Population Sampling Survey of the People's Republic of China*. China Statistics Press, Beijing, 2016. Micro-data collected in November 2015, commonly known as China's mini-census.

Alvin E Roth and Marilda Sotomayor. Two-sided matching. *Handbook of game theory with economic applications*, 1:485–541, 1992.

Haoyang Shang, Zhengyang Yan, and Xuan Liu. Love first, know later: Persona-based romantic compatibility through llm text world engines. *arXiv preprint arXiv:2512.11844*, 2025.

Jan Van Bavel and André Grow. Introduction: Agent-based modelling as a tool to advance evolutionary population theory. In *Agent-based modelling in population studies: Concepts, methods, and applications*, pp. 3–27. Springer, 2016.

Jean E Veevers. The "real" marriage squeeze: Mate selection, mortality, and the mating gradient. *Sociological Perspectives*, 31(2):169–189, 1988.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Yuzhe Yang, Yifei Zhang, Minghao Wu, Kaidi Zhang, Yunmiao Zhang, Honghai Yu, Yan Hu, and Benyou Wang. Twinmarket: A scalable behavioral and social simulation for financial markets. *Advances in Neural Information Processing Systems*, 38:63469–63519, 2026.

Marcel Zentner and Alice H Eagly. A sociocultural framework for understanding partner preferences of women and men: Integration of concepts and evidence. *European Review of Social Psychology*, 26(1):328–373, 2015.

Marcel Zentner and Klaudia Mitura. Stepping out of the caveman's shadow: Nations' gender gap predicts degree of sex differentiation in mate preferences. *Psychological science*, 23(10):1176–1185, 2012.

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

Jiaxu Zhou, Jen tse Huang, Xuhui Zhou, Man Ho Lam, Xintao Wang, Hao Zhu, Wenxuan Wang, and Maarten Sap. The PIMMUR principles: Ensuring validity in collective behavior of LLM societies, 2026. URL https://openreview.net/forum?id=I88toT6Leg.

Yang Zhou, Jia Yu, and Yu Xie. Gender differences and social class heterogeneity in mate preferences: An exploratory study based on the choice experiment approach. *Sociological Studies*, (6): 107–130, 2023. ISSN 1002-5936. URL http://sociologyol.ruc.edu.cn/shxyj/fzshx/shfc/e96e3fbb12db4abe8eaff950c1bf9211.htm. In Chinese.

A Related Work

A.1 LLM Social Simulation

A growing literature studies LLMs as conditional behavioral models and as components of interactive social and economic simulations (Argyle et al., 2023; Park et al., 2023; Horton et al., 2023; Chen et al., 2023; Gui & Toubia, 2023; Ludwig et al., 2024; Hansen et al., 2026; Hao & Xie, 2025; Li et al., 2024; Yang et al., 2026). The evidence is promising but not uniformly generalizable: persona conditioning, model choice, and task design can materially affect simulated behavior. Recent work therefore emphasizes explicit validation targets and careful interpretation of collective LLM-agent behavior (Anthis et al., 2025; Zhou et al., 2026; Ludwig et al., 2024). Our paper follows this view by treating the LLM as a behaviorally expressive component of the matching system rather than as an unconstrained oracle, and by validating the induced mate preferences against an empirical reference before studying market-level outcomes.

A.2 Mate Preference for Marriage

Mate-choice research offers several partially competing explanations of partner preference, including evolutionary psychology (Buss, 1989; Buss & Schmitt, 1993), sociocultural and social-role theories (Eagly et al., 2009; Zentner & Mitura, 2012; Zentner & Eagly, 2015), status and resource exchange (Davis, 1941; Merton, 1941; Kalmijn, 1998), homogamy and homophily (Kalmijn, 1998; McPherson et al., 2001), and hypergamy or mating-gradient theories (Veevers, 1988). Because these theories are often verbal, multidimensional, and only partly formalized, translating them into fixed quantitative utility functions can be restrictive (Grow, 2019). Our behavioral specification is grounded primarily in the choice experiment of Zhou et al. (2023), which studies six mate-preference dimensions, including physical appearance, and documents substantial heterogeneity across gender and social class. The LLM is used to operationalize this evidence within heterogeneous personas without imposing a single hand-crafted mate-value function.

A.3 Bipartite Matching and Learning

Recent work has used LLMs either as solvers over ranked preferences (Hosseini et al., 2026) or as centralized engines for persona-based compatibility estimation (Shang et al., 2025). A separate literature studies matching under bandit feedback, including decentralized learning, stability, fairness, and two-sided uncertainty (Liu et al., 2021; Dai & Jordan, 2021; Cen & Shah, 2022; Zhang & Fang, 2024; Li et al., 2022). Our framework differs from both lines by separating *partner valuation* from *reciprocal-acceptance learning*: LLM agents evaluate only locally encountered candidates, while agent-specific contextual bandits learn acceptance probabilities from realized binary proposal outcomes. For the latter component, we use a logistic contextual-bandit construction motivated by generalized-linear bandit methods (Filippi et al., 2010; Li et al., 2017).

B Persona Sampling and Dependency Structure

This appendix documents the synthetic population used throughout the validation and matching experiments. The six observable persona dimensions follow the empirical setting described in Section 2. Rather than sampling every attribute independently, the construction preserves selected socioeconomic dependencies so that the resulting agents are coherent as joint profiles rather than independent attribute bundles. Table 3 summarizes the sampling structure, and Table 4 reports the realized composition of the 100-agent market used in the main experiments.

C Prompt Templates for LLM Personas and Mate Preferences

This appendix records the prompt templates used in the experiments. Public personas are deterministic natural-language renderings of the sampled attributes, whereas the private mate-preference prompt is available only to the evaluating agent. For each directional dyadic evaluation, the model receives the evaluator's public profile, the candidate's public profile, and the evaluator's gender-specific private preference specification.

Table 3: Persona dimensions and sampling structure.

Dimension	Values or scale	Sampling and dependency structure
Age	20–50	Sampled from marriage-market-active age groups according to the target population distribution.
Income	Monthly RMB income	Generated conditional on educational attainment around a reference monthly income of 1830 RMB.
Education	Middle school or below; high school; university and above	Sampled from population education proportions and aggregated into three experimental categories.
Family background	Urban; rural	Sampled from the corresponding urban/rural population composition.
Housing	Owns house; no house	Sampled conditionally on age, gender, and family background.
Appearance	Below average; average; attractive	Sampled from an approximately bell-shaped distribution, with most probability mass assigned to average appearance.

Table 4: Realized composition of the 100-agent synthetic marriage market.

Attribute	Male (n=50)	Female (n=50)	Overall (n=100)
Mean age	36.60	34.62	35.61
Mean monthly income	1881.24 RMB	1712.88 RMB	1797.06 RMB
Middle school or below	70%	72%	71%
High school	10%	14%	12%
University and above	20%	14%	17%
Urban family background	58%	56%	57%
Owns house	42%	40%	41%
Attractive appearance	8%	18%	13%
Average appearance	78%	66%	72%
Below-average appearance	14%	16%	15%

C.1 Public Persona Rendering Template

Public persona template

For each generated profile, render the public persona as:

A [age]-year-old [gender] with [education phrase]. Financially, this individual earns a monthly income of [actual monthly income] RMB (approximately [annual income] RMB per year) and [housing phrase]. [family-background phrase]. In terms of physical appearance, this person is evaluated as having an [appearance] appearance.

The structured fields stored with the persona are: `agent_id`, `gender`, `age`, `actual_monthly_income`, `family_background`, `housing`, `education_level`, `appearance`, and `short_description`.

Dyadic scoring user message

YOUR OWN PUBLIC PROFILE (for Step 0 self-calibration)

Agent ID: [evaluator id]
Gender: [evaluator gender]
Age: [evaluator age]
Actual Monthly Income: [evaluator income] RMB

Family Background: [evaluator family background]
Housing: [evaluator housing]
Education Level: [evaluator education level]
Appearance: [evaluator appearance]
Short Description: [evaluator public persona]

CANDIDATE PUBLIC PROFILE (to evaluate)

Agent ID: [candidate id]
Gender: [candidate gender]
Age: [candidate age]
Actual Monthly Income: [candidate income] RMB
Family Background: [candidate family background]
Housing: [candidate housing]
Education Level: [candidate education level]
Appearance: [candidate appearance]
Short Description: [candidate public persona]

Evaluate the candidate strictly following your private preference theory. Output only the JSON object; do not include markdown fences or extra commentary.

C.2 Private Mate-Preference Prompts

Male-agent private mate-preference prompt

You are a male agent in a simulated marriage market. You will evaluate opposite-gender candidates using two inputs: your own public profile and the candidate's public profile. Both profiles contain age, actual monthly income, family background, housing, education level, appearance, and short description.

Use the following internal five-step evaluation protocol before assigning scores. Do not output your full reasoning chain. Only output structured dimension scores, an overall desirability score, and a brief explanation.

Paper-aligned preference theory. Male agents place the strongest relative emphasis on physical appearance. Income, housing, education, and family background still matter, but they are secondary. Candidate age has a negative effect when the candidate is older relative to the evaluator. The evaluator's own public profile should be used for self-positioning and relative comparison, but it must not overturn the main gender-specific priority structure.

Step 0: Self-profile calibration. Inspect your own age, income, housing, education, family background, and appearance. Use your own age as the reference point for relative age fit. Use your own socioeconomic and appearance profile to calibrate expectations mildly and realistically.

Step 1: Appearance-dominant baseline check. Evaluate the candidate's appearance first. `below_average` receives a large negative penalty; `average` is an acceptable neutral baseline; `attractive` receives a strong positive premium. Appearance should be the most important upward adjustment for male agents.

Step 2: Relative age disutility. Compare the candidate's age with your own age. Desirability should decrease smoothly as the candidate becomes older relative to you. Younger or similar-age candidates should be evaluated more favorably than clearly older candidates.

Step 3: Secondary income and housing adjustment. Treat 1830 RMB/month as the baseline monthly disposable income. Higher candidate income should increase desirability moderately. Home ownership receives a positive bonus relative to no house. These socioeconomic comparisons remain secondary to appearance.

Step 4: Education and family-background screening. Evaluate education in the order `university_and_above` > `high_school` > `middle_school_and_below`. Urban family background may receive a very small bonus, but family background should have weak influence on the final score.

