

Mechanism Alignment for Socially Embedded Decision Making in LLMs

Jiixin Liu^{1*}, Yuhe Wu^{1*}, Yixuan Xie², Hanzhi Xiao¹,
Zeyu Ma¹, Wenqing Gu¹, Zeheng Xian¹, Guang Zhang¹

¹The Hong Kong University of Science and Technology (Guangzhou)

²The Hong Kong University of Science and Technology

✉: {jliu043@connect., guangzhang@}hkust-gz.edu.cn

Abstract

LLM agents are widely used in social simulation, yet fluent social reasoning does not guarantee valid decision mechanisms. We test whether models implement socially embedded utility, where actions depend on peer-weighted references rather than isolated private preference. We introduce a theory-guided benchmark from network utility theory covering utility recovery, social-comparison gating, and structural transfer across two out-of-distribution decision forms. Across open-source and proprietary models, we find a gap between cue recognition and peer aggregation, spurious social activation when comparison should be inactive, and declining accuracy as social dependence rises. We propose theory-supervised mechanism alignment trained on latent utility states. Fine-tuning yields +8.3 and +13.9 utility-recovery points on Llama-3.1-8B and Qwen3-8B, raises gating above 95%, and improves transfer to unseen structures. Multi-agent simulations show that repairing local utility shifts macro outcomes, including stronger financial amplification and more stable labor stratification, which indicates that realistic macro-level simulations require adjustments at the individual level. Data and code are available at github.com/saaranLiu/SEDM-alignment.

1 Introduction

Investment, consumption, career choice, and coordination often depend on peer behavior, network position, and positional payoffs. A credible LLM social simulator should therefore map local peer profiles and closeness weights into reference-dependent utility, not only produce plausible social language. We call this capability socially em-

bedded decision making (SEDM), following network utility models of social comparison (Gali, 1994; Langtry, 2023). This requirement is easy to miss because dominant alignment and reasoning pipelines often treat agents as preference-stable and only weakly affected by peers (Ouyang et al., 2022; Liu et al., 2023).

Recent generative-agent systems can express preferences, interact at scale, and support counterfactual analysis of social or market settings (Li et al., 2024; Manning and Horton, 2026; Horton et al., 2023; Taillandier et al., 2025). However, fluent social reasoning is not the same as correct utility implementation. A model may describe conformity while choosing from a private baseline or a single salient peer. This distinction matters because macro patterns such as cascades, conformity, and relative-wealth dynamics should arise from local incentives rather than from narrative priors, as shown in Figure 1 (Latané, 1981; Banerjee, 1992; Bikhchandani et al., 1992; DeMarzo et al., 2008; Li and Tao, 2026; Zhou et al., 2024).

These limitations motivate a central question for LLM-based social simulation:

Research Question *Can language agents implement socially embedded utility at the local level, so that macro social dynamics arise from network-conditioned incentives rather than narrative priors?*

This question is particularly important because current post-training pipelines often reward consistency, harmlessness, and resistance to manipulation across broad instruction distributions (Ouyang et al., 2022; Liu et al., 2023). Under socially embedded utility, however, adaptive responses to peer behavior are frequently normatively rational rather than signs of instability. Therefore, behaviors re-

*Equal contribution.

warded as robustness under generic instruction tuning may correspond to irrational coordination failure in graph-structured environments. A model that cannot implement socially embedded utility at the node level cannot be trusted to generate scientifically meaningful macro social behavior.

We construct a theory-guided framework that separates directional social sensitivity from recovery of the peer-weighted aggregation rule. Across small- and large-scale systems, models often infer the direction of social pressure without recovering the utility rule that determines the action. Errors increase as comparison intensity rises, and several models over-apply social aggregation when the comparison channel should be inactive. We then introduce theory-supervised mechanism alignment, which trains models to represent the utility state before action selection, and test whether this repair changes recursive asset-market and labor-sorting dynamics.

Overall, We contribute (i) a SEDM benchmark for peer-weighted utility recovery, social gating, and structural transfer, (ii) utility-state supervised alignment for 8B backbones, and (iii) multi-agent tests linking local mechanism repair to financial and labor macro dynamics.

2 Related Work

2.1 Decision Making with Social Comparison

Utility-based choice is a standard foundation for individual decision making under uncertainty (Abel, 1990; Campbell and Cochrane, 1999). Keeping up with the Joneses models extend this view by making utility depend on social reference points, so optimal behavior is shaped by both private fundamentals and peer comparisons (Gali, 1994; Bakshi and Chen, 1996; Robson, 1992). Empirical and theoretical work shows that status competition changes risk-taking, positional consumption, and portfolio decisions (Frank, 1985; DeMarzo et al., 2004; Carlsson, 2007; Roussanov, 2010; Hopkins and Kornienko, 2004; Clark et al., 2008; Cohn et al., 2014; Liu and Sun, 2023). Recent network formulations further specify how localized peer weights enter the utility function (Langtry, 2023). These micro-level mechanisms connect to aggregate outcomes in social learning, cascades, and relative-wealth models (Latané, 1981; Banerjee, 1992; Bikhchandani et al., 1992, 2024; DeMarzo et al., 2008). However, this theory has rarely been used as a diagnostic target for LLM agents. These models provide explicit

micro-level utility rules for socially embedded behavior, including peer-weighted comparison and localized network influence. In this work, we use these utility formulations as behavioral targets for evaluating language agents.

2.2 LLM-based Social Simulation

LLM agents are increasingly used to simulate social and economic systems. Macro-oriented frameworks report that agent populations can reproduce selected aggregate patterns in structured environments (Li et al., 2024). Benchmark studies also measure financial decision performance at the task level (Li et al., 2025). Work on trust, social learning control, and recommendation bias shows that outcomes are sensitive to prompt design and model tendencies (Jia et al., 2024; Arghal et al., 2026; Filandrianos et al., 2025). However, realistic dialogue and plausible aggregate patterns do not imply valid mechanism recovery (Zhou et al., 2024; Chen et al., 2025; Wu et al., 2025). This distinction is especially important in recursive environments, where small local misspecifications can accumulate into different macro equilibria.

2.3 LLM Alignment & Preference Learning

Alignment methods often begin with supervised instruction tuning, then use preference optimization to shape response behavior. RLHF combines supervised demonstrations with reward-based optimization (Ouyang et al., 2022), while DPO-style methods simplify the preference-optimization stage and raise questions about scaling, reference policies, and objective design (Rafailov et al., 2023, 2024; Amini et al., 2024; Azar et al., 2024; Liu et al., 2025b). Studies also documents risks such as reward misspecification, optimization instability, and alignment tax (Liu et al., 2023; Lin et al., 2024; Niu et al., 2025; Lambert et al., 2025). In parallel, studies on safety and values examine trade-offs in moral reasoning and multi-value steering (Lu et al., 2025; Chen and Shen, 2026; Lau et al., 2025; An and Du, 2026; Zhu et al., 2025; Wang et al., 2024). Recent work further moves toward economically grounded calibration, including risk-preference evaluation and investment-oriented alignment (Wang et al., 2025; Liu et al., 2025a; Zhang et al., 2025). Most alignment objectives therefore operate at the level of preferred responses or behavioral preferences. Much less work studies whether alignment changes the internal utility representation that generates a socially embedded decision.

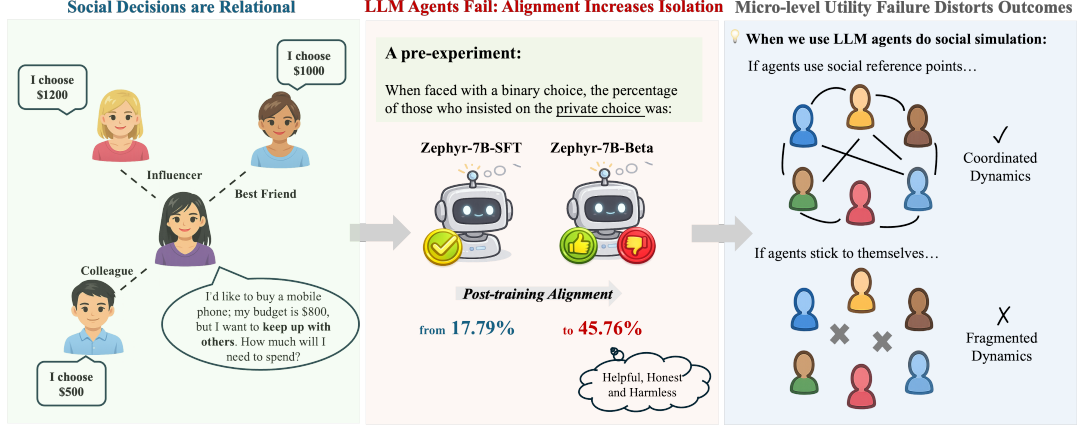


Figure 1: Overview of socially embedded decisions, alignment-induced preference for private choices over peer-weighted utility, and fragmented vs. coordinated simulation dynamics.

3 Benchmark and Preliminary Analysis

3.1 Problem Definition

We study whether language agents can recover and apply socially embedded utility functions rather than relying on narrative social heuristics. Formally, the problem is to infer how peer-conditioned information enters an agent’s decision rule, when that social reference should be activated, and whether the resulting utility structure transfers across different social environments. We operationalize socially embedded utility recovery through three benchmark families: **Utility Recovery**, **Social Gating**, and **Structural Transfer**. Structural Transfer further includes Matching Transfer and Rank Transfer settings, which preserve socially conditioned utility while changing the decision form itself. Appendix A gives the full derivations.

Utility Recovery. The core task is peer-weighted choice. For agent i , action x_i is evaluated against a local social reference formed from peer actions $\{x_j\}_j$ and closeness weights $\{g_{ij}\}_j$. Following (Langtry, 2023), we use

$$u_i = f\left(x_i - \alpha_i \sum_j g_{ij} x_j\right) - c_i x_i + b_i \alpha_i \sum_j g_{ij}, \quad (1)$$

where α_i is comparison sensitivity, $f(\cdot)$ is a concave utility kernel, c_i is marginal cost, and b_i is social value. The final term captures utility derived from social connectedness. Let

$$R_i = \sum_j g_{ij} x_j, \quad (2)$$

where R_i is the peer-weighted reference. With private baseline $F = f'^{-1}(c_i)$, the best response is

$$x_i^* = F + \alpha_i R_i. \quad (3)$$

The key challenge is therefore recovering the correct aggregation operator over socially embedded information. Utility Recovery tests whether models recover this rule rather than choosing a private baseline, a closest-peer shortcut, an equal-weight average, or another heuristic.