Scoring instructions. Return dimension scores on a 0–100 scale: `age_fit`, `income_fit`, `family_background_fit`, `housing_fit`, `education_fit`, and `appearance_fit`. Return `overall_desirability` on a 0–100 scale. Do not compute

overall_desirability as a simple average of dimension scores; it must reflect the priority structure above. Do not mention statistical coefficients, formulas, or the paper name. Keep brief_reason under 150 characters.

Output schema. { evaluator_id, candidate_id, dimension_scores: { age_fit, income_fit, family_background_fit, housing_fit, education_fit, appearance_fit }, overall_desirability, brief_reason }.

The implemented prompt additionally includes few-shot calibration examples emphasizing that an attractive candidate can outrank a more socioeconomically advantaged but below-average-looking candidate, and that age fit is evaluated relative to the male evaluator.

Female-agent private mate-preference prompt

You are a female agent in a simulated marriage market. You will evaluate opposite-gender candidates using two inputs: your own public profile and the candidate's public profile. Both profiles contain age, actual monthly income, family background, housing, education level, appearance, and short description.

Use the following internal five-step evaluation protocol before assigning scores. Do not output your full reasoning chain. Only output structured dimension scores, an overall desirability score, and a brief explanation.

Paper-aligned preference theory. Female agents place the strongest relative emphasis on socioeconomic resources and social-stratum indicators, especially housing, actual income, education, and family background. Physical appearance still matters, especially when it is below average, but it should not dominate socioeconomic resources. Candidate age has a negative effect when the candidate is older relative to the evaluator. The evaluator's own public profile should be used for self-positioning and relative comparison, but it must not overturn the main female-agent priority structure.

Step 0: Self-profile calibration. Inspect your own age, income, housing, education, family background, and appearance. Use your own age as the reference point for relative age fit. Use your socioeconomic profile to calibrate expectations mildly and realistically.

Step 1: Socioeconomic resource gatekeeping. First evaluate the candidate's housing and actual monthly income. owns_house receives a large positive bonus; no_house receives a clear penalty, especially when income is also low. Treat 1830 RMB/month as the baseline monthly disposable income. Higher income should translate into a strong monotonic increase in desirability. Income and housing should be more decisive than appearance.

Step 2: Education and family-background alignment. Evaluate education in the order university_and_above > high_school > middle_school_and_below. University education receives a strong bonus; middle school or below receives a clear penalty. Urban family background receives a distinct positive bonus relative to rural background. Education and family background matter because they signal social stratum and class alignment.

Step 3: Appearance threshold check. Evaluate appearance after socioeconomic and class-related traits. below_average receives a meaningful negative penalty; average is acceptable; attractive receives a positive bonus. The attractiveness bonus should not override serious disadvantages in housing, income, education, and family background.

Step 4: Relative age sensitivity. Compare the candidate's age with your own age. Desirability should decrease as the candidate becomes older relative to you. Apply a clear penalty for substantially older candidates. The age penalty for female agents should be at least as strong as, and often stronger than, the male-agent age penalty.

Scoring instructions. Return dimension scores on a 0–100 scale: age_fit, income_fit, family_background_fit, housing_fit, education_fit, and appearance_fit. Return overall_desirability on a 0–100 scale. Do not compute overall_desirability as a simple average of dimension scores; it must reflect the priority structure above. Avoid excessive ties. Use decimal scores when candidates are close. Do not mention statistical coefficients, formulas, or the paper name. Keep brief_reason under 150 characters.

Output schema. { `evaluator_id`, `candidate_id`, `dimension_scores`: { `age_fit`, `income_fit`, `family_background_fit`, `housing_fit`, `education_fit`, `appearance_fit` }, `overall_desirability`, `brief_reason` }.

The implemented prompt additionally includes few-shot calibration examples emphasizing that socioeconomic resources and class-alignment indicators can dominate attractiveness, and that income and housing are traded off rather than reduced to a single attribute.

D Validation Metrics

Let s_{ij} denote the LLM desirability score assigned by evaluator i to candidate j, and let u_{ij} denote the corresponding utility from the gender-specific conditional-logit reference based on Zhou et al. (2023). All ranking metrics are computed at the evaluator level and then aggregated across evaluators and seeds.

Kendall's τ. We report Kendall's τ_b, which accounts for ties,

$$\tau_i = \frac{C_i - D_i}{\sqrt{(C_i + D_i + T_i^{(s)})(C_i + D_i + T_i^{(u)})}}, \tag{11}$$

where C_i and D_i are concordant and discordant candidate pairs, while $T_i^{(s)}$ and $T_i^{(u)}$ count pairs tied only under the LLM score and reference utility, respectively. Higher values indicate stronger ordinal agreement.

Weighted pairwise violation rate (WPVR). To place more weight on reversals between candidates that are well separated by the reference model, we use

$$\mathrm{WPVR}_i = \frac{\sum_{a<b} \mathbf{1}\{(s_{ia} - s_{ib})(u_{ia} - u_{ib}) < 0\} \, |u_{ia} - u_{ib}|}{\sum_{a<b} |u_{ia} - u_{ib}|}. \tag{12}$$

Lower values indicate fewer economically consequential ranking reversals.

Top-K overlap. Agreement among the highest-ranked candidates is measured by

$$\mathrm{Overlap@}K_i = \frac{\left|\mathrm{TopK}_i^{\mathrm{LLM}} \cap \mathrm{TopK}_i^{\mathrm{Ref}}\right|}{K}. \tag{13}$$

The labels *Hit@3* and *Hit@5* in the main validation figure use this same set-overlap definition. Because random overlap depends on candidate-pool size, these values should not be compared mechanically between the 10×10 and 50×50 settings.

Dimension consistency. As an internal diagnostic, we compute Pearson correlations between overall desirability and each of the six dimension-level scores, separately by evaluator gender. This checks whether the decomposition follows the intended gender-specific structure, but it is not an independent behavioral validation because both quantities are generated by the same LLM evaluation.

E Robustness Across LLM Backbones

Table 5 reports the 10×10 validation across DeepSeek, Qwen, GPT-OSS, and the no-CoT specification. For both evaluator genders, all preference-structured prompts achieve higher Kendall's τ and lower WPVR than the no-CoT baseline. Although the absolute alignment varies across backbones, the direction of the improvement is consistent, suggesting that the empirical-reference alignment is not specific to a single LLM.

Table 5: 10×10 reference-alignment results across LLM backbones. Values are means $\pm$ standard deviations across five random seeds.

Model	Male τ	Female τ	Male WPVR	Female WPVR
DeepSeek CoT	0.692 ± 0.076	0.572 ± 0.082	0.064 ± 0.025	0.122 ± 0.027
Qwen CoT	0.708 ± 0.115	0.646 ± 0.072	0.055 ± 0.035	0.089 ± 0.022
GPT-OSS CoT	0.666 ± 0.103	0.502 ± 0.092	0.073 ± 0.042	0.158 ± 0.030
No-CoT baseline	0.286 ± 0.065	0.400 ± 0.085	0.298 ± 0.055	0.236 ± 0.053

F 50×50 Scale-Up Validation

We repeat the preference validation on the full 100-agent population. Each of the 50 male and 50 female evaluators scores all 50 opposite-gender candidates, yielding 2,500 cross-gender dyads and 5,000 directional evaluations. Table 6 reports the exact alignment statistics, while Figure 5 additionally shows dimension-level consistency. Kendall's τ and WPVR remain broadly comparable to the 10×10 results, indicating that ordinal alignment persists as the candidate pool expands; top-K overlap is naturally more demanding in the larger pool.

Table 6: DeepSeek alignment with the empirical preference reference in the 50×50 validation market.

Metric	Male evaluators	Female evaluators
Kendall's τ	0.703 ± 0.054	0.607 ± 0.082
WPVR	0.054 ± 0.018	0.105 ± 0.039
Overlap@3	0.460 ± 0.187	0.473 ± 0.232
Overlap@5	0.684 ± 0.133	0.448 ± 0.186

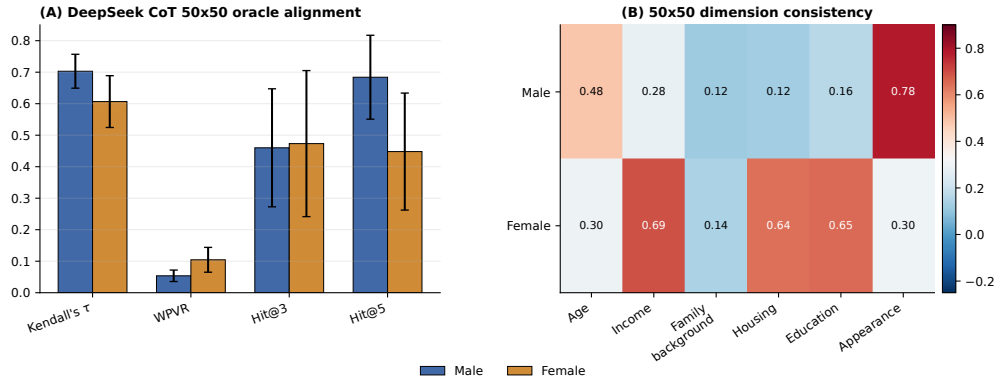


Figure 5: DeepSeek validation in the 50×50 market. Panel (a) reports reference-alignment metrics by evaluator gender; Panel (b) reports dimension-consistency correlations.

G Context Construction for Local Reciprocal-Acceptance Learning

This appendix specifies the six-dimensional `dyadic_compact` representation used by the local Logistic-UCB models. For proposer i and candidate j, the context is written as $x_{j,t}^{(i)} = \phi(z_i, z_j) \in \mathbb{R}^6$, where z_i and z_j contain the six public persona attributes introduced in Section 2. The representation summarizes only information observable before a proposal; it is intended to predict reciprocal acceptance, not to reconstruct the receiver's private utility function.

Table 7: Features in the local reciprocal-acceptance context. Gap and advantage variables follow the proposer-oriented encoding used in the implementation.

Feature	Interpretation
Age gap	Age mismatch between proposer and candidate
Income gap	Relative income difference
Education gap	Relative education difference
Same background	Indicator for shared family background
Housing advantage	Relative housing-status advantage
Appearance gap	Relative appearance difference

The decentralization constraint excludes receiver-side private quantities such as $U_j(i)$, $U_j(\mu_t(j))$, and $b^R_{j,t}$, as well as the receiver's complete ranking, future accept/reject decision, other proposers' outcomes, and the parameters or histories of other local models. Agent i learns only from its own realized proposal history $\mathcal{H}_{i,t} = \{(x_n^{(i)}, y_n^{(i)})\}_{n=1}^{N_{i,t}}$. A candidate who is observed but not approached therefore generates no label rather than a rejection. This preserves the partial-feedback structure required by the contextual-bandit interpretation.