Social Gating. Social Gating tests whether models suppress social comparison when socially embedded utility should be inactive. In these settings, optimal behavior reduces to the private baseline without peer-conditioned adjustment. The task therefore distinguishes valid social reasoning from indiscriminate conformity or social over-activation.

Structural Transfer. Structural Transfer tests whether models generalize socially conditioned utility beyond the original peer-weighted aggregation rule while the decision form changes. Matching Transfer asks models to select partners with compatible benefit–cost ratios,

$$x_i^* = \frac{b_i}{c_i}, \quad (4)$$

and Rank Transfer asks whether absolute salary gains offset relative-rank costs,

$$x_{S,H} - x_{S,L} \geq \alpha_{2i}(\bar{x}_H - \bar{x}_L), \quad (5)$$

where $x_{S,m}$ is own salary at firm m and $\bar{x}_m$ is the firm average. Together, these settings test whether

socially embedded utility is learned as a transferable decision mechanism rather than as a surface-level narrative heuristic.

3.2 Benchmark Construction and Data Synthesis

Latent construction. The benchmark follows a two-stage pipeline: we sample structured oracle states first and render narratives afterward.

For Utility Recovery, each record draws a private baseline F , marginal-cost scalar c , peer count $N_i \in \{3, 4, 5\}$, comparison intensity α_i , peer actions $\{x_j\}$, and closeness weights $\{g_{ij}\}$, then computes $R_i = \sum_j g_{ij}x_j$ and $x_i^* = F + \alpha_i R_i$. The main grid is a $3 \times 3 \times 3$ factorial over three latent factors—comparison intensity, peer-weight dispersion, and peer-action skew—yielding 27 cells, six of which are withheld from training. Each factor takes low, mid, and high buckets whose intervals appear in Table S1: α_i controls social pull; dispersion is set by a top peer weight $g_{\max}$ with Dirichlet-split residual weights; skew sets the mean peer-to-baseline ratio $\mathbb{E}[x_j]/F$ with log-normal noise.

Scenarios use visible consumption and effort domains in different life scenarios; local peer cards verbalize x_j and g_{ij} as action levels and tie-strength cues without formal notation, rule names, or latent-cell labels. Training-side source pools contain 1,512 Utility Recovery, 560 Social Gating, and 500 records per transfer setting; the final 3,000-example alignment corpus is specified in Section 4.

Options and synthesis. For Utility Recovery and Social Gating, each MCQ contains the gold rule plus three confusable distractors sampled after numeric separability checks. Structured records are rendered into natural scenarios by DeepSeek-R1-671B under domain-specific writer prompts; the teacher receives public scene anchors and style constraints, while the structured oracle state is retained for scoring. Student-facing prompts expose only the narrative, peer descriptions, and answer interface. Detailed distractor templates, surface-form separation, and leakage controls appear in Appendix B and Appendix C.

3.3 Empirical Findings

Setup. We evaluate seven models on the core benchmark families: three Qwen3 variants spanning a 235B API model and 14B/8B Ollama deployments, DeepSeek-V4-Flash, Llama-3.1-8B via Ollama, Zephyr-7B SFT, and Zephyr-7B DPO (β).

All runs share the same structured records and code-templated options. Following the canonical mixed protocol, Utility Recovery and Social Gating use a utility-mechanism response format, while Structural Transfer and the auxiliary probe use a concise choice format. We score exact-match accuracy on the parsed choice letter and count unparsed outputs as incorrect. Table 1 reports core benchmark accuracy.

Task-level findings. Table 1 shows that recovering the underlying utility mechanism remains the central challenge. Even the strongest model, Qwen3-235B, reaches only 61.6% on Utility Recovery, far below ceiling despite the structured and fully observable setup. By contrast, the auxiliary Comparative Statics Probe averages 60.36% accuracy across the same seven models; full results appear in Appendix G. This gap suggests that models often detect the direction of social influence while failing to recover the aggregation rule that determines the action itself. In other words, directional social sensitivity does not imply valid utility recovery.

The benchmark also separates socially conditioned reasoning from indiscriminate social activation. Social Gating accuracy remains low for several open models, indicating that social comparison is frequently over-applied even when the optimal action reduces to the private baseline. Notably, DPO on Zephyr substantially improves Social Gating while leaving Utility Recovery weak, suggesting that preference-style alignment may partially suppress social over-activation without repairing the underlying utility mechanism.

Finally, structural transfer remains limited. Matching Transfer shows moderate performance for several models, especially DeepSeek-V4-Flash, but Rank Transfer remains near binary chance across nearly all systems. This asymmetry suggests that models may reuse shallow relational heuristics in familiar matching settings while failing to generalize socially conditioned utility to new decision structures such as threshold trade-offs between pay-off and rank.

Diagnostic analysis. Utility Recovery failures vary with latent conditions and distractor family. Table 2 stratifies accuracy by comparison intensity α_i , peer-weight dispersion, and peer-action skew. Accuracy falls monotonically with α_i for every model—for example, DeepSeek-V4-Flash falls from 77.6% to 48.9% and Qwen3-235B

Model	Util. Rec.	Gating	Matching	Rank
Qwen3-235B	61.57	79.52	68.06	50.00
Qwen3-14B	54.17	10.71	59.03	50.17
Qwen3-8B	55.09	22.42	57.64	48.78
DeepSeek-V4-Flash	60.89	75.46	78.15	50.00
Llama-3.1-8B	56.68	13.69	54.86	49.83
Zephyr-7B-SFT	33.73	32.14	59.03	50.17
Zephyr-7B- β	38.49	53.77	63.37	48.26

Table 1: Accuracy (%) on the core benchmark splits. Utility Recovery tests peer-weighted rule identification under Eq. 3. Social Gating tests private-baseline gating when $\alpha_i \approx 0$. Matching and Rank Transfer test Eq. 4 and Eq. 5 in binary form. Bold marks the best score in each column; tied best scores are both marked.

from 71.8% to 51.4%—consistent with stronger social pull favoring single-peer shortcuts over weighted aggregation. Dispersion shifts accuracy only slightly, whereas skew separates top systems: low-skew items are hardest when peer actions cluster near F , and high-skew items raise Qwen3-14B and DeepSeek-V4-Flash by roughly 15 pp.

Figure 2 shows how mistakes distribute across distractor rules among incorrect answers. Errors are not dominated by a single shortcut. Among stronger open models, Qwen3-235B and DeepSeek-V4-Flash mainly confuse peer-weighted aggregation with closest-peer mimicry and top-anchor substitution, at 35.11% and 33.22% for Qwen3-235B and 29.66% and 30.56% for DeepSeek, while pure-private mistakes stay below 8%. Mid-scale Qwen3-14B and Qwen3-8B instead favor equal-weight mixing, at 41.24% and 34.90%, together with median-anchor and uniform-average heuristics. Llama-3.1-8B shows a similar mixing-heavy profile, with 35.42% equal mix and 14.81% uniform average. Among weaker systems, Zephyr-7B-SFT spreads errors across several shortcuts, with pure-private confusions at 16.37%; after DPO (β), Utility Recovery accuracy rises modestly from 33.73% to 38.49%, but pure-private becomes the largest error share at 23.44%, indicating that preference-based alignment can push MCQ mistakes toward private-baseline collapse even when social aggregation remains the target. Overall, the latent gradients and heterogeneous error profiles show that failures reflect misspecified utility operators and motivate the mechanism alignment in Section 4.

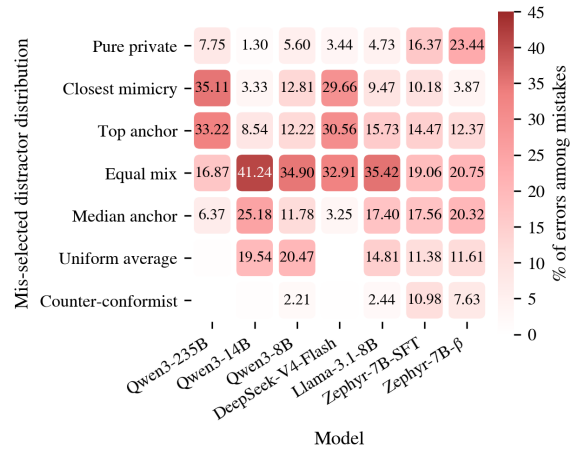


Figure 2: Mis-selected distractor distribution on Utility Recovery. Cell values are the share (%) of each wrong rule among all incorrect answers on gold peer-weighted items.

4 Methodology: Theory-Supervised Mechanism Alignment

Theory-supervised mechanism alignment trains the model to recover the utility state behind a decision, rather than only imitate a final answer style. We instantiate it on Llama-3.1-8B-Instruct and Qwen3-8B with LoRA adapters (Grattafiori et al., 2024; Hu et al., 2022), keeping the base parameters fixed. Let s_i denote a natural-language scenario and z_i the structured utility state derived from the oracle variables. For social records, z_i includes the private baseline, peer-weighted reference, comparison intensity, and commitment implied by Eq. 3. For social gating records, the comparison channel is inactive and the target reduces to $x_i^* \approx F_i$. Alignment supervises the following process beyond only the final choice:

$$s_i \mapsto z_i \mapsto x_i^* \quad (6)$$

Model	α_i bucket			Dispersion bucket			Skew bucket		
	L	M	H	L	M	H	L	M	H
Qwen3-235B	71.83	61.51	51.39	65.08	59.13	60.52	55.95	64.48	64.29
Qwen3-14B	69.84	49.60	43.06	55.56	56.35	50.60	46.23	54.96	61.31
Qwen3-8B	66.87	50.79	47.62	57.14	55.16	52.98	45.04	59.13	61.11
DeepSeek-V4-Flash	77.59	55.60	48.93	63.20	57.72	61.80	49.18	63.77	70.24
Llama-3.1-8B	62.30	56.75	50.99	55.36	61.31	53.37	49.60	57.54	62.90
Zephyr-7B-SFT	38.69	31.55	30.95	34.52	34.52	32.14	34.13	32.74	34.33
Zephyr-7B- β	43.06	35.71	36.71	39.29	39.88	36.31	34.33	40.28	40.87

Table 2: Utility Recovery accuracy (%) by latent bucket (504 items per bucket). Model order matches Table 1; bold marks the best score in each bucket column.

The alignment corpus contains 3,000 training examples drawn from the in-domain training split. Each example presents only the scenario and asks for an open-ended commitment decision; it does not expose MCQ options, rule labels, or answer letters. The assistant target contains a structured utility trace and a final commitment. Appendix B details the split construction, held-out latent cells, surface-form separation, and leakage controls. Given chat-format pairs (u_i, y_i) , we minimize the autoregressive negative log-likelihood over assistant tokens:

$$\mathcal{L}_{\text{SFT}}(\phi) = - \sum_i \sum_{t=1}^{|y_i|} \log p_{\theta_0 + \Delta_\phi}(y_{i,t} \mid u_i, y_{i,<t}), \quad (7)$$

where θ_0 is the frozen backbone and Δ_ϕ denotes the adapter parameters. The main alignment setting excludes Matching and Rank Transfer supervision, so transfer to Eq. 4 and Eq. 5 tests whether the learned utility-state representation generalizes beyond the supervised interface.

5 Micro-Level Evaluation and Analysis

We fine-tune Llama-3.1-8B-Instruct and Qwen3-8B with theory-supervised utility-state supervision and evaluate whether alignment repairs node-level decision rules. All tables compare the base checkpoint (Pre) with the aligned model (Aligned). This section focuses on static micro-level validity before the dynamic simulations in Section 6.

5.1 Main Results

Table 3 reports in-domain and OOD accuracy for both 8B backbones.

Utility-state supervision improves both small-parameter backbones across all evaluation splits.

Llama-3.1-8B improves Utility Recovery by 8.3 percentage points and Qwen3-8B by 13.9 points; gains appear on held-out latent cells of +4.1 and +10.7 pp, suggesting that models learn part of the latent aggregation structure rather than only memorizing observed cells. The largest in-domain improvements are on Social Gating, where accuracy rises from 13.7% to 95.2% for Llama-3.1-8B and from 22.4% to 98.0% for Qwen3-8B: both base models over-apply social aggregation when the comparison channel should be inactive, and alignment teaches them to gate it correctly. Beyond aggregate accuracy, alignment also reshapes the *type* of remaining Utility Recovery errors: Appendix F, Table S9, and Figure S1 show that both backbones shift away from equal-weight and uniform shortcuts toward private-baseline confusions, with a stronger private-baseline spike for Llama-3.1-8B—consistent with the Social Gating gains above. Structural transfer is also positive despite the alignment corpus excluding Matching and Rank Transfer supervision entirely. Matching Transfer improves by 14.7 pp for Llama-3.1-8B and 10.1 pp for Qwen3-8B; Rank Transfer improves by 9.4 and 10.7 pp respectively. These gains suggest that utility-state supervision instills a structural representation of the aggregation mechanism that transfers to equation-form shifts beyond the supervised Utility Recovery interface.

5.2 Representation-Level Analysis

To verify that accuracy gains reflect representational change rather than output-format tuning, we train linear probes on last-token hidden states $h_l^{(i)} \in \mathbb{R}^{4096}$ from each layer of Llama-3.1-8B for all 1,512 Utility Recovery records. The probes predict the discretized α_i bucket and the oracle quantities $\log R_i$ and $\log \delta x = \log(\alpha_i R_i)$ with 5-fold

Model	Util. Rec.	Held-in	Held-out	Gating	Matching	Rank
Llama-3.1-8B (Pre)	56.68	56.63	56.85	13.69	54.86	49.83
Llama-3.1-8B (Aligned)	64.95	66.07	61.01	95.24	69.62	59.20
Δ	+8.27	+9.44	+4.16	+81.55	+14.76	+9.37
Qwen3-8B (Pre)	55.09	55.02	55.36	22.42	57.64	48.78
Qwen3-8B (Aligned)	68.98	69.81	66.07	98.02	67.71	59.55
Δ	+13.89	+14.79	+10.71	+75.60	+10.07	+10.77

Table 3: Alignment results for two 8B backbones. Accuracies (%) are means over five random evaluation seeds; standard deviations for evaluation, ablation, and simulation are in Appendix F. Utility Recovery uses $N=1,512$ with 21 held-in cells and 6 held-out cells; Social Gating uses $N=504$; transfer splits use $N=576$ each. Matching and Rank Transfer use equation forms not seen in alignment training, namely Eq. 4 and Eq. 5. Δ in percentage points is Aligned minus Pre.

cross-validation; Appendix D gives details.

Figure 3 shows that pre-alignment α_i classification stays near chance, whereas the aligned model exceeds chance from layer 8 onward. For regression, the aligned model closes the $\log \delta x$ gap at layer 31 to $\Delta=+0.005$ even while the $\log R_i$ gap remains negative. This dissociation is consistent with a shift from passive numerical storage toward computation of the social adjustment term.

5.3 Robustness Analysis

Table 3 reports seed-averaged accuracies; across five random seeds, Pre-to-Aligned gains are stable and standard deviations remain small on every core split, ablation, and simulation metric, as detailed in Appendix F. Alignment also improves hard Utility Recovery slices. High- α_i accuracy rises from 51.0% to 61.9% for Llama-3.1-8B and from 47.6% to 68.5% for Qwen3-8B; low-skew accuracy rises from 49.6% to 62.9% and from 45.0% to 67.9%, respectively, as reported in Table S8.

5.4 Ablation Study

Table 4 isolates which supervision component carries the repair. Full SFT uses a utility state, mechanism narrative, and final commitment; answer-only keeps only the commitment, while w/o-latent removes the explicit utility block but keeps the narrative and commitment. Both ablations retain some in-domain benefit but lose the full profile: answer-only drops on Social Gating and Matching Transfer, and w/o-latent drops sharply on Matching Transfer. This suggests that final-answer imitation and natural-language rationales teach parts of the interface, but explicit utility-state supervision is needed for the strongest transferable repair.

Training target	Util.	Gating	Match	Rank
Full SFT	64.95	95.24	69.62	59.20
Answer-only	58.80	72.02	44.27	55.90
w/o latent	66.14	83.13	35.59	52.08

Table 4: Llama-3.1-8B ablation accuracy (%) on Utility Recovery, Social Gating, Matching Transfer, and Rank Transfer.

6 Macro-Level Social Dynamics

Static benchmark accuracy does not guarantee valid macro-social dynamics under recursive interaction (Epstein, 2012; Li and Tao, 2026; Ashery et al., 2025). We therefore test whether local decision-rule errors alter aggregate outcomes in two environments: a financial coordination setting based on Eq. 3 and a labor-stratification setting based on Eq. 5. Each environment compares the base LLM and the aligned LLM under matched prompts and seeds. We report case-specific macro metrics in Table 5; Appendix E gives implementation details.

In the financial environment, agents adjust asset exposure through endogenous prices; peer-weighted targets create a coordination channel in which joint demand raises both prices and future references (DeMarzo et al., 2008). Private choice dampens this channel, producing a price peak of 2.00 and terminal wealth Gini of 0.093. The base LLM generates a larger bubble with a price peak of 14.30, while alignment raises it further to 19.38 and return volatility from 0.329 to 0.424, closing much of the gap to the oracle peak of 22.66. Figure 5 shows the corresponding time paths: after the Hub Pump phase at $t=5$, which is an exogenous demand shock centered on highly connected agents,

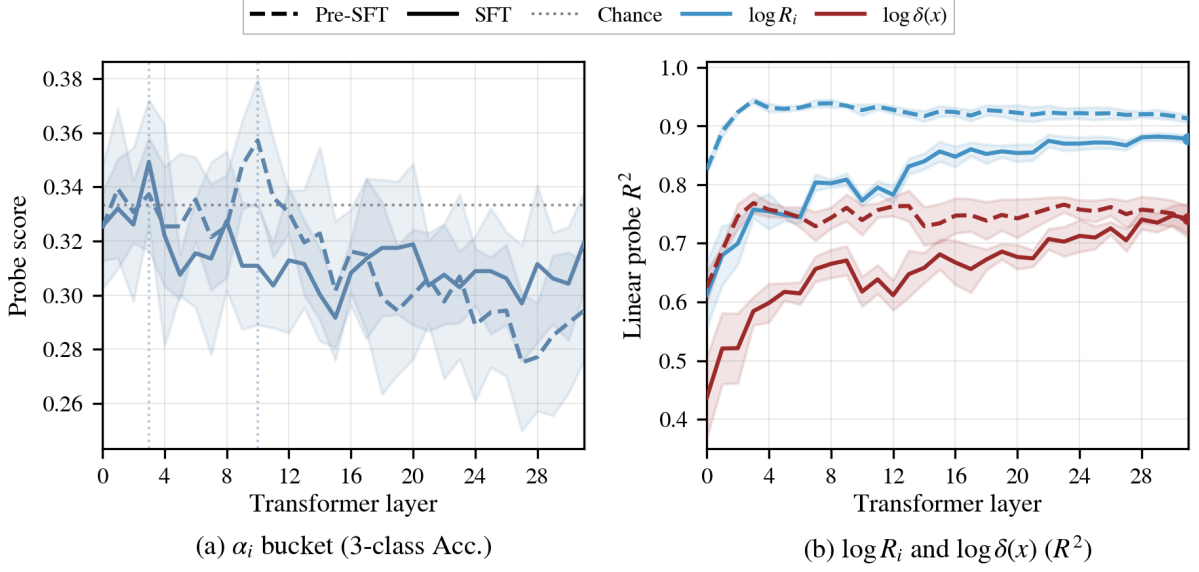


Figure 3: Latent utility probing across all 32 transformer layers of Llama-3.1-8B before and after alignment; shaded bands show ± 1 s.d. under 5-fold CV. **(a)** Three-class probe accuracy for comparison-intensity bucket α_i ; the dotted line marks chance at 33.3%. **(b)** Ridge-regression R^2 for $\log R_i$ in blue and $\log \delta(x) = \log(\alpha_i R_i)$ in red; solid curves are Pre and dashed curves are Aligned.