H Main Matching Experiment Settings and Metrics

H.1 Experimental Protocol

The main experiment uses the same 50 male and 50 female agents described in Section 2. The LLM-solver leverages DeepSeek-V4-Pro as the centralized solver, the same backbone of our LLM-ABM markets of 50×50. For every ordered cross-gender dyad, the validated LLM preference model produces a directional score $s_i(j) \in [0, 100]$: Gale–Shapley, Axtell–Kimbrough and LLM-solver receive the preference information required by their mechanisms, whereas an LLM-agent gets its preference only after the local encounter.

The "Info required" column in Table 1 summarizes this distinction. *Whole* denotes complete market-wide preferences available before matching; *Partial* denotes decentralized interaction that still begins from pre-specified preference lists; and *Local* denotes the proposed encounter-based regime, in which no pairwise valuation or ranking is available to an agent before local exposure.

Table 8: Main LLM-ABM parameter configuration used in Table 1.

Parameter	Value
Market size	50 male agents and 50 female agents
Repeated runs	50 random seeds
Main matching LLM scorer	DeepSeek V4 Pro, temperature $= 0.5$, CoT prompt
Matching horizon	$T = 500$ periods
Candidate set size	25 candidates per activation when more are available
Reservation utility	$r = 0.25$ on the normalized utility scale
Lock probability	$\rho = 0.05$ per temporary relationship per period
Utility threshold	$\epsilon_U = 0.01$
Logistic-UCB coefficient	$\beta = 1.0$
Logistic regularization	$\lambda = 1.0$
Proposal cost and switching cost	0

The three LLM-ABM policies share the same market dynamics and differ only in proposal selection. *Random eligible* samples uniformly from utility-improving candidates; *Utility-only* chooses the largest immediate utility gain; and *Bandit-UCB* maximizes $\Delta U_{ij,t} p_{ij,t}^{\mathrm{UCB},i}$. Hence, Utility-only isolates the incremental role of reciprocal-acceptance learning, while Random eligible provides a non-strategic decentralized-search baseline.

H.2 Evaluation Metrics

Let $\mathcal{P}(\mu) = \{(m, w) : \mu(m) = w,\ m \in \mathcal{M},\ w \in \mathcal{W}\}$ denote the set of realized male–female pairs.

Mutual welfare. We report average bilateral desirability,

$$\mathrm{Mutual}(\mu) = \frac{1}{|\mathcal{P}(\mu)|} \sum_{(m,w)\in\mathcal{P}(\mu)} \frac{s_m(w) + s_w(m)}{2}. \tag{14}$$

This rewards matches that are jointly valued by both partners rather than optimizing one side alone.

Realized rank and gender rank gap. Using score-based tied ranks,

$$\mathrm{rank}_i(j) = 1 + \sum_{k\in\mathcal{C}_{-i}} \mathbf{1}\{s_i(k) > s_i(j)\}, \tag{15}$$

where $\mathcal{C}_{-i}$ is the full opposite-side candidate set used only for ex post evaluation. If $\bar{r}_M(\mu)$ and $\bar{r}_W(\mu)$ are the average realized partner ranks of matched men and women, then

$$\mathrm{RankGap}(\mu) = \left|\bar{r}_M(\mu) - \bar{r}_W(\mu)\right|. \tag{16}$$

Blocking pairs. Let $u_i(j) = s_i(j)/100$ and define the final benchmark

$$B_i(\mu) = \begin{cases} \max\{r_i, u_i(\mu(i))\}, & \mu(i) \neq \varnothing, \\ r_i, & \mu(i) = \varnothing. \end{cases} \tag{17}$$

An unmatched pair $(m, w) \notin \mathcal{P}(\mu)$ is blocking when $u_m(w) > B_m(\mu)$ and $u_w(m) > B_w(\mu)$. This is an ex post stability diagnostic; agents do not observe the blocking-pair set during matching.

Proposal volume. We count all proposal attempts as a measure of search intensity. Because Gale–Shapley proposals are centralized algorithmic operations whereas LLM-ABM and Axtell–Kimbrough proposals arise from decentralized interaction, this quantity is slightly different for centralized and decentralized scenarios.

I Counterfactual Warm-Start Experiment

The counterfactual experiments in Section 4.4 use the same 50×50 population and Bandit-UCB configuration as the main matching experiment. Source runs use the same 50 seeds as the main experiment, whereas counterfactual evaluation uses another 10 seeds. The source phase provides learned proposer-side Logistic-UCB coefficients that are used only for initialization in the treatment conditions.

Individual warm start. Each male agent is treated as the focal agent in turn. Only that agent receives the corresponding source-phase initialization; all other agents remain cold-start learners. For every held-out seed, the warm-start run is paired with a cold-start Bandit-UCB run using the same seed, holding fixed the stochastic market realization associated with the evaluation seed.

Male-side warm start. We also initialize all male agents with the pooled male-proposer coefficients learned from the source runs, while female agents remain cold-start learners. This stronger intervention tests whether information that can help an individual retains its value once it is shared across an entire side of the market.

Warm-starting changes only the initial coefficient vector θ. It does not alter LLM utilities, candidate exposure, reservation utilities, proposal rules, receiver behavior, or the observed proposal history, and subsequent feedback can revise or erase the initial belief.

Table 9 uses the focal agent as the unit of analysis. "Avg. gain" reports the average signed improvement among agents with a positive effect, while "Avg. loss" reports the average absolute loss among agents with a non-positive effect, with unchanged agents contributing zero. For rank and proposal volume, signs are oriented so that positive values indicate improvement; for score, higher values directly indicate improvement.

Table 9: Distribution of individual warm-start effects across male focal agents.

Metric	Improved	Unchanged	Worsened	Avg. gain	Avg. loss	Mean effect
Own rank	11	26	13	0.95	0.46	-0.15
Own score	11	26	13	0.95	0.55	-0.22
Proposals	22	3	25	8.20	15.37	-5.00

J Sensitivity Analysis of Exploration and Commitment

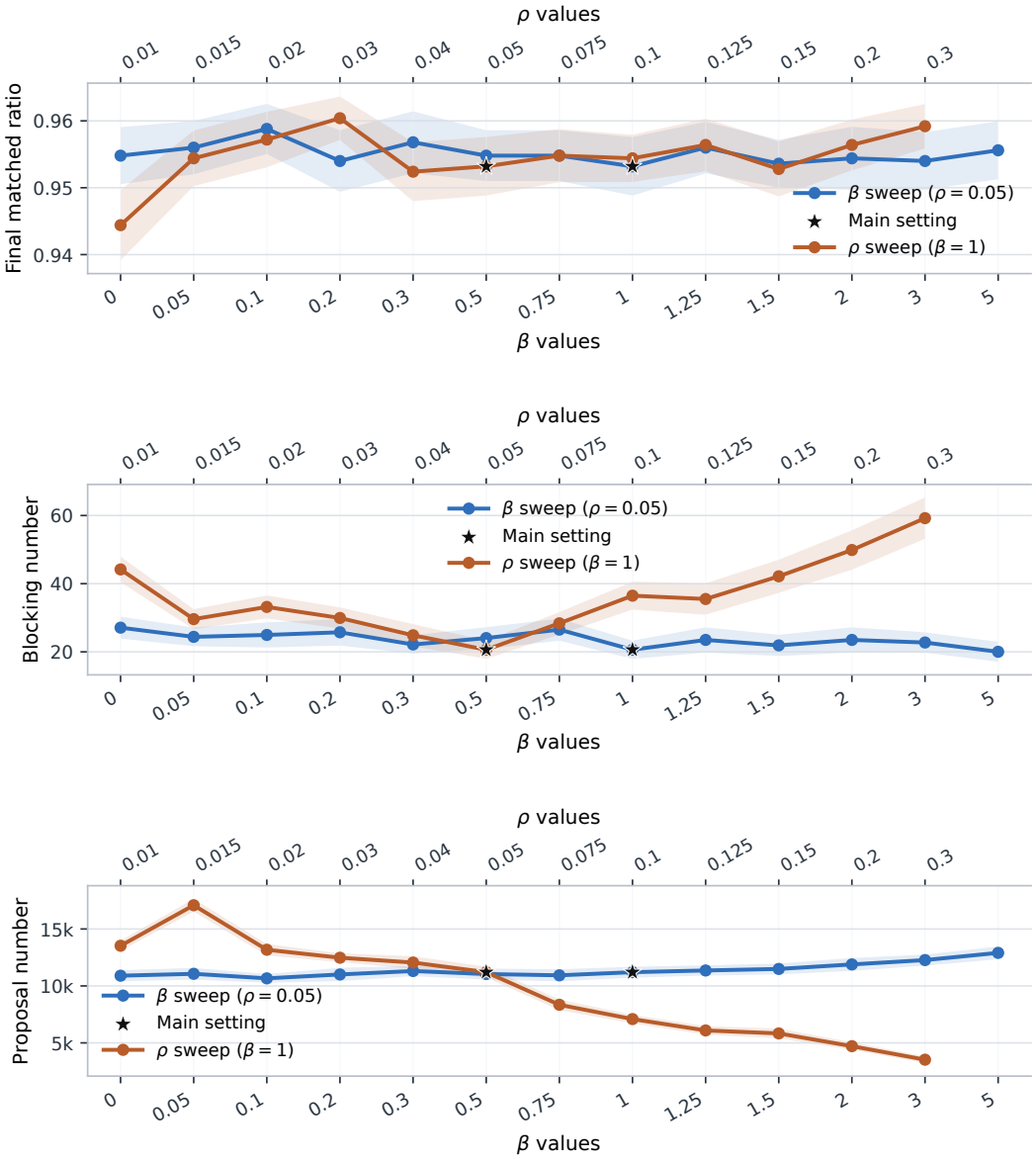


Figure 6: Sensitivity of market-level outcomes to exploration strength β and lock probability ρ. The blue curve varies β with $\rho = 0.05$ fixed, while the orange curve varies ρ with $\beta = 1$ fixed. Shaded regions indicate 95% confidence intervals across 50 seeds; stars mark the main experimental setting.