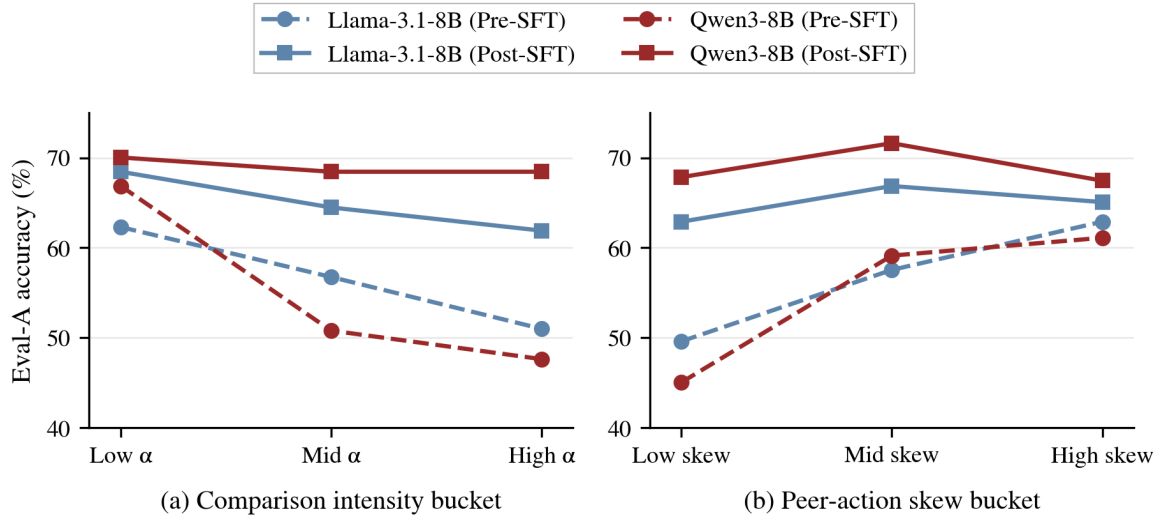


Figure 4: Utility Recovery accuracy across latent buckets for the two 8B backbones before and after theory-supervised alignment. **(a)** Comparison-intensity (α_i) buckets. **(b)** Peer-action skew buckets. Dashed lines: Pre; solid lines: Aligned; color denotes backbone.

the aligned policy reaches a higher peak before the crash and leaves wealth Gini more elevated.

In the labor environment, workers sort between elite and regional tiers while pool means update with membership. Eq. 5 trades off own pay against relative standing, so each sorting decision reshapes future incentives (Schelling, 1971). The private baseline collapses to the absolute-pay option, placing all agents in H and producing a pool fill gap

of 0.800. The base LLM still places 69% of agents in H . The aligned policy lowers the elite share to 55%, reduces the pool fill gap from 0.397 to 0.215, and moves sorting skew from -0.158 to -0.021, cutting distance to the oracle on elite share and pool fill gap. Thus, the same local repair increases co-ordinated exposure in the asset market but reduces elite-tier crowding in the labor market.

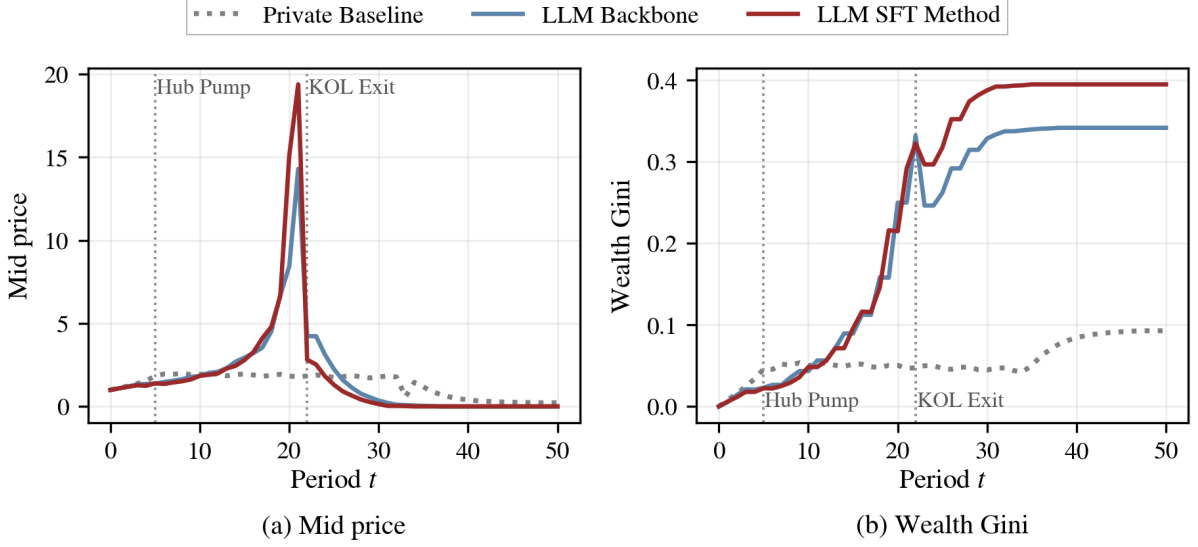


Figure 5: Financial amplification trajectories for $N=100$ and $T=50$. **(a)** Mid price for the LLM Backbone and LLM SFT Method policies; the vertical marker at $t=5$ denotes the Hub Pump phase. **(b)** Cross-agent wealth Gini. Both panels compare Private Baseline, LLM Backbone, and LLM SFT Method under matched prompts.

	Oracle	Priv.	Base	Aligned
<i>(a) Financial amplification</i>				
Price peak	22.66	2.00	14.30	19.38
Return volatility	0.59	0.23	0.33	0.42
Wealth Gini	0.41	0.09	0.34	0.39
<i>(b) Stratification</i>				
Elite tier share	0.58	1.00	0.69	0.55
Pool fill gap	0.25	0.80	0.40	0.22
Sorting skew	-0.31	0.00	-0.16	-0.02

Table 5: Macro outcomes for oracle, private baseline, base LLM, and aligned LLM policies. **Priv.** denotes the private baseline policy. Financial amplification uses $N=100$, $T=50$; stratification uses $N=100$, $T=20$. Return volatility is mid-price log-return volatility; pool fill gap is $\text{fill}_H - \text{fill}_L$; sorting skew summarizes tier-composition imbalance.

7 Conclusion

This paper studies whether LLM agents implement socially embedded utility or merely produce socially plausible reasoning. We formulate SEDM as a mechanism-recovery problem under reference-dependent network utility and build a benchmark that separates directional social sensitivity from correct peer-weighted aggregation. The results show that current models often detect social pressure without recovering the utility rule that should govern action selection, especially when comparison intensity is high or when the comparison chan-

nel should be inactive. We then introduce theory-supervised mechanism alignment, which trains models on structured utility states beyond response style alone. Across Llama-3.1-8B and Qwen3-8B, this supervision improves rule recovery, social gating, and structural OOD transfer. Dynamic simulations further show that local utility repair changes aggregate behavior in mechanism-specific directions, producing stronger financial amplification and more balanced labor stratification.

Limitations

The benchmark adopts one reference-dependent utility family; bargaining, institutional constraints, and coalition formation require additional theory modules. Alignment is evaluated on two 8B backbones only, and broader model families remain to be tested. Simulations use few paper-scale runs for illustration rather than population-level estimates, and OOD tasks do not cover all social environments in agent-based simulation.

Acknowledgments

Guang Zhang acknowledges support from the 2024 Guangzhou Municipal Education Bureau Scientific Research Projects for the Higher-educational Institution (Project No. 2024312032).

References

- Andrew B Abel. 1990. Asset prices under habit formation and catching up with the joneses. *The American Economic Review*, 80(2):38–42.
- Afra Amini, Tim Vieira, and Ryan Cotterell. 2024. Direct preference optimization with an offset. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 9954–9972.
- Zhiyu An and Wan Du. 2026. Moralreason: Generalizable moral decision alignment for llm agents using reasoning-level reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 37232–37239.
- Raghu Arghal, Kevin He, Shirin Saeedi Bidokhti, and Saswati Sarkar. 2026. Steering the herd: A framework for LLM-based control of social learning. In *International Conference on Learning Representations (ICLR)*.
- Ariel Flint Ashery, Luca Maria Aiello, and Andrea Baronchelli. 2025. Emergent social conventions and collective bias in llm populations. *Science Advances*, 11(20):eadu9368.
- Mohammad Gheshlaghi Azar, Zhaohan Daniel Guo, Bilal Piot, Remi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. 2024. A general theoretical paradigm to understand learning from human preferences. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 4447–4455. PMLR.
- Gurdip S Bakshi and Zhiwu Chen. 1996. The spirit of capitalism and stock-market prices. *The American Economic Review*, pages 133–157.
- Abhijit V Banerjee. 1992. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.
- Sushil Bikhchandani, David Hirshleifer, Omer Tamuz, and Ivo Welch. 2024. Information cascades and social learning. *Journal of Economic Literature*, 62(3):1040–1093.
- Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026.
- John Y Campbell and John H Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205–251.
- Mikael Carlsson. 2007. Investment and uncertainty: a theory-based empirical approach. *Oxford Bulletin of Economics and Statistics*, 69(5):603–617.
- Chien Hung Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2025. Self-augmented preference alignment for sycophancy reduction in llms. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 12390–12402.
- Jiajun Chen and Hua Shen. 2026. Value alignment tax: Measuring value trade-offs in llm alignment. *arXiv preprint arXiv:2602.12134*.
- Andrew E Clark, Paul Frijters, and Michael A Shields. 2008. Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1):95–144.
- Alain Cohn, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. 2014. Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association*, 12(4):877–898.
- Peter M DeMarzo, Ron Kaniel, and Ilan Kremer. 2004. Diversification as a public good: Community effects in portfolio choice. *The Journal of Finance*, 59(4):1677–1716.
- Peter M DeMarzo, Ron Kaniel, and Ilan Kremer. 2008. Relative wealth concerns and financial bubbles. *The Review of Financial Studies*, 21(1):19–50.
- Joshua M Epstein. 2012. *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.
- Giorgos Filandrianos, Angeliki Dimitriou, Maria Lymperaio, Konstantinos Thomas, and Giorgos Stamou. 2025. Bias beware: The impact of cognitive biases on llm-driven product recommendations. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 22408–22437.
- Robert H Frank. 1985. The demand for unobservable and other nonpositional goods. *The American Economic Review*, 75(1):101–116.
- Jordi Gali. 1994. Keeping up with the joneses: Consumption externalities, portfolio choice, and asset prices. *Journal of Money, Credit and Banking*, 26(1):1–8.
- Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Ed Hopkins and Tatiana Kornienko. 2004. Running to keep in the same place: Consumer choice as a game of status. *American Economic Review*, 94(4):1085–1107.
- John J Horton, Apostolos Filippas, and Benjamin S Manning. 2023. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research.

- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, and 1 others. 2022. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3.
- Feiran Jia, Ziyu Ye, Shiyang Lai, Kai Shu, Jindong Gu, Adel Bibi, Ziniu Hu, David Jurgens, James Evans, Philip H Torr, and 1 others. 2024. Can large language model agents simulate human trust behavior? *Advances in Neural Information Processing Systems*, 37:15674–15729.
- Nathan Lambert, Valentina Pyatkin, Jacob Morrison, LJ Miranda, Bill Yuchen Lin, Khyathi Chandu, Nouha Dziri, Sachin Kumar, Tom Zick, Yejin Choi, and 1 others. 2025. Rewardbench: Evaluating reward models for language modeling. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 1755–1797.
- Alastair Langtry. 2023. Keeping up with “the joneses”: Reference-dependent choice with social comparisons. *American Economic Journal: Microeconomics*, 15(3):474–500.
- Bibb Latané. 1981. The psychology of social impact. *American psychologist*, 36(4):343.
- Gabriel Rongyang Lau, Wei Yan Low, Seow Min Koh, and Andree Hartanto. 2025. Evaluating ai alignment in eleven llms through output-based analysis and human benchmarking. *arXiv preprint arXiv:2506.12617*.
- Haohang Li, Yupeng Cao, Yangyang Yu, Shashidhar Reddy Javaji, Zhiyang Deng, Yueru He, Yuechen Jiang, Zining Zhu, Kp Subbalakshmi, Jimin Huang, and 1 others. 2025. Investorbench: A benchmark for financial decision-making tasks with llm-based agent. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL), Volume 1: Long Papers*, pages 2509–2525.
- Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. 2024. Econagent: large language model-empowered agents for simulating macroeconomic activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL), Volume 1: Long Papers*, pages 15523–15536.
- Yiming Li and Dacheng Tao. 2026. Position: Ai agents are not (yet) a panacea for social simulation. *arXiv preprint arXiv:2603.00113*.
- Yong Lin, Hangyu Lin, Wei Xiong, Shizhe Diao, Jianmeng Liu, Jipeng Zhang, Rui Pan, Haoxiang Wang, Wenbin Hu, Hanning Zhang, and 1 others. 2024. Mitigating the alignment tax of rlhf. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 580–606.
- Jiixin Liu, Yixuan Tang, Yi Yang, and Kar Yan Tam. 2025a. Evaluating and aligning human economic risk preferences in llms. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 18185–18199.
- Qi Liu and Bo Sun. 2023. Relative wealth concerns, executive compensation, and managerial risk-taking. *American Economic Journal: Microeconomics*, 15(2):568–598.
- Tianqi Liu, Yao Zhao, Rishabh Joshi, Misha Khalman, Mohammad Saleh, Peter J Liu, and Jialu Liu. 2023. Statistical rejection sampling improves preference optimization. *arXiv preprint arXiv:2309.06657*.
- Yixin Liu, Pengfei Liu, and Arman Cohan. 2025b. Understanding reference policies in direct preference optimization. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 8022–8037.
- Haoran Lu, Luyang Fang, Ruidong Zhang, Xinliang Li, Jiazhang Cai, Huimin Cheng, Lin Tang, Ziyu Liu, Zeliang Sun, Tao Wang, and 1 others. 2025. Alignment and safety in large language models: Safety mechanisms, training paradigms, and emerging challenges. *arXiv preprint arXiv:2507.19672*.
- Benjamin S Manning and John J Horton. 2026. General social agents. Technical report, National Bureau of Economic Research.
- Yifan Niu, Han Xiao, Dongyi Liu, Nuo Chen, and Jia Li. 2025. Mitigating the safety alignment tax with null-space constrained policy optimization. *arXiv preprint arXiv:2512.11391*.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.
- Rafael Rafailov, Yaswanth Chittapu, Ryan Park, Harshit Sushil Sikchi, Joey Hejna, Brad Knox, Chelsea Finn, and Scott Niekum. 2024. Scaling laws for reward model overoptimization in direct alignment algorithms. *Advances in Neural Information Processing Systems*, 37:126207–126242.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36:53728–53741.
- Arthur J Robson. 1992. Status, the distribution of wealth, private and social attitudes to risk. *Econometrica: Journal of the Econometric Society*, pages 837–857.
- Nikolai Roussanov. 2010. Diversification and its discontents: Idiosyncratic and entrepreneurial risk in the quest for social status. *The Journal of Finance*, 65(5):1755–1788.
- Thomas C Schelling. 1971. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Patrick Taillandier, Jean Daniel Zucker, Arnaud Grignard, Benoit Gaudou, Nghi Quang Huynh, and Alexis Drogoul. 2025. Integrating llm in agent-based social simulation: Opportunities and challenges. *arXiv preprint arXiv:2507.19364*.

Huisheng Wang, Zhuoshi Pan, Hangjing Zhang, Mingxiao Liu, Hanqing Gao, and H Vicky Zhao. 2025. Investalign: Overcoming data scarcity in aligning large language models with investor decision-making processes under herd behavior. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL), Volume 1: Long Papers*, pages 10021–10052.

Xinran Wang, Qi Le, Ammar Ahmed, Enmao Diao, Yi Zhou, Nathalie Baracaldo, Jie Ding, and Ali Anwar. 2024. Map: Multi-human-value alignment palette. *arXiv preprint arXiv:2410.19198*.

Yue Wu, Zhiqing Sun, Huizhuo Yuan, Kaixuan Ji, Yiming Yang, and Quanquan Gu. 2025. [Self-play preference optimization for language model alignment](#). In *International Conference on Learning Representations (ICLR)*.

Lijun Zhang, Lin Li, Yajie Qi, Huizhong Song, Yaodong Yang, Jun Wang, and Wei Wei. 2025. Risk-aware direct preference optimization under nested risk measure. *arXiv preprint arXiv:2505.20359*.

Xuhui Zhou, Zhe Su, Tiwalayo Eisape, Hyunwoo Kim, and Maarten Sap. 2024. Is this the real life? is this just fantasy? the misleading success of simulating social interactions with llms. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 21692–21714.

Jian-Qiao Zhu, Haijiang Yan, and Thomas L Griffiths. 2025. Steering risk preferences in large language models by aligning behavioral and neural representations. *arXiv preprint arXiv:2505.11615*.

A Model Derivations

This appendix records the closed-form rules used by the benchmark oracle. The main peer-weighted rule follows (Langtry, 2023), who studies reference-dependent utility with network-determined reference points. Our benchmark instantiates that theory in three decision interfaces: action choice (Utility Recovery), partner matching (Matching Transfer), and firm sorting (Rank Transfer).

A.1 Utility Recovery

Setup. For agent i , Eq. 1 in the main text is the Langtry utility

$$u_i(x_i, x_{-i}) = f\left(x_i - \alpha_i \sum_j g_{ij} x_j\right) - c_i x_i + b_i \alpha_i \sum_j g_{ij}, \quad (8)$$

where f is strictly increasing and concave, $c_i > 0$ is marginal cost, $b_i > 0$ is the benefit of social links, and $\{g_{ij}\}_j$ are closeness weights with $\sum_j g_{ij} = 1$. Write the peer-weighted reference

$$R_i = \sum_j g_{ij} x_j \quad (9)$$

and the reference-adjusted action $z_i = x_i - \alpha_i R_i$. We assume $f'(0) > c_i > f'(\infty)$ so that the no-comparison optimum is interior.

First-order condition. The link-value term $b_i \alpha_i \sum_j g_{ij}$ is independent of x_i , so

$$\frac{\partial u_i}{\partial x_i} = f'(x_i - \alpha_i R_i) - c_i. \quad (10)$$

At an interior optimum, $\partial u_i / \partial x_i = 0$, hence

$$f'(z_i) = c_i, \quad z_i = f'^{-1}(c_i) \equiv F_i, \quad (11)$$

where F_i is the private baseline (Langtry’s $f'^{-1}(c)$). Because f is strictly concave, u_i is strictly concave in x_i for fixed (x_j, g_{ij}) , so this first-order condition is sufficient.

Best response. Substituting $z_i = x_i - \alpha_i R_i$ yields

$$x_i - \alpha_i R_i = F_i, \quad (12)$$

and therefore

$$x_i^* = F_i + \alpha_i R_i, \quad (13)$$

which is Eq. 3 with $F \equiv F_i$. This is the scalar best-response function in Langtry’s Eq. (3), restricted to a single decision stage with exogenous peer actions $\{x_j\}$. When $\alpha_i \rightarrow 0$, Eq. 13 collapses to $x_i^* = F_i$, which is the gold rule for Social Gating. Distractors in Utility Recovery perturb one component of Eq. 13 (unweighted mean, closest peer, private baseline only, or wrong-signed social term).

A.2 Matching Transfer

From consumption choice to tie intensity. Matching Transfer keeps the Langtry preference structure but changes the choice variable to partner selection. Each agent i chooses tie intensity $x_i \geq 0$ with the same kernel as Eq. 8, now interpreted as effort or time invested in a relationship. The first-order condition again implies that any interior best response satisfies $f'(x_i - \alpha_i R_i) = c_i$.

Pairwise stability and preferred intensity. Following Langtry’s endogenous-network analysis (Proposition 4), consider a pairwise-stable link between i and j . A necessary condition for i to keep the link is that the marginal value of the connection is nonnegative. Using Langtry’s Lemma on link derivatives, this requires $-x_j^* c_i + b_i \geq 0$ for all stable links, and symmetrically for j . If $b_i \neq b_j$, the higher- b agent wants to sever the link, so stable friendships can only form between agents with common b . Within a homogeneous class, if $x_i^* > b_i/c_i$, agent j would prefer to cut the link because $-x_i^* c_j + b_j < 0$. Hence every stable mutual tie satisfies

$$x_i^* = \frac{b_i}{c_i}, \quad (14)$$

which is Eq. 4. The benchmark exposes semantic benefit and cost tiers rather than numeric (b_i, c_i) , but the oracle uses the same ratio logic.

Partner-selection rule. Given focal agent i and candidates A, B , Matching Transfer asks which candidate has the preferred intensity closest to b_i/c_i . The oracle therefore chooses the smaller ratio distance,

$$\left| \frac{b_i}{c_i} - \frac{b_A}{c_A} \right| < \left| \frac{b_i}{c_i} - \frac{b_B}{c_B} \right|, \quad (15)$$

equivalently, the candidate whose b/c ratio is nearest to the protagonist’s. This is the discrete analogue of matching on compatible socially embedded intensities when the continuous action x_i is not directly chosen in the narrative.