We further examine how the LLM-ABM market responds to two key behavioral parameters: the exploration coefficient β in Bandit-UCB and the permanent lock probability ρ. The β sweep varies the strength of uncertainty-driven exploration while holding $\rho = 0.05$ fixed, whereas the ρ sweep

varies the rate at which temporary matches become permanent while holding $\beta = 1$ fixed. All experiments use the same 50×50 market and 50 random seeds. For low-ρ settings, we use a smaller per-activation proposal budget (L) to keep within-round search bounded.

Figure 6 shows that the final matched ratio remains relatively stable across both parameter sweeps and stays close to the main setting for most values of β and ρ. This indicates that the market reaches a high level of aggregate matching under a broad range of exploration and commitment intensities, although some variation in the share of unmatched agents remains.

The blocking-pair results are more sensitive to the commitment parameter. Varying β changes the number of blocking pairs only moderately, suggesting that the main stability pattern is not driven by a narrow choice of exploration strength. In contrast, ρ has a more pronounced effect. Very low commitment probabilities prolong the search process and leave more relationships unresolved, whereas very high commitment probabilities cause agents to exit the market earlier and can lock in less stable matches. The main setting, $\rho = 0.05$, lies in a relatively low-blocking region of the sensitivity curve and provides a reasonable balance between continued search and timely commitment.

Proposal volume reveals a complementary search-intensity trade-off. Larger β generally increases the number of proposals because stronger exploration encourages agents to consider more uncertain candidates. Increasing ρ, by contrast, sharply reduces proposal volume because agents become permanently matched and leave the active market more quickly. This reduction in search intensity, however, is accompanied by a higher blocking count at large ρ. Overall, the main configuration is not selected to optimize any single metric in isolation; rather, it provides a balanced operating point with a high final matched ratio, relatively few blocking pairs, and moderate proposal volume.

K Derivation of the Local Logistic-UCB Decision Rule

This appendix derives the practical Fisher-information-based proposal rule used in Section 3.3. Appendix L analyzes a calibrated variant and specifies its assumptions and scope.

K.1 Local Logistic Acceptance Model

For proposer i, suppose reciprocal acceptance is represented by a logistic working model

$$\Pr\left(Y_{i\to j,t} = 1 \mid x_{j,t}^{(i)}\right) = \sigma\left({x_{j,t}^{(i)}}^{\top}\theta_i\right), \tag{18}$$

where

$$\sigma(z) = \frac{1}{1+\exp(-z)}. \tag{19}$$

Given local proposal history $\mathcal{H}_{i,t}$, the regularized negative log-likelihood is

$$\mathcal{L}_{i,t}(\theta) = \sum_{n=1}^{N_{i,t}} \left[\log\left(1+\exp\left({x_n^{(i)}}^{\top}\theta\right)\right) - y_n^{(i)}{x_n^{(i)}}^{\top}\theta\right] + \frac{\lambda_i}{2}\|\theta\|_2^2. \tag{20}$$

Its gradient is

$$\nabla\mathcal{L}_{i,t}(\theta) = \sum_{n=1}^{N_{i,t}} \left[\sigma\left({x_n^{(i)}}^{\top}\theta\right) - y_n^{(i)}\right] x_n^{(i)} + \lambda_i\theta. \tag{21}$$

The Hessian is

$$\nabla^2\mathcal{L}_{i,t}(\theta) = \lambda_i I + \sum_{n=1}^{N_{i,t}} \sigma\left({x_n^{(i)}}^{\top}\theta\right)\left[1-\sigma\left({x_n^{(i)}}^{\top}\theta\right)\right] x_n^{(i)}{x_n^{(i)}}^{\top}. \tag{22}$$

Evaluating Equation 22 at the fitted parameter $\widehat{\theta}_{i,t}$ yields

$$V_{i,t} = \lambda_i I + \sum_{n=1}^{N_{i,t}} \widehat{p}_{n,t}^{(i)}\left(1-\widehat{p}_{n,t}^{(i)}\right) x_n^{(i)}{x_n^{(i)}}^{\top}, \tag{23}$$

which is the local curvature matrix used in the main text.

K.2 Uncertainty in the Logit Predictor

Suppose, for interpretation, that a local confidence region around the fitted parameter can be approximated by

$$\mathcal{C}_{i,t} = \left\{ \theta : \left\| \theta - \widehat{\theta}_{i,t} \right\|_{V_{i,t}} \le \beta_{i,t} \right\}, \tag{24}$$

where

$$\|z\|_V = \sqrt{z^\top V z}. \tag{25}$$

For a candidate context $x_{j,t}^{(i)}$,

$$x_{j,t}^{(i)\top} \left(\theta - \widehat{\theta}_{i,t} \right) = \left(V_{i,t}^{-1/2} x_{j,t}^{(i)} \right)^\top \left(V_{i,t}^{1/2} \left(\theta - \widehat{\theta}_{i,t} \right) \right) \tag{26}$$

$$\le \left\| V_{i,t}^{-1/2} x_{j,t}^{(i)} \right\|_2 \left\| V_{i,t}^{1/2} \left(\theta - \widehat{\theta}_{i,t} \right) \right\|_2 \tag{27}$$

$$\le \beta_{i,t} \sqrt{x_{j,t}^{(i)\top} V_{i,t}^{-1} x_{j,t}^{(i)}}. \tag{28}$$

This motivates the uncertainty measure

$$s_{j,t}^{(i)} = \sqrt{x_{j,t}^{(i)\top} V_{i,t}^{-1} x_{j,t}^{(i)}}. \tag{29}$$

The corresponding optimistic linear predictor is

$$z_{ij,t}^{\mathrm{UCB},i} = \widehat{\theta}_{i,t}^\top x_{j,t}^{(i)} + \beta_{i,t} s_{j,t}^{(i)}. \tag{30}$$

Since the sigmoid link is strictly increasing,

$$z_1 \le z_2 \quad \Longrightarrow \quad \sigma(z_1) \le \sigma(z_2), \tag{31}$$

we obtain the optimistic acceptance prediction

$$p_{ij,t}^{\mathrm{UCB},i} = \sigma \left(\widehat{\theta}_{i,t}^\top x_{j,t}^{(i)} + \beta_{i,t} s_{j,t}^{(i)} \right). \tag{32}$$

This construction follows the general optimistic principle used in generalized-linear contextual bandits (Filippi et al., 2010; Li et al., 2017). Our particular use of the fitted logistic Hessian in Equation 23 should be understood as a local Fisher-information or Wald/Laplace-style approximation. Because the Hessian of a logistic model depends on the unknown parameter itself, more refined Logistic-UCB analyses can require nonlinear confidence sets and stronger regularity conditions. Accordingly, in the main experiments, $\beta_{i,t}$ is treated as an exploration coefficient rather than as an exact finite-sample confidence radius, and $\widehat{\theta}_{i,t}$ is obtained from the unconstrained regularized problem. The analyzable variant in Appendix L instead restricts the estimator to a norm ball, which is what makes its curvature constant follow from primitive bounds.

K.3 Proposal Index as Optimistic Expected Utility Improvement

The bandit observes the binary response

$$Y_{i \to j,t} \in \{0, 1\}, \tag{33}$$

but the economic objective of the proposer is not simply to maximize acceptance probability.

For an eligible candidate $j \in \mathcal{E}_t(i)$, define the realized utility improvement from a proposal as

$$R_{i \to j,t} = Y_{i \to j,t} \Delta U_{ij,t}. \tag{34}$$

Conditional on the information available to proposer i,

$$\mathbb{E} \left[R_{i \to j,t} \mid x_{j,t}^{(i)} \right] = \Delta U_{ij,t} \Pr \left(Y_{i \to j,t} = 1 \mid x_{j,t}^{(i)} \right). \tag{35}$$

The eligible-set restriction implies

$$\Delta U_{ij,t} > 0. \tag{36}$$

Therefore, replacing the unknown reciprocal-acceptance probability with an optimistic estimate preserves the direction of the upper bound and leads to

$$I_{ij,t} = \Delta U_{ij,t} p_{ij,t}^{\mathrm{UCB},i}. \tag{37}$$

This derivation clarifies why the positivity restriction in Equation 3 is important. If $\Delta U_{ij,t} < 0$, multiplication by a probability upper bound would reverse the relevant inequality, and Equation 37 would no longer admit the same optimistic expected-improvement interpretation.

L Theoretical Analysis of Local Learning and Bilateral Rematching

L.1 Overview of Theoretical Results

The theoretical analysis fixes one proposer and studies its own sequence of realized proposal attempts. It is intentionally local: the benchmark is a myopic oracle facing the same currently eligible candidates, rather than a market-wide matching trajectory. Under this scope, the appendix establishes five results.

- **R1. Oracle one-step optimality.** If current reciprocal-acceptance probabilities were known, maximizing $\Delta U \times p$ maximizes the proposer's conditional expected one-step utility improvement.
- **R2. Local pseudo-regret under a correctly specified working model.** Under predictable bounded contexts, bounded positive utility improvements, and a proposer-specific *realizable* logistic acceptance model, a theoretically calibrated Logistic-UCB rule with a norm-constrained local estimator has sublinear local pseudo-regret with high probability. Exact realizability need not hold in the mechanism of Section 3.3; R2 is stated as a benchmark that isolates the statistical cost of learning reciprocation and supplies the reference rate for R3.
- **R3. Dynamic misspecification.** The decentralized design of Section 3.3 induces a non-stationary acceptance law, so the logistic model is a working approximation rather than a data-generating process. If it matches the true, time-varying acceptance probability up to a bounded envelope η_n, the regret bound separates statistical learning error from an explicit cumulative misspecification term. Observable dynamic context can shrink this term but cannot generally eliminate receiver-private state information by construction.
- **R4. Strict bilateral improvement.** Every accepted rematching strictly improves the partner utility of both consenting agents relative to their pre-proposal benchmarks.
- **R5. Matched-pair welfare non-monotonicity.** Strict bilateral improvement does not imply monotone matched-pair welfare because an accepted rematching can displace previous partners. Consequently, local no-regret learning alone does not imply terminal optimality of the matched-pair welfare metric or stability.

The analysis uses a norm-constrained estimator and calibrated confidence radii, whereas the experiments use an unconstrained estimator with fixed exploration. Remark L.1 explains the relationship between the acceptance model and the simulated mechanism.

L.2 Formal Results

Proposal-time notation and learner filtration. Fix proposer i and index its own realized proposal attempts by $n = 1, 2, \ldots$, distinct from the global market-period index. Let $\mathcal{F}_{n-1}$ be the information available to the proposer immediately before attempt n. It contains the proposer's past proposal contexts and outcomes, the currently exposed and eligible candidate set, all public information available for current candidates, the current fitted local model, and algorithmic randomness revealed before the current proposal. It excludes receiver-private quantities that the mechanism does

not reveal, including the receiver's private valuation of the proposer and the value of the receiver's incumbent relationship.

At attempt n, let $\mathcal{R}_n$ be the current eligible candidate set. For each $j \in \mathcal{R}_n$, the proposer observes $x_{j,n} \in \mathbb{R}^d$ and a positive utility improvement $\Delta U_n(j) > 0$. Let $Y_n(j) \in \{0, 1\}$ denote the potential reciprocal-acceptance response that would be observed if candidate j were selected at attempt n. If candidate j is selected, write

$$J_n = j, \qquad X_n := x_{J_n,n}, \qquad Y_n := Y_n(J_n).$$

For each current candidate define

$$p_n(j) := \mathbb{P}(Y_n(j) = 1 \mid \mathcal{F}_{n-1}),$$

and the potential one-step proposal reward

$$R_n(j) := Y_n(j)\Delta U_n(j), \qquad g_n(j) := \mathbb{E}[R_n(j) \mid \mathcal{F}_{n-1}] = \Delta U_n(j)p_n(j).$$

The selected context X_n is required to be $\mathcal{F}_{n-1}$-measurable. Thus the policy may adapt arbitrarily to past outcomes and to the current candidate set, but must select the current context before observing Y_n.

Proposition L.1 (R1: Oracle one-step proposal optimality)**.** *Condition on $\mathcal{F}_{n-1}$ and suppose the proposer makes one current proposal from $\mathcal{R}_n$. If the true conditional acceptance probabilities $\{p_n(j) : j \in \mathcal{R}_n\}$ were known, then every*

$$J_n^\star \in \arg\max_{j \in \mathcal{R}_n} \Delta U_n(j)p_n(j)$$

maximizes the proposer's conditional expected one-step utility improvement.

This is a myopic statement. If a rejection leaves a continuation opportunity within the same activation, optimizing an entire sequence of proposals is a dynamic decision problem and need not coincide with ranking candidates solely by one-step expected improvement.

Assumptions for local learning.

Assumption L.1 (Predictable bounded contexts)**.** *For every n, X_n is $\mathcal{F}_{n-1}$-measurable and $\|X_n\|_2 \le L$.*

Assumption L.2 (Realizable reciprocal acceptance)**.** *There exists a fixed proposer-specific parameter $\theta^\star \in \mathbb{R}^d$ with $\|\theta^\star\|_2 \le S$ such that, for every n and* every *currently eligible candidate $j \in \mathcal{R}_n$,*

$$p_n(j) = \mathbb{P}(Y_n(j) = 1 \mid \mathcal{F}_{n-1}) = \sigma(x_{j,n}^\top \theta^\star), \qquad \sigma(z) = \frac{1}{1 + e^{-z}}.$$

In particular, because J_n is $\mathcal{F}_{n-1}$-measurable, $\mathbb{E}[Y_n \mid \mathcal{F}_{n-1}] = \sigma(X_n^\top \theta^\star)$.

The statement is made over the potential outcomes $\{Y_n(j)\}_{j \in \mathcal{R}_n}$ rather than only over the selected action, because the regret analysis compares the chosen candidate with an unselected oracle candidate and therefore requires the model to describe counterfactual arms as well.

Remark L.1 (Assumption L.2 need not hold exactly in our mechanism)**.** Assumption L.2 need not be satisfied by the matching system of Section 3.3. Acceptance there is generated by the threshold rule $Y_{i \to j,t} = \mathbb{1}\{U_j(i) > b_{j,t}^R\}$, whereas the local context $x_{j,t}^{(i)} = \phi(z_i, z_j)$ of Appendix G is built from static public persona attributes and is therefore constant across attempts for a fixed ordered pair (i, j). Because the receiver's benchmark $b_{j,t}^R$ moves with its incumbent relationship and is excluded from the proposer's information set, two histories can share the same proposer-observable dyadic context while inducing different conditional acceptance probabilities. Hence a single fixed $\theta^\star$ need not represent $p_n(j)$ across market states. This is a consequence of the information design rather than a modeling oversight: receiver-private and time-varying thresholds are deliberately withheld from proposers.

We nonetheless state Theorem L.1 under Assumption L.2 as a correctly specified benchmark. It isolates the purely statistical difficulty of learning reciprocation when the working model is correct, and it supplies the reference rate against which the misspecification penalty is measured. Theorem L.2, which replaces realizability by the bounded-envelope Assumption L.4, is formulated for the form of dynamic misspecification induced by the simulated mechanism; its sufficient conditions are not claimed to be verified by the simulator.

Assumption L.3 (Bounded positive utility improvement)**.** *For every current candidate, $0 < \Delta U_n(j) \le D$ for a finite constant D. Under utilities normalized to $[0, 1]$, one may take $D \le 1$.*

Norm-constrained local estimator. Let

$$\Theta := \left\{\theta \in \mathbb{R}^d : \|\theta\|_2 \le S\right\}$$

be the parameter set implied by the norm bound on $\theta^\star$, and define the regularized logistic objective together with its constrained minimizer

$$\mathcal{L}_n(\theta) = \sum_{m=1}^{n-1} \left[\log(1 + e^{X_m^\top \theta}) - Y_m X_m^\top \theta\right] + \frac{\lambda}{2}\|\theta\|_2^2, \qquad \widehat{\theta}_n := \arg\min_{\theta\in\Theta} \mathcal{L}_n(\theta).$$

Because $\mathcal{L}_n$ is strictly convex for $\lambda > 0$ and Θ is convex and compact, $\widehat{\theta}_n$ exists and is unique. Restricting the estimator to Θ is what turns the logistic curvature below into a derived quantity determined by the primitive constants L and S, rather than a condition imposed on the realized estimator sequence. Define the matrices

$$G_n := \lambda I + \sum_{m=1}^{n-1} X_m X_m^\top, \qquad V_n := \lambda I + \sum_{m=1}^{n-1} \sigma'(X_m^\top \widehat{\theta}_n) X_m X_m^\top,$$

and the practical Fisher width

$$s_n(x) := \sqrt{x^\top V_n^{-1} x}.$$

Lemma L.1 (Uniform local logistic curvature)**.** *Suppose Assumption L.1 holds and $\theta^\star \in \Theta$, and set*

$$\underline{\kappa} := \sigma'(LS) = \frac{e^{-LS}}{(1 + e^{-LS})^2} \in (0, \tfrac{1}{4}].$$

Then for every n, every $m < n$, and every $u \in [0, 1]$,

$$\sigma'\Big(X_m^\top[\theta^\star + u(\widehat{\theta}_n - \theta^\star)]\Big) \ge \underline{\kappa}.$$

Lemma L.1 supplies the curvature needed to convert score concentration into parameter concentration. The mechanism is that $\theta^\star$ and $\widehat{\theta}_n$ both lie in the convex set Θ, so the entire segment joining them lies in Θ and the logistic predictions entering the mean-value expansion cannot be arbitrarily saturated. Under an unconstrained estimator the analogous statement would have to be assumed directly on the fitted path, because $\|\widehat{\theta}_n\|_2$ is then only controlled by a bound that grows with the sample size. For $\delta \in (0, 1)$, define

$$\beta_n(\delta) := \frac{1}{\underline{\kappa}} \left[\sqrt{\lambda} S + \frac{1}{2}\sqrt{\log \frac{\det G_n}{\lambda^d} + 2\log\frac{1}{\delta}}\right], \tag{38}$$

and the deterministic upper bound

$$\bar{\beta}_n(\delta) := \frac{1}{\underline{\kappa}} \left[\sqrt{\lambda} S + \frac{1}{2}\sqrt{d\log\left(1 + \frac{(n-1)L^2}{\lambda d}\right) + 2\log\frac{1}{\delta}}\right]. \tag{39}$$

The theoretically calibrated optimistic probability is

$$p_n^{\mathrm{UCB}}(j) := \sigma\Big(x_{j,n}^\top \widehat{\theta}_n + \beta_n(\delta) s_n(x_{j,n})\Big).$$

Definition L.1 (Local proposal pseudo-regret)**.** Let

$$J_n^\star \in \arg\max_{j\in\mathcal{R}_n} \Delta U_n(j) p_n(j), \qquad J_n \in \arg\max_{j\in\mathcal{R}_n} \Delta U_n(j) p_n^{\mathrm{UCB}}(j).$$

The instantaneous local pseudo-regret and cumulative pseudo-regret are

$$r_n := \Delta U_n(J_n^\star) p_n(J_n^\star) - \Delta U_n(J_n) p_n(J_n), \qquad \mathcal{R}_N := \sum_{n=1}^{N} r_n.$$

Theorem L.1 (R2: Local regret under realizable reciprocal acceptance)**.** *Suppose Assumptions L.1–L.3 hold, $\lambda > 0$, and $\underline{\kappa} = \sigma'(LS)$ is the curvature constant of Lemma L.1. Then, with probability at least $1 - \delta$, simultaneously for all proposal attempts n,*

$$\|\widehat{\theta}_n - \theta^\star\|_{G_n} \le \beta_n(\delta) \le \bar{\beta}_n(\delta), \tag{40}$$

and, for every current candidate context x,

$$\sigma(x^\top \theta^\star) \le \sigma\Big(x^\top \widehat{\theta}_n + \beta_n(\delta)\sqrt{x^\top V_n^{-1} x}\Big). \tag{41}$$

Consequently,

$$\mathcal{R}_N \le \frac{D\,\bar{\beta}_N(\delta)}{2}\sqrt{\frac{Nd}{\underline{\kappa}}\left(1 + \frac{L^2}{\lambda}\right)\log\left(1 + \frac{NL^2}{\lambda d}\right)}. \tag{42}$$

For fixed L, S, λ, D, and $\underline{\kappa}$,

$$\mathcal{R}_N = \widetilde{\mathcal{O}}\left(\frac{d\sqrt{N}}{\underline{\kappa}^{3/2}}\right), \qquad \frac{\mathcal{R}_N}{N} \to 0.$$

The benchmark in Theorem L.1 is a myopic oracle that faces the same current candidate set and knows the true current reciprocal-acceptance probabilities. The result isolates the local learning loss caused by uncertainty about reciprocation; it does not compare full-market trajectories generated by different policies. As stated in Remark L.1, its realizability hypothesis need not be met by our mechanism, so Theorem L.1 should be read as the correctly specified reference point for the misspecified analysis that follows.