A.3 Rank Transfer

Firm choice setup. Rank Transfer follows Langtry’s labour-market extension (Section 4), where an agent compares herself to coworkers inside a firm as well as to friends outside it. Let $m \in \{H, L\}$ index firms, $x_{S,m}$ denote i ’s own wage at firm m , and $\bar{x}_m$ the firm average coworker wage.

Friend-side references are held fixed when i contemplates a job move, so only the coworker channel varies across m . With coworker-comparison weight $\alpha_{2i} \geq 0$, period utility at firm m can be written as

$$u_{i,m} = f(x_{S,m} - \alpha_{2i} \bar{x}_m) - c_i x_{S,m} + \text{terms independent of } m, \quad (16)$$

where the omitted terms capture friend comparisons that do not change with the firm choice in a short-run sorting problem.

Utility ranking across firms. Agent i prefers H to L when $u_{i,H} \geq u_{i,L}$, or

$$\begin{aligned} f(x_{S,H} - \alpha_{2i} \bar{x}_H) - c_i x_{S,H} \\ \geq f(x_{S,L} - \alpha_{2i} \bar{x}_L) - c_i x_{S,L}. \end{aligned} \quad (17)$$

Define the own-salary gain and average-pay gap

$$\Delta_S = x_{S,H} - x_{S,L}, \quad \Delta_{\bar{x}} = \bar{x}_H - \bar{x}_L. \quad (18)$$

A sufficient condition. When f is linear on the relevant wage range, $f(y) = \kappa y$, Eq. 17 becomes $\kappa(\Delta_S - \alpha_{2i} \Delta_{\bar{x}}) \geq c_i \Delta_S$. If the marginal gain-loss term equals the marginal cost ($\kappa = c_i$, the condition Langtry imposes at an interior optimum), this inequality simplifies to $\Delta_S \geq \alpha_{2i} \Delta_{\bar{x}}$. More generally, rearranging Eq. 17 and using the same reference-point logic as in Langtry’s proof of Proposition 6 yields the cutoff rule used by the oracle,

$$\Delta_S \geq \alpha_{2i} \Delta_{\bar{x}}. \quad (19)$$

Equivalently,

$$x_{S,H} - x_{S,L} \geq \alpha_{2i} (\bar{x}_H - \bar{x}_L), \quad (20)$$

which is Eq. 5; choose L when the inequality is reversed. Intuitively, a higher absolute wage at H must compensate for falling in rank relative to higher-paid coworkers; stronger α_{2i} raises the relative-rank cost and makes the low-average firm L more attractive (the “big fish in a small pond” logic in (Langtry, 2023)). The benchmark samples wages and firm averages so that one firm strictly dominates under Eq. 20, yielding a unique gold label.

B Dataset Construction and Split Control

Structured state generation. Each example is generated from a structured oracle state before

it is rendered as natural language. For the main peer-weighted task, the state includes a private baseline F , comparison intensity α_i , peer actions $\{x_j\}$, closeness weights $\{g_{ij}\}$, and the oracle target $x_i^* = F + \alpha_i \sum_j g_{ij} x_j$. The main split uses a $3 \times 3 \times 3$ grid over comparison intensity, peer-weight dispersion, and peer-action skew. Utility Recovery contains 1,512 items over all 27 cells, with 504 items per marginal bucket for each latent factor. Social Gating contains 504 items in which the comparison channel is inactive. Comparative Statics Probe contains 280 paired perturbation units. Matching Transfer and Rank Transfer each contain 576 binary-choice items. Table S1 summarizes the latent variables used in oracle construction.

The table separates latent variables by their role in the oracle computation. F anchors the private action level, α_i controls how strongly social information enters the choice, g_{ij} specifies which peers matter most, and x_j supplies the visible peer actions to be aggregated. The scenario writer converts these quantities into natural cues such as baseline budgets, effort levels, peer commitments, and tie-strength descriptions; the scoring code continues to use the structured state rather than the generated prose.

Relationship-cue rendering and quality control. Peer closeness weights are exposed to the teacher only as coarse cue labels, not as numeric g_{ij} values (Table S2).

DeepSeek-R1-671B renders each structured record into a 140–220 word scenario with domain-specific writer prompts (Appendix I). Failed anchor or leakage checks trigger retry; the structured oracle, not the prose, is used for scoring.

Training and held-out cells. The alignment corpus contains 3,000 open-ended commitment examples sampled from the in-domain construction process. Six latent cells from the main $3 \times 3 \times 3$ grid are reserved as held-out cells for Utility Recovery analysis, while the remaining 21 cells form the held-in portion. Held-out assignment is made at the latent-cell level, not at the surface-text level, so the held-out evaluation checks generalization across combinations of α_i , dispersion, and skew rather than memorization of individual scenarios.

Surface-form separation. Training and evaluation examples use disjoint surface vocabulary pools for names, occupations, item domains, and phrases where possible. The student-facing prompts

expose scenario cues and answer options but do not expose record identifiers, latent-cell labels, or oracle rule names. The structured oracle state is retained only for scoring and for producing the supervised trace in the training corpus.

Leakage prevention. Training examples are open-ended commitment tasks, whereas Utility Recovery and Social Gating use multiple-choice rule identification and the Structural Transfer splits use binary choices. The training target does not contain option letters, MCQ distractors, or rule labels. Matching Transfer and Rank Transfer are excluded from the main alignment corpus. Automated consistency checks verify that each supervised trace matches the oracle state and that evaluation labels are not copied into the training target format.

C MCQ Option Design

Rule pool. The benchmark uses eight candidate rules. In Utility Recovery, rule A is correct; in Social Gating, rule D is correct.

$$A: x = F + \alpha \sum_j g_{ij} x_j$$

$$B: x = F + \alpha x_{\text{closest}}$$

$$C: x = F + \alpha \text{mean}(x_j)$$

$$D: x = F$$

$$E: x = x_{\text{closest}}$$

$$F: x = F + \alpha \text{median}(x_j)$$

$$G: x = F - \alpha \sum_j g_{ij} x_j$$

$$H: x = 0.5F + 0.5 \sum_j g_{ij} x_j$$

Design rationale. Distractors cover the main shortcut families: B and E (closest-peer anchoring), C (equal-weight aggregation), D (private-baseline collapse), F (median-anchor substitution), G (sign inversion under social influence), H (alpha-free linear mixing).

Filtering and balancing. For each item, three distractors are sampled with global frequency balancing. Candidates pass a numeric separability check: $|x_k - x_{\text{gold}}| / \max(F, 10^{-9}) \geq \tau$, with $\tau = 0.10$ for Utility Recovery and $\tau = 0.08$ for Social Gating. Pairwise separability among the three selected distractors uses $\tau_{\text{pair}} = 0.05$ (Utility Recovery) and 0.04 (Social Gating). Option letter assignment is randomized per record to reduce position bias.

Quantity	Sampling role
F, c, N_i	Scene-level private baseline and cost scale; $N_i \in \{3, 4, 5\}$ peers.
α_i	Comparison intensity: low (0.05, 0.20), mid (0.40, 0.60), high (0.80, 0.95); Social Gating uses a near-zero bucket.
g_{ij}	Peer closeness weights. A top weight $g_{\max}$ is drawn from low [0.30, 0.40], mid [0.45, 0.60], or high [0.70, 0.85] dispersion; remaining mass is Dirichlet-split.
x_j	Peer action levels. A skew bucket sets $E[x_j]/F$ in low [1.00, 1.25], mid [1.40, 1.85], or high [2.20, 3.20], with log-normal noise and a boost for the closest peer.
R_i, x_i^*	Oracle aggregate and target commitment, computed as $R_i = \sum_j g_{ij}x_j$ and $x_i^* = F + \alpha_i R_i$.
Transfer latents	Matching Transfer samples benefit–cost buckets and match distance; Rank Transfer samples salary gaps, firm averages, and rank-sensitivity α_{2i} .

Table S1: Latent variables used to construct benchmark records. Student-facing prompts verbalize these states but do not expose formal notation, bucket labels, or oracle rule identifiers.

Cue	Weight range	Example phrasing
tight	$g_{ij} \geq 0.45$	close friend, roommate, mentor, sibling
mid	$0.18 \leq g_{ij} < 0.45$	regular teammate, study-buddy, neighbour
peripheral	$g_{ij} < 0.18$	feed acquaintance, Slack lurker, distant peer

Table S2: Closeness-cue labels shown to the scenario writer (not numeric g_{ij}).

D Training and Probing Setup

Both backbones are fine-tuned with LoRA adapters applied to all attention projection matrices (query, key, value, and output). We use rank $r = 16$, scaling $\alpha_{\text{LoRA}} = 32$, and dropout 0.05. Training runs for 3 epochs with a cosine learning-rate schedule (peak 2×10^{-4}), effective batch size 16, and AdamW with weight decay 10^{-2} . The alignment corpus contains 3,000 training examples drawn from the in-domain split, excluding Matching Transfer and Rank Transfer examples. Llama-3.1-8B uses its instruction-tuned checkpoint; Qwen3-8B uses its chat checkpoint.

Last-token hidden states are extracted at each of the 32 transformer layers of Llama-3.1-8B by processing the scenario in the same format used during alignment training (system prompt + scenario, with a generation-prompt token appended;

the model reads but does not generate). All 1,512 Utility Recovery records are processed in bfloat16 with batch size 4.

Three probes are trained per layer independently for the pre-alignment and aligned checkpoints:

1. A logistic-regression classifier predicting the α_i bucket (low / mid / high).
2. A ridge-regression model predicting $\log R_i = \log\left(\sum_j g_{ij}x_j\right)$.
3. A ridge-regression model predicting $\log \delta x = \log(\alpha_i R_i)$.

Regression targets are log-transformed and standardized. All probes use 5-fold stratified cross-validation; reported values are mean and standard deviation of R^2 or accuracy across folds.

E Dynamic Simulation Environments

E.1 Shared Protocol

At each decision round, every agent observes a local state, selects an action under a fixed policy, and the environment applies aggregate feedback before the next round. Policies are compared on identical random seeds so that network topology, heterogeneity draws, and shock timing are held constant within each case. Four policy types are compared: *oracle* (closed-form best response under the environment’s utility rule), *private* (non-social rationality), *base LLM*, and *aligned LLM*. Tables S3 and S4 summarize the two environments and shared hyperparameters.

Case	Oracle rule; feedback channel; macro outcomes
Financial	Eq. 3; endogenous price; price peak, return volatility, wealth Gini
Stratif.	Eq. 5; tier pool means; elite share, fill gap, sorting skew

Table S3: The two simulation environments reported in the main text.