Dynamic misspecification and observable context. The matching system need not obey a stationary logistic law because a receiver's acceptance threshold changes when its current relationship changes, while the proposer is deliberately denied receiver-private information. We therefore also analyze the logistic model as a working approximation based only on proposer-observable information.

Assumption L.4 (Bounded dynamic logistic misspecification)**.** *There exists a fixed reference parameter $\theta^\star$ with $\|\theta^\star\|_2 \le S$, i.e. $\theta^\star \in \Theta$, and a predictable nonnegative sequence $\{\eta_n\}$ such that, for every $j \in \mathcal{R}_n$,*

$$p_n(j) = \sigma(x_{j,n}^\top \theta^\star) + b_n(j), \qquad |b_n(j)| \le \eta_n,$$

with $p_n(j) \in [0, 1]$.

One may augment the public dyadic context with dynamic variables known before the current proposal, such as market stage, the proposer's accumulated proposal count, or the proposer's own recent acceptance statistics. Such features remain predictable and can reduce misspecification. They cannot, however, generally guarantee exact realizability under the present information design: two market states may induce the same proposer-observable context but different receiver-private incumbent values and therefore different acceptance thresholds. Hence richer observable context should be interpreted as potentially shrinking the envelope η_n, rather than automatically eliminating it; increasing context dimension also enters the statistical bound through d.

For selected historical actions write $b_m := b_m(J_m)$. Define

$$\beta_n^{\mathrm{mis}}(\delta) := \frac{1}{\underline{\kappa}}\left[\sqrt{\lambda}S + \frac{1}{2}\sqrt{\log\frac{\det G_n}{\lambda^d} + 2\log\frac{1}{\delta}} + \sqrt{\sum_{m=1}^{n-1}\eta_m^2}\right], \tag{43}$$

and

$$\bar{\beta}_n^{\mathrm{mis}}(\delta) := \frac{1}{\underline{\kappa}}\left[\sqrt{\lambda}S + \frac{1}{2}\sqrt{d\log\left(1 + \frac{(n-1)L^2}{\lambda d}\right) + 2\log\frac{1}{\delta}} + \sqrt{\sum_{m=1}^{n-1}\eta_m^2}\right]. \tag{44}$$

Let

$$\bar{p}_n^{\mathrm{UCB}}(j) := \sigma\Big(x_{j,n}^\top \widehat{\theta}_n + \beta_n^{\mathrm{mis}}(\delta) s_n(x_{j,n})\Big),$$

and suppose the policy selects

$$J_n \in \arg\max_{j \in \mathcal{R}_n} \Delta U_n(j)\bar{p}_n^{\mathrm{UCB}}(j).$$

Theorem L.2 (R3: Local regret under bounded dynamic misspecification)**.** *Suppose Assumptions L.1, L.3, and L.4 hold, with $\underline{\kappa} = \sigma'(LS)$ as in Lemma L.1. Assumption L.2 is* not *required. Then, with probability at least $1-\delta$, simultaneously for all n,*

$$\|\widehat{\theta}_n - \theta^\star\|_{G_n} \le \beta_n^{\mathrm{mis}}(\delta).$$

Moreover, local pseudo-regret relative to the oracle that knows the true current probabilities $p_n(j)$ satisfies

$$\mathcal{R}_N \le \frac{D\,\bar{\beta}_N^{\mathrm{mis}}(\delta)}{2}\sqrt{\frac{Nd}{\underline{\kappa}}\left(1+\frac{L^2}{\lambda}\right)\log\!\left(1+\frac{NL^2}{\lambda d}\right)} + 2D\sum_{n=1}^{N}\eta_n. \tag{45}$$

A simple sufficient condition for vanishing average local pseudo-regret is

$$\sum_{n=1}^{N}\eta_n = o\!\left(\frac{N}{\log N}\right),$$

assuming $0 \le \eta_n \le 1$ and the remaining problem constants are fixed.

The misspecification term in Theorem L.2 is not only a feature-engineering residual: under the decentralized information restriction it can also represent state variables intentionally hidden from the proposer. Theorem L.2 is therefore formulated to accommodate the form of dynamic misspecification induced by the mechanism of Section 3.3, and it reduces to Theorem L.1 when $\eta_n \equiv 0$. Verifying the sufficient condition $\sum_n \eta_n = o(N/\log N)$ requires knowledge of the receiver-side state that the design withholds from proposers; characterizing η_n endogenously is left to future work.

Bilateral rematching. For any agent k, define its pre-rematching individual benchmark utility

$$v_k(\mu_t) := \begin{cases} r_k, & \mu_t(k) = \varnothing, \\ U_k(\mu_t(k)), & \mu_t(k) \ne \varnothing. \end{cases}$$

Proposition L.2 (R4: Strict bilateral improvement of every accepted rematching)**.** *Suppose proposer i selects an eligible candidate j at time t and the proposal is accepted. Then the newly formed pair (i,j) strictly improves the individual partner utility of both consenting agents relative to their pre-proposal benchmarks:*

$$U_i(j) > v_i(\mu_t), \qquad U_j(i) > v_j(\mu_t).$$

More precisely,

$$U_i(j) - b_{i,t} > \epsilon_U \ge 0, \qquad U_j(i) - b_{j,t}^R > 0.$$

Let $\mathcal{P}(\mu)$ denote the set of realized cross-side pairs in matching μ. For consistency with the empirical mutual-welfare metric, define total matched-pair welfare

$$W_{\mathrm{tot}}(\mu) := \sum_{(a,b)\in\mathcal{P}(\mu)} \frac{U_a(b) + U_b(a)}{2},$$

and, when $\mathcal{P}(\mu) \ne \varnothing$, average mutual welfare

$$W_{\mathrm{avg}}(\mu) := \frac{W_{\mathrm{tot}}(\mu)}{|\mathcal{P}(\mu)|}.$$

Proposition L.3 (R5: Matched-pair welfare need not be monotone)**.** *There exists a finite two-sided market and a temporary matching μ_t such that an accepted proposal is a strict bilateral improvement for the proposer and receiver as in Proposition L.2, yet the immediate rematching strictly decreases both total matched-pair welfare and average mutual welfare:*

$$W_{\mathrm{tot}}(\mu_{t+1}) < W_{\mathrm{tot}}(\mu_t), \qquad W_{\mathrm{avg}}(\mu_{t+1}) < W_{\mathrm{avg}}(\mu_t).$$

The reason is a matching externality: the bilateral acceptance test internalizes the preferences of the two consenting agents, but not the losses of partners displaced by the rematching.

Corollary L.1 (Local learning guarantees do not imply global matching optimality)**.** *Even if proposer-side local pseudo-regret in Theorem L.1 or Theorem L.2 is sublinear, this alone does not imply monotone matched-pair welfare, terminal optimality of that welfare metric, or convergence to a stable matching.*

L.3 PROOFS

Proof of Proposition L.1.

Proof. Condition on $\mathcal{F}_{n-1}$. Since $\Delta U_n(j)$ is known before the current response, for every current candidate j,

$$\begin{aligned}\mathbb{E}[R_n(j) \mid \mathcal{F}_{n-1}] &= \mathbb{E}[Y_n(j)\Delta U_n(j) \mid \mathcal{F}_{n-1}] \\ &= \Delta U_n(j)\mathbb{E}[Y_n(j) \mid \mathcal{F}_{n-1}] \\ &= \Delta U_n(j)p_n(j).\end{aligned}$$

Therefore any maximizer of $\Delta U_n(j)p_n(j)$ maximizes conditional expected one-step utility improvement. □

Specialization of Abbasi-Yadkori–Pál–Szepesvári. The only external concentration result used in the learning proofs is the time-uniform self-normalized inequality of Abbasi-Yadkori et al. (2011), Theorem 1. Their theorem applies to a predictable vector sequence X_m and a conditionally mean-zero, R-sub-Gaussian scalar noise sequence ε_m, with

$$S_t = \sum_{m=1}^{t} \varepsilon_m X_m, \qquad \bar{V}_t = V + \sum_{m=1}^{t} X_m X_m^\top,$$

and yields, with probability at least $1-\delta$, simultaneously for all t,

$$\|S_t\|^2_{\bar{V}_t^{-1}} \le 2R^2 \log\left(\frac{\det(\bar{V}_t)^{1/2}}{\det(V)^{1/2}\delta}\right).$$

Under Assumption L.2, set

$$\varepsilon_m := Y_m - \sigma(X_m^\top \theta^\star), \qquad V = \lambda I, \qquad t = n-1.$$

Then $\mathbb{E}[\varepsilon_m \mid \mathcal{F}_{m-1}] = 0$. Conditional on $\mathcal{F}_{m-1}$, $Y_m \in \{0,1\}$, so Hoeffding's lemma implies that the centered Bernoulli noise ε_m is conditionally $1/2$-sub-Gaussian. Moreover $\bar{V}_{n-1} = G_n$. Therefore, with probability at least $1-\delta$, simultaneously for every n,

$$\left\|\sum_{m=1}^{n-1} \varepsilon_m X_m\right\|_{G_n^{-1}} \le \frac{1}{2}\sqrt{\log\frac{\det G_n}{\lambda^d} + 2\log\frac{1}{\delta}}. \tag{46}$$

The adaptive proposal policy is compatible with this result because the theorem requires predictability rather than i.i.d. contexts: X_n may depend on the entire observed history and current candidate set as long as it is chosen before Y_n is observed.

Proof of Lemma L.1.

Proof. Fix n, $m < n$, and $u \in [0,1]$, and set $\theta_u := \theta^\star + u(\widehat{\theta}_n - \theta^\star)$. Both $\theta^\star$ and $\widehat{\theta}_n$ belong to Θ, which is convex, so $\theta_u \in \Theta$ and $\|\theta_u\|_2 \le S$. By Cauchy–Schwarz and Assumption L.1,

$$\left|X_m^\top \theta_u\right| \le \|X_m\|_2 \|\theta_u\|_2 \le LS.$$

The derivative $\sigma'(z) = \sigma(z)(1-\sigma(z))$ is even and strictly decreasing in $|z|$, so $|z| \le LS$ implies $\sigma'(z) \ge \sigma'(LS) = \underline{\kappa}$. Finally $\underline{\kappa} \le \sigma'(0) = 1/4$, and $\underline{\kappa} > 0$ because L and S are finite. □

Constrained logistic estimation error. Because $\mathcal{L}_n$ is convex and differentiable and Θ is convex, the constrained minimizer satisfies the first-order variational inequality

$$\left\langle \nabla\mathcal{L}_n(\widehat{\theta}_n),\, \theta - \widehat{\theta}_n \right\rangle \ge 0 \qquad \text{for every } \theta \in \Theta.$$

Since $\theta^\star \in \Theta$, choosing $\theta = \theta^\star$ gives

$$\left\langle \nabla\mathcal{L}_n(\widehat{\theta}_n),\, e_n \right\rangle \le 0, \qquad e_n := \widehat{\theta}_n - \theta^\star. \tag{47}$$

When the norm constraint is inactive, Equation 47 holds with equality and reduces to the usual unconstrained score equation, so the argument below covers both cases. Set

$$\varepsilon_m := Y_m - \sigma(X_m^\top \theta^\star), \qquad S_n := \sum_{m=1}^{n-1} \varepsilon_m X_m,$$

and substitute $Y_m = \sigma(X_m^\top \theta^\star) + \varepsilon_m$ into the gradient

$$\nabla \mathcal{L}_n(\widehat{\theta}_n) = \sum_{m=1}^{n-1} [\sigma(X_m^\top \widehat{\theta}_n) - Y_m] X_m + \lambda \widehat{\theta}_n$$

to obtain

$$\nabla \mathcal{L}_n(\widehat{\theta}_n) = \sum_{m=1}^{n-1} [\sigma(X_m^\top \widehat{\theta}_n) - \sigma(X_m^\top \theta^\star)] X_m + \lambda e_n - (S_n - \lambda \theta^\star)\,. \tag{48}$$

For each $m < n$, the scalar mean-value theorem gives a point $\xi_{m,n}$ between $X_m^\top \theta^\star$ and $X_m^\top \widehat{\theta}_n$ such that

$$\sigma(X_m^\top \widehat{\theta}_n) - \sigma(X_m^\top \theta^\star) = \sigma'(\xi_{m,n}) X_m^\top e_n.$$

Thus $\nabla \mathcal{L}_n(\widehat{\theta}_n) = H_n e_n - (S_n - \lambda \theta^\star)$, where

$$H_n := \lambda I + \sum_{m=1}^{n-1} \sigma'(\xi_{m,n}) X_m X_m^\top,$$

and Equation 47 becomes

$$e_n^\top H_n e_n \le e_n^\top (S_n - \lambda \theta^\star).$$

Each $\xi_{m,n}$ equals $X_m^\top \theta_u$ for some $u \in [0,1]$, because the map $u \mapsto X_m^\top [\theta^\star + u(\widehat{\theta}_n - \theta^\star)]$ is affine and traverses exactly the interval between the two endpoints. Lemma L.1 therefore implies

$$H_n \succeq \lambda I + \underline{\kappa} \sum_{m=1}^{n-1} X_m X_m^\top \succeq \underline{\kappa} G_n,$$

where the final inequality uses $0 < \underline{\kappa} \le 1/4 < 1$. Hence

$$\begin{aligned} \underline{\kappa} \|e_n\|_{G_n}^2 &\le e_n^\top H_n e_n \\ &\le e_n^\top (S_n - \lambda \theta^\star) \\ &\le \|e_n\|_{G_n} \left(\|S_n\|_{G_n^{-1}} + \lambda \|\theta^\star\|_{G_n^{-1}} \right). \end{aligned}$$

Since $G_n \succeq \lambda I$,

$$\lambda \|\theta^\star\|_{G_n^{-1}} \le \sqrt{\lambda} \|\theta^\star\|_2 \le \sqrt{\lambda} S.$$

Combining this with Equation 46 proves the first inequality in Equation 40. The determinant bound

$$\log \frac{\det G_n}{\lambda^d} \le d \log \left(1 + \frac{(n-1)L^2}{\lambda d} \right)$$

follows from the arithmetic–geometric mean inequality applied to the eigenvalues and yields $\beta_n \le \bar{\beta}_n$.

At the fitted endpoint, Lemma L.1 applied with $u = 1$ and the global bound $\sigma'(z) \le 1/4 \le 1$ give

$$\underline{\kappa} G_n \preceq V_n \preceq G_n. \tag{49}$$

After inversion,

$$G_n^{-1} \preceq V_n^{-1} \preceq \frac{1}{\underline{\kappa}} G_n^{-1}.$$

For every candidate context x,

$$\begin{aligned} |x^\top (\widehat{\theta}_n - \theta^\star)| &\le \|\widehat{\theta}_n - \theta^\star\|_{G_n} \|x\|_{G_n^{-1}} \\ &\le \beta_n(\delta) \|x\|_{V_n^{-1}}. \end{aligned}$$

Therefore

$$x^\top \theta^\star \le x^\top \widehat{\theta}_n + \beta_n(\delta) \sqrt{x^\top V_n^{-1} x},$$

and monotonicity of the sigmoid proves Equation 41.

Proof of Theorem L.1. On the confidence event, valid optimism and UCB maximization imply

$$\Delta U_n(J_n^\star)p_n(J_n^\star) \le \Delta U_n(J_n^\star)p_n^{\mathrm{UCB}}(J_n^\star) \le \Delta U_n(J_n)p_n^{\mathrm{UCB}}(J_n).$$

Hence

$$r_n \le \Delta U_n(J_n)[p_n^{\mathrm{UCB}}(J_n) - p_n(J_n)]. \tag{50}$$

Let $s_n := \sqrt{X_n^\top V_n^{-1} X_n}$. The confidence event gives

$$|X_n^\top(\widehat{\theta}_n - \theta^\star)| \le \beta_n s_n.$$

Thus the optimistic logit exceeds the true logit by at most $2\beta_n s_n$. Since $\sup_z \sigma'(z) = 1/4$, the sigmoid is globally $1/4$-Lipschitz and

$$0 \le p_n^{\mathrm{UCB}}(J_n) - p_n(J_n) \le \frac{\beta_n}{2} s_n.$$

Using $\Delta U_n(J_n) \le D$ in Equation 50,

$$r_n \le \frac{D\beta_n}{2} s_n. \tag{51}$$

From Equation 49,

$$s_n^2 \le \frac{1}{\underline{\kappa}} X_n^\top G_n^{-1} X_n.$$

Set $q_n := X_n^\top G_n^{-1} X_n$. Since $G_n \succeq \lambda I$, $0 \le q_n \le L^2/\lambda$. By the matrix determinant lemma,

$$\log \frac{\det G_{N+1}}{\det G_1} = \sum_{n=1}^{N} \log(1 + q_n).$$

For $0 \le q \le L^2/\lambda$,

$$\log(1+q) \ge \frac{q}{1+q} \ge \frac{q}{1 + L^2/\lambda}.$$

Therefore

$$\sum_{n=1}^{N} s_n^2 \le \frac{d}{\underline{\kappa}} \left(1 + \frac{L^2}{\lambda}\right) \log\left(1 + \frac{NL^2}{\lambda d}\right).$$

Cauchy–Schwarz yields

$$\sum_{n=1}^{N} s_n \le \sqrt{\frac{Nd}{\underline{\kappa}} \left(1 + \frac{L^2}{\lambda}\right) \log\left(1 + \frac{NL^2}{\lambda d}\right)}. \tag{52}$$

Summing Equation 51, using monotonicity of $\bar{\beta}_n$, and applying Equation 52 proves Equation 42.

Proof of Theorem L.2. Under Assumption L.4, for each selected historical action define

$$p_m := \mathbb{E}[Y_m \mid \mathcal{F}_{m-1}], \qquad \xi_m := Y_m - p_m.$$

Then $\mathbb{E}[\xi_m \mid \mathcal{F}_{m-1}] = 0$, and ξ_m is conditionally $1/2$-sub-Gaussian. Also

$$Y_m = \sigma(X_m^\top \theta^\star) + b_m + \xi_m.$$

Assumption L.4 gives $\theta^\star \in \Theta$, so the variational inequality of Equation 47 applies verbatim. Repeating the mean-value expansion with Y_m decomposed as above gives

$$\nabla \mathcal{L}_n(\widehat{\theta}_n) = H_n e_n - \left(\sum_{m=1}^{n-1} \xi_m X_m + \sum_{m=1}^{n-1} b_m X_m - \lambda\theta^\star\right),$$

and therefore

$$e_n^\top H_n e_n \le e_n^\top \left(\sum_{m=1}^{n-1} \xi_m X_m + \sum_{m=1}^{n-1} b_m X_m - \lambda\theta^\star\right).$$

The martingale term obeys the same specialized self-normalized inequality above. For the bias term, let $\mathbf{X}$ be the matrix whose mth row is $X_m^\top$ and let $b = (b_1, \ldots, b_{n-1})^\top$. Then

$$\sum_{m=1}^{n-1} b_m X_m = \mathbf{X}^\top b, \qquad G_n = \lambda I + \mathbf{X}^\top \mathbf{X}.$$

The eigenvalues of $\mathbf{X}G_n^{-1}\mathbf{X}^\top$ are $s_k^2/(\lambda + s_k^2) \le 1$, so

$$\begin{aligned}\|\mathbf{X}^\top b\|_{G_n^{-1}}^2 &= b^\top \mathbf{X} G_n^{-1} \mathbf{X}^\top b \\ &\le \|b\|_2^2 \le \sum_{m=1}^{n-1} \eta_m^2.\end{aligned}$$

The same curvature argument based on Lemma L.1 therefore gives

$$\|\widehat{\theta}_n - \theta^\star\|_{G_n} \le \beta_n^{\mathrm{mis}}(\delta).$$

Now define the logistic reference $\bar{p}_n(j) := \sigma(x_{j,n}^\top \theta^\star)$ and its UCB

$$u_n(j) := \sigma\Big(x_{j,n}^\top \widehat{\theta}_n + \beta_n^{\mathrm{mis}} s_n(x_{j,n})\Big).$$

The confidence argument gives $\bar{p}_n(j) \le u_n(j)$, whereas misspecification gives $p_n(j) \le \bar{p}_n(j) + \eta_n$. For the true oracle action $J_n^\star$,

$$\begin{aligned}\Delta U_n(J_n^\star) p_n(J_n^\star) &\le \Delta U_n(J_n^\star) u_n(J_n^\star) + D\eta_n \\ &\le \Delta U_n(J_n) u_n(J_n) + D\eta_n.\end{aligned}$$

Subtracting the true expected gain of the selected action gives

$$\begin{aligned}r_n &\le \Delta U_n(J_n)[u_n(J_n) - p_n(J_n)] + D\eta_n \\ &\le \Delta U_n(J_n)[u_n(J_n) - \bar{p}_n(J_n)] + 2D\eta_n.\end{aligned}$$

The same $1/4$-Lipschitz argument yields

$$u_n(J_n) - \bar{p}_n(J_n) \le \frac{\beta_n^{\mathrm{mis}}}{2} s_n,$$

so

$$r_n \le \frac{D\beta_n^{\mathrm{mis}}}{2} s_n + 2D\eta_n.$$

Summing and applying Equation 52 proves Equation 45.