Parameter	Case 1	Case 2
Agents N	100	100
Horizon T	50	20
Decision interval	every step	every step
Mobility μ	$\lambda=1$	0.60
Benchmark split	Util. Recovery	Rank Trans.
Shock step	5	15

Table S4: Simulation hyperparameters. Case 1 uses full rebalancing ($\lambda=1$); Case 2 uses partial mobility μ for tier switches.

Table S4 lists the shared numerical settings; Case 1 rebalances fully each period whereas Case 2 applies partial mobility when workers switch tiers.

E.2 Case 1: Financial Amplification

$N=100$ agents trade a positional asset on a limit order book. Visible peer exposure is holdings times mid price. Agents sit on a Watts–Strogatz network ($k=6$, $p=0.10$) with up to five visible neighbors per period; roughly 8% are designated as key opinion leaders (KOLs) with early stakes. Initial wealth is \$2,500 per agent; the private anchor F is zero with probability 0.35, otherwise uniform on [\$80, \$2,000]. Each period, agents fully rebalance toward a target exposure. Net buy flow moves the mid price at 5 ticks per \$1,000 traded.

From $t=5$ to $t=19$, 18% of agents act as hubs receiving cash injections and placing market buys combined with exogenous hype drift. When mid price drops 3% below the running peak after a minimum exit step, KOLs exit with amplified sell impact; retail agents exit with a lag. Post-exit price decays toward \$0.01.

E.3 Case 2: Labor Stratification

$N=100$ workers choose between an elite tier H (5 firms, 25 seats each) and a regional tier L (10 firms, 20 seats each). Oracle sorting follows Eq. 5. The

elite wage anchor is \$180,000; the regional anchor is 52% of that value. Pool means $\bar{x}_H$ and $\bar{x}_L$ update endogenously from current memberships. At $t=15$, the elite anchor increases by 12%. When a worker’s target tier differs from the current assignment, the switch succeeds with probability $\mu=0.6$.

E.4 LLM Inference

All other simulation parameters are identical. Each policy is evaluated under five random seeds; Table S7 in Appendix F reports mean $\pm$ s.d. for the LLM columns in Table 5. Across seeds, macro outcomes vary modestly relative to the base-versus-aligned gaps reported in the main text, so the qualitative amplification and stratification patterns are stable under reruns.

F Additional Robustness Results

All static evaluations in the main text (Tables 3 and 4) report means over five random seeds. This section reports seed-level variability, performance on hard Utility Recovery slices, and post-alignment error-type shifts. Table S5 lists the corresponding standard deviations for the main alignment evaluation; standard deviations remain small relative to the Pre versus Aligned gaps, so the ranking of policies is stable across seeds.

Table S6 gives the same breakdown for Llama-3.1-8B ablation targets (Table 4). Seed-level spread is largest for answer-only and latent-ablated targets on structural splits, consistent with weaker mechanism supervision rather than evaluation noise.

Table S7 reports seed-level variability for the dynamic simulations in Table 5. Financial outcomes show the largest absolute spread in price peak and return volatility; stratification metrics vary less but preserve the base-versus-aligned ordering.

F.1 Hard Utility Recovery Slices

Table S8 summarizes accuracy on the two hardest Utility Recovery slices for the aligned 8B backbones (base-model stratification appears in Table 2 in the main text). Both backbones improve on high- α and low-skew items after alignment, with Qwen3-8B showing larger post-alignment gains on the low-skew slice.

F.2 Utility Recovery Error-Type Distribution

Figure 2 in the main text reports the distractor distribution among Utility Recovery mistakes for seven base models. Table S9 and Figure S1 extend this analysis to the two aligned 8B backbones on the

Model	Util.	Gating	Match.	Rank
Llama Pre	56.68±0.45	13.69±0.62	54.86±0.55	49.83±0.38
Llama Aligned	64.95±0.35	95.24±0.28	69.62±0.42	59.20±0.51
Qwen Pre	55.09±0.48	22.42±0.75	57.64±0.46	48.78±0.35
Qwen Aligned	68.98±0.32	98.02±0.18	67.71±0.48	59.55±0.45

Table S5: Alignment accuracy (%) over five evaluation seeds (mean ± s.d.). Corresponds to Table 3.

Target	Util.	Gating	Match.	Rank
Full SFT	64.95±0.35	95.24±0.28	69.62±0.42	59.20±0.51
Answer-only	58.80±0.65	72.02±0.88	44.27±0.72	55.90±0.58
w/o latent	66.14±0.52	83.13±0.75	35.59±0.85	52.08±0.44

Table S6: Llama-3.1-8B ablation accuracy (%) over five evaluation seeds (mean ± s.d.). Corresponds to Table 4.

	Base	Aligned
<i>(a) Financial amplification</i>		
Price peak	14.30±1.85	19.38±2.15
Return volatility	0.33±0.04	0.42±0.05
Wealth Gini	0.34±0.03	0.39±0.04
<i>(b) Stratification</i>		
Elite tier share	0.69±0.05	0.55±0.04
Pool fill gap	0.40±0.06	0.22±0.05
Sorting skew	−0.16±0.03	−0.02±0.02

Table S7: Macro simulation outcomes over five seeds (mean±s.d.) for base and aligned LLM policies. Private baselines in Table 5 are deterministic.

Model	High- α		Low-skew	
	Pre	Post	Pre	Post
Llama-3.1-8B	50.99	61.90	49.60	62.90
Qwen3-8B	47.62	68.45	45.04	67.86

Table S8: Accuracy (%) on the two hardest Utility Recovery slices. High- α : strong comparison pressure. Low-skew: peer actions close to the private baseline.

same 1,512-item split. Each cell reports the share (%) of mis-selected distractor rules among all incorrect answers on gold peer-weighted items; values sum to 100% within each row.

Pre-alignment, both backbones rely heavily on equal-weight mixing (Llama: 35.42%; Qwen: 34.90%) and other aggregation shortcuts (uniform average, median, top anchor). Post-alignment, equal-mix and uniform shares fall for both models, while private-baseline confusions rise sharply—

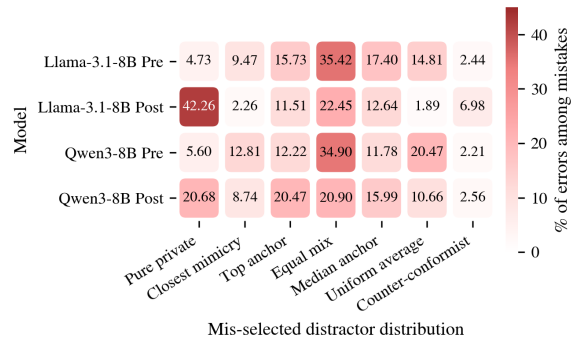


Figure S1: Utility Recovery mis-selected distractor distribution for Llama-3.1-8B and Qwen3-8B before (Pre) and after (Post) alignment. Cell values: share (%) of each distractor among incorrect answers ($N=1,512$).

especially for Llama-3.1-8B (4.73%→42.26%)—consistent with stronger Social Gating and a residual tendency to collapse social influence when peer-weighted aggregation is still missed. Qwen3-8B shows the same directional shift but retains more top-anchor and equal-mix errors (20.47% and 20.90%, respectively), indicating a less extreme re-balancing toward private reasoning.

G Comparative Statics Probe Results

We treat Comparative Statics Probe as an auxiliary directional-response check. It measures whether models choose the correct direction under paired single-factor perturbations. Table S10 reports the full per-model breakdown. The mean accuracy across the seven evaluated models is 60.36%, indicating that directional monotonicity is easier than recovering the peer-weighted utility mechanism tested by Utility Recovery. DeepSeek-V4-Flash

Model	N_{err}	Acc	Priv.	Cl.	Top	Eq.	Med.	Unif.	Ctr.
Llama Pre	655	56.68	4.73	9.47	15.73	35.42	17.40	14.81	2.44
Llama Post	530	64.95	42.26	2.26	11.51	22.45	12.64	1.89	6.98
Qwen Pre	679	55.09	5.60	12.81	12.22	34.90	11.78	20.47	2.21
Qwen Post	469	68.98	20.68	8.74	20.47	20.90	15.99	10.66	2.56

Table S9: Utility Recovery error-type shares (% of mistakes) before/after alignment. Acc: overall accuracy (%). Cl.: closest-peer; Ctr.: counter-aggregation.

Model	Acc. (%)
Qwen3-235B	69.64
Qwen3-14B	60.00
Qwen3-8B	58.93
DeepSeek-V4-Flash	70.71
Llama-3.1-8B	56.07
Zephyr-7B-SFT	51.79
Zephyr-7B- β	55.36
Mean	60.36

Table S10: Comparative Statics Probe accuracy (%). Not used as a primary success criterion for mechanism alignment.

Stage	Prompt configuration	Acc.
Algorithmic oracle	Eq. 3 + matching	100.00
Main eval (Aligned)	narrative only (full split)	64.95
Numeric format	latents + formula; numeric MCQ	87.90
State & mechanism	latents + formula; long-text MCQ	92.70

Table S11: Utility Recovery ceiling calibration for Llama-3.1-8B. Algorithmic oracle and main evaluation: full 1,512-item split. Oracle-assisted rows: stratified 300-item subsample, parsed-only.

and Qwen3-235B score highest on this auxiliary probe, but the spread is modest and does not track Utility Recovery ranking.

H Oracle Ceiling and In-Context Learning Control

Running the closed-form oracle of Eq. 3 through the same option-matching pipeline used for model evaluation yields 100.00% accuracy on all 1,512 Utility Recovery items, with zero ambiguous matches under display rounding. Distractor construction already enforces numeric separability ($\tau=0.10$; Section 3; Appendix C), so this is a construction guarantee. To separate narrative inference from arithmetic and option matching for an aligned 8B model, we evaluate Llama-3.1-8B (Aligned) under progressively stronger oracle injection on a stratified 300-item Utility Recovery subsample. Table S11 reports the ladder alongside the full-split main evaluation from Table 3.

Injecting latents and the aggregation formula with numeric options reaches **87.90%** (an arithmetic ceiling for this backbone on the task). Keeping long-text rule options under the same injection raises accuracy to **92.70%** (+4.80 pp), indicating

that semantic distractors do not confuse the model and can recover minor calculation errors. The remaining gap between closed-book aligned evaluation (64.95%) and the oracle-assisted upper bound is therefore primarily the cost of zero-shot latent recovery from narrative. To test whether exposing the utility-state format in context is sufficient for mechanism repair without parameter updates, we evaluate base Llama-3.1-8B with $k \in \{2, 4, 8\}$ in-domain utility-state demonstrations and no SFT. Demonstrations are drawn from the training pool with held-in latent cells and scene domains disjoint from the evaluated item. Table S12 compares these ICL runs to the Pre→Aligned SFT numbers from Table 3.