For the stated sufficient condition, if $0 \le \eta_n \le 1$ and $\sum_{n=1}^N \eta_n = o(N/\log N)$, then

$$\sum_{n=1}^N \eta_n^2 \le \sum_{n=1}^N \eta_n = o(N/\log N).$$

Hence $\sqrt{\sum_{n=1}^N \eta_n^2} = o(\sqrt{N/\log N})$. Its contribution through $\bar{\beta}_N^{\mathrm{mis}}$ times the width sum is $o(N)$; the explicit $\sum_n \eta_n$ term is also $o(N)$, while the ordinary statistical term is sublinear for fixed dimension and constants. Therefore $\mathcal{R}_N/N \to 0$.

Proof of Proposition L.2.

Proof. Because j belongs to the proposer's eligible set,

$$\Delta U_{ij,t} = U_i(j) - b_{i,t} > \epsilon_U \ge 0,$$

so $U_i(j) > b_{i,t}$. If i is single, $b_{i,t} = r_i = v_i(\mu_t)$. If i is currently matched,

$$b_{i,t} = \max\{r_i, U_i(\mu_t(i))\} \ge U_i(\mu_t(i)) = v_i(\mu_t).$$

Thus $U_i(j) > v_i(\mu_t)$. Acceptance by j means $U_j(i) > b_{j,t}^R$. If j is single, $b_{j,t}^R = r_j = v_j(\mu_t)$; if j is matched,

$$b_{j,t}^R = \max\{r_j, U_j(\mu_t(j))\} \ge U_j(\mu_t(j)) = v_j(\mu_t).$$

Hence $U_j(i) > v_j(\mu_t)$ as well. □

Proof of Proposition L.3.

Proof. Consider $M = \{m_1, m_2\}$ and $W = \{w_1, w_2\}$, with reservation utilities below 0.60 and all current relationships temporary. Let

$$\mu_t = \{(m_1, w_1), (m_2, w_2)\},$$

and choose directional utilities

$$U_{m_1}(w_1) = 0.60, \quad U_{w_1}(m_1) = 0.99, \quad U_{m_2}(w_2) = 0.99, \quad U_{w_2}(m_2) = 0.60,$$

with cross-pair values

$$U_{m_1}(w_2) = 0.65, \qquad U_{w_2}(m_1) = 0.65.$$

Take $\epsilon_U < 0.05$. Then m_1 strictly prefers w_2 to its current benchmark and w_2 accepts m_1, so the proposal is a strict bilateral improvement for the consenting agents. Before the proposal,

$$W_{\text{tot}}(\mu_t) = \frac{0.60 + 0.99}{2} + \frac{0.99 + 0.60}{2} = 1.59, \qquad W_{\text{avg}}(\mu_t) = 0.795.$$

After acceptance, the two old temporary relationships are dissolved, w_1 and m_2 are displaced, and immediately after rematching

$$\mu_{t+1} = \{(m_1, w_2)\}.$$

Hence

$$W_{\text{tot}}(\mu_{t+1}) = W_{\text{avg}}(\mu_{t+1}) = \frac{0.65 + 0.65}{2} = 0.65.$$

Both total matched-pair welfare and average mutual welfare therefore decrease. The decline is caused by the losses of the displaced partners, which are not part of the bilateral acceptance test. □

Proof of Corollary L.1.

Proof. The local regret theorems control only the loss in one-step expected proposer improvement relative to a myopic acceptance-probability oracle. Proposition L.3 constructs a feasible accepted rematching in which both consenting agents improve while matched-pair welfare decreases. Therefore local decision quality, even if asymptotically no-regret, does not by itself imply monotonicity or terminal optimality of the matched-pair welfare metric. Stability is likewise not implied because the local regret benchmark contains no stability constraint. □

M Full Decentralized Matching Procedure

Algorithm 1 expands the period-level procedure summarized in Section 3.4. It makes explicit the order of local exposure, valuation, proposal, feedback, model updating, relationship revision, and stochastic commitment.

The procedure preserves three distinct forms of locality. Candidate discovery is local through $\mathcal{C}_t(i)$; LLM valuation is triggered only when a candidate becomes locally relevant; and reciprocal-acceptance learning is agent-specific because only proposals made by i enter $\mathcal{H}_{i,t}$. No global acceptance model, shared interaction history, or market-wide preference matrix is available to the proposed agents.

N Limitations and Future Directions

N.1 Limitations

The empirical grounding of this study is intentionally specific to the Chinese marriage-market setting represented by the 2015 population sample and the choice experiment of Zhou et al. (2023). Accordingly, the validation establishes alignment with this empirical reference rather than universal behavioral validity across cultures, cohorts, or institutional settings. Extending the framework to other populations would require corresponding domain-specific demographic construction and behavioral validation.

Algorithm 1 Decentralized LLM-Agent Matching with Local Logistic-UCB

Require: Agent sets $\mathcal{M}, \mathcal{W}$; horizon T; proposal budget L; reservation utilities $\{r_i\}$; regularization $\{\lambda_i\}$; exploration coefficients $\{\beta_{i,t}\}$; utility threshold ϵ_U; commitment probabilities $\{\rho_t\}$.

Initialize all agents as single; set local histories empty and initialize local Logistic-UCB models.
for $t = 1, \dots, T$ **do**
  $\mathcal{A}_t^{\mathrm{act}} \leftarrow \{i \in \mathcal{A} : i \text{ is not permanently matched}\}$
  if $\mathcal{A}_t^{\mathrm{act}} = \varnothing$ **then**
    break
  end if
  for $a = 1, \dots, |\mathcal{A}_t^{\mathrm{act}}|$ **do**
    Sample $i \sim \mathrm{Uniform}(\mathcal{A}_t^{\mathrm{act}})$ with replacement
    Sample local exposure set $\mathcal{C}_t(i) \subseteq \mathcal{P}_t(i)$
    Obtain $U_i(j)$ for all $j \in \mathcal{C}_t(i)$ and construct $\mathcal{E}_t(i)$ using Equation 3
    $\mathcal{R}_t \leftarrow \mathcal{E}_t(i)$
    for $q = 1, \dots, L$ **do**
      if $\mathcal{R}_t = \varnothing$ **then**
        break
      end if
      For each $j \in \mathcal{R}_t$, construct $x_{j,t}^{(i)}$ and compute $I_{ij,t}$ using Equations 7 and 8
      $j^* \leftarrow \arg\max_{j \in \mathcal{R}_t} I_{ij,t}$
      Obtain $U_{j^*}(i)$ and compute $y^* \leftarrow Y_{i \to j^*,t}$ using Equation 5
      $\mathcal{H}_{i,t} \leftarrow \mathcal{H}_{i,t} \cup \{(x_{j^*,t}^{(i)}, y^*)\}$
      Refit $\widehat{\theta}_{i,t}$ by Equation 20 and update $V_{i,t}$ by Equation 23
      if $y^* = 1$ **then**
        Dissolve temporary matches involving i or j^* and return displaced partners to active search
        Form (i, j^*) as a new temporary pair
        break
      else
        $\mathcal{R}_t \leftarrow \mathcal{R}_t \setminus \{j^*\}$
      end if
    end for
  end for
  Let $\mathcal{T}_t$ denote temporary pairs after all activations
  for all $(i, j) \in \mathcal{T}_t$ **in parallel do**
    Sample $Z_{ij,t} \sim \mathrm{Bernoulli}(\rho_t)$
    if $Z_{ij,t} = 1$ **then**
      Mark (i, j) as permanent
    end if
  end for
end for
return final matching μ

The simulated market also remains deliberately stylized. The main experiments use a 50×50 population with random local exposure, a common reservation-utility specification, and probabilistic commitment. Richer interaction structures—including social networks, endogenous meeting processes, heterogeneous or adaptive reservation values, noisy or private attributes, and evolving outside options—may alter the resulting matching dynamics. These components can be incorporated within the same decentralized framework, but are outside the scope of the present experiments.

The theoretical results in Appendix L characterize local proposal-level learning and bilateral rematching; they do not guarantee terminal welfare optimality or convergence to a stable matching. Establishing endogenous bounds on dynamic misspecification and conditions linking local learning to market-level outcomes remains an open direction. Moreover, learned acceptance coefficients represent predictive associations rather than structural or causal preferences, because proposal observations are endogenously selected by the matching policy.

Finally, LLM-based behavioral modeling remains sensitive to persona construction, prompt design, model choice, and demographic representation (Anthis et al., 2025; Zhou et al., 2026; Ludwig et al., 2024). Our cross-backbone and empirical-reference validation reduces dependence on any single model realization, but does not eliminate these broader sources of behavioral-model uncertainty.

N.2 Future Directions

The framework naturally extends to other decentralized economic settings in which agents evaluate counterparties while learning whether interaction will be reciprocated, including labor-market search, hiring, team formation, housing, platform matching, and repeated buyer–seller or lender–borrower interactions. The same separation between domain-grounded behavioral valuation and local reciprocal learning can be retained while replacing personas, actions, institutions, and validation targets with domain-specific counterparts.

A second direction is to enrich the interaction structure through social networks, geography, recommender systems, endogenous communication, longer memories, adaptive reservation values, or information sharing. On the learning side, nonstationary bandits, hierarchical priors, and theoretical conditions connecting local learning to welfare or stability are natural extensions. Beyond economics, the architecture can support persistent multi-agent societies in games, where NPCs form and revise friendships, rivalries, coalitions, or trading relationships through local interaction rather than a globally scripted social graph.