ICL yields at most a modest Utility Recovery gain (61.86% at $k=4$ vs. 56.68% pre-trained; +5.18 pp) and does not recover Social Gating or structural OOD transfer. On every split, ICL remains well below aligned SFT. Thus, placing the full utility-state template in context is insufficient to replicate the mechanism-level repair achieved by theory-supervised alignment.

Split	Pre→SFT	$k=2$	$k=4$	$k=8$
Utility Recovery	56.68→64.95	50.96	61.86	57.05
Gating	13.69→95.24	3.85	6.09	9.94
Matching	54.86→69.62	52.24	50.96	47.76
Rank	49.83→59.20	50.96	48.72	51.28

Table S12: Llama-3.1-8B ICL control (accuracy, %) versus Pre→SFT. ICL uses k in-domain utility-state demonstrations without fine-tuning. **Bold**: best $k \in \{2, 4, 8\}$ in each row.

I Prompt Templates and Data Examples

I.1 Scenario writer prompts

DeepSeek-R1-671B receives a domain-specific system prompt plus a structured user brief. The system prompt requires 140–220 words of natural prose, verbatim numeric anchors, no oracle leakage, and JSON output. Positional domains additionally map each peer’s closeness hint (tight / mid / peripheral) to narrative voice and embed the private baseline once. Placebo, Matching Transfer, and Rank Transfer domains use parallel templates with domain-specific framing (private-only investment, semantic benefit–cost tiers, or firm salary/rank cues).

I.2 Rendering quality control

Quality checks require valid JSON fields, verbatim inclusion of all required numeric anchors (F , each visible x_j , and career salaries/averages when applicable), no gold answer or latent notation, no terms such as “alpha”, “bucket”, “weight”, “best response”, or “reference aggregate”, and no extra final-answer number.

I.3 Representative data examples

The training corpus uses open-ended commitment prompts with oracle utility traces. Utility Recovery and Social Gating use multiple-choice rule identification, Comparative Statics Probe uses paired comparative statics, and the Structural Transfer splits use binary choices. The examples below show one complete representative item for each data type.

Training Example

User: Our college friend group is pooling wedding gifts on Splitwise for Mia's big day. I have my own budget in mind, what I'd spend if nobody was watching: \$331. But then I see what everyone else is putting in. Jake, who I grab lunch with every other week, logged \$1,364. Emma, my old study-buddy from sophomore year, added \$781. And then there's my roommate Sam, who I've lived with for three years. He just entered \$2,182. The group chat is buzzing, and the shared ledger shows these numbers in bold. I stare at the screen, my \$331 looking tiny next to theirs. Do I stick to my private baseline and risk looking cheap, or do I stretch to match the crowd?

Question: Based on this scene, what gift spend should the protagonist settle on, and why?

Assistant target:

Utility: private anchor \$331; comparison intensity high; peer incentives: Jake \$1,364 with moderate tie weight, Emma \$781 with moderate tie weight, and Sam \$2,182 with the strongest tie weight; weighted social reference \$1,659; net utility pressure up; optimal commitment \$1,793. Mechanism: the private floor is \$331. The peer reference is a weighted blend in which Sam carries the largest share. Because comparison intensity is high, the social reference creates a strong upward pull. The resulting settlement is the private anchor plus the comparison-driven pull toward the weighted reference. **Final commitment: \$1,793 for the gift spend.**

Utility Recovery Example

Scenario: Our engineering crew keeps a shared Google Calendar tracking system-design mocks. Privately, I figured 3.8 sessions monthly would prep me fine without burnout. But last week's Slack recap pinged my phone: my mentor Eli, who has prepped me for promotions since my first day, crushed 13.0 sessions. Sophia from the infra team shared that she had logged 6.2 sessions during our coffee run. Two others popped up on the calendar feed: Raj from the Bangalore office scheduled 5.4 sessions, and Maya in analytics posted 7.5 sessions in our #mock-masters thread. Now Eli is nudging me to level up, but Sophia's pace feels more sustainable. My cursor hovers over next month's slots. Do I match my mentor's grind or find my own rhythm?

Question: Which rule best captures the optimal monthly mock sessions?

- (A) Combine fifty-fifty the private floor and the weighted peer aggregate, disregarding any individual peer-sensitivity parameter, around 6.9 sessions.
- (B) Anchor on the middle coworker of the group and scale up the private floor by that median pull, around 4.7 sessions.
- (C) Elevate the baseline by a weighted peer aggregate where tighter bonds contribute more share than looser ones, around 5.0 sessions.
- (D) Invert the peer signal: the harder coworkers push up, the further the commitment dips beneath the personal baseline, near 2.6 sessions.

Oracle: $F = 3.81$, $\alpha_i = 0.124$, $R_i = 0.2636 \times 6.2 + 0.1157 \times 5.4 + 0.5559 \times 13.0 + 0.0647 \times 7.5 \approx 9.98$, so $x_i^* = 3.81 + 0.124 \times 9.98 \approx 5.0$ sessions. The gold rule is peer-weighted aggregation. **Answer: C**

Comparative Statics Probe Example

Scenario A: At our neighborhood nursery, the pickup line doubles as an EV showcase among us new parents. I'd been eyeing a modest \$30,489 model, what I'd grab if no eyes were on me. My closest friend Sarah just parked her \$50,847 luxury ride by the gates. My neighbor Ben, whose toddler shares playdates with mine, mentioned his \$40,872 purchase over coffee. Two peripheral faces from the parent app arrived in a \$37,887 compact and a \$46,691 SUV.

Scenario B: As a new parent, I'd planned to spend \$49,969 on an EV, what felt right when nobody was watching. Nursery pickup has become a subtle showroom. My closest friend Jamie parks her glossy Model Y daily and mentioned it ran her \$50,847. Peripheral acquaintances Riley and Alex reported \$37,887 and \$46,691, and my neighbor Taylor chose a \$40,872 crossover.

Question: Which scenario implies the higher optimal EV price?

(A) Scenario A

(B) Scenario B

Oracle: Both records share $\alpha_i = 0.478$, the same peer actions, and the same weights, giving $R_i \approx \$46,515$. Scenario A has $F = \$30,489$ and $x_A^* \approx \$52,719$; Scenario B has $F = \$49,969$ and $x_B^* \approx \$72,199$. The perturbation raises the private baseline, so the correct direction is $B > A$. **Answer: B**

Social Gating Example

Scenario: October winds rattled the maple seeds loose, and Alex knew the gutters needed clearing before Pacific storms hit. The quote from ClearFlow Services was clear: \$381 for the single-story house. At the neighborhood watch meeting, Robert from three blocks over mentioned that his two-story colonial cost \$588 to clean. Linda, Alex's pottery class buddy, texted the group chat about her \$573 bill for clogged downspouts. Dave grumbled about his \$507 charge on a townhome he is selling. Chloe, Alex's hiking partner, called panicked about squirrels nesting in her attic; her emergency cleanout hit \$810. Mr. Patel at the corner store said he had paid \$743 for his vintage home's copper troughs. Alex studied the \$381 estimate again, knowing only their own roof's pitch and oak debris mattered, even though those numbers buzzed like wasps in the eaves.

Question: Which rule best captures the optimal maintenance cost?

- (A) Pool the other people's visible choices with identical weights, then stack that on the private floor, near \$647.
- (B) Set the maintenance cost equal to whatever the most intimate other person picked, about \$810.
- (C) Let the personal floor be the whole answer and treat visible peer amounts as information without pull, \$381.
- (D) Route the peer term through the sole nearest-in-life other person's move above the baseline, near \$715.

Oracle: This is a non-positional maintenance decision. Although a distractor calculation can form a peer reference, the comparison channel is inactive for the gold label. The correct target is therefore the private quote $F = \$381$. **Answer: C**

Matching Transfer Example

Scenario: At the Riverside Tennis Club, you are debating who to partner with for weekly sessions. Booking courts requires matching commitment, and both hours and fees add up. For you, tennis is more than exercise; it is a cornerstone of your week that strengthens real connections. However, \$45 hourly court fees mean adjusting your budget. First option is Mia, a consistent baseline player. She radiates enthusiasm about tennis bonds, calling them essential for mental reset and real friendships. Yet she hesitates before confirming sessions and admits she skips streaming subscriptions to afford it. Second is Jake, whose powerful serve complements your net game. He sees real value in teaming up, saying regular play could make the friendship click. However, his jaw tightens discussing fees; he once canceled last minute when rent was due. Court reservations open tomorrow. Staring at both names, you wonder whose rhythm and reality truly sync with yours.

Question: Which candidate is the more stable weekly tennis partner?

- (A) Mia: strong value for the tie and mid-level cost pressure.
- (B) Jake: clear value for the tie and high cost pressure.

Oracle: The protagonist has $b_i/c_i \approx 2.19$. Mia has $b/c \approx 5.806/4.376 = 1.33$, while Jake has $b/c \approx 3.479/5.722 = 0.61$. Pairwise stability favors the closer ratio, so Mia is closer to the protagonist's preferred intensity. **Answer: A**

Rank Transfer Example

Scenario: Fresh MBA graduate Elena stared at her dual offers. Bellweather Global, a prestigious MBB firm, proposed \$167,251, but during the mixer she had overheard their associates averaged \$180,863. That put her near the bottom of a high-pay cohort. Cascade Partners, a boutique regional consultancy, offered \$115,058. HR mentioned their team average hovered around \$105,139, positioning her at the top of a lower-pay cohort as their star hire. Her friend Priya urged her to take Bellweather because the salary speaks volumes. Her mentor Raj countered that at Cascade she would be their golden child from day one. Elena traced the numbers: Bellweather promised immediate comfort, while Cascade came with instant recognition.

Question: Which firm should Elena choose?

- (A) Bellweather Global, the high-pay firm H : own salary \$167,251 and firm average \$180,863.
- (B) Cascade Partners, the lower-pay firm L : own salary \$115,058 and firm average \$105,139.

Oracle: The absolute salary gain is $\Delta_S = 167,251 - 115,058 = \$52,193$. The average-pay gap is $\Delta_{\bar{x}} = 180,863 - 105,139 = \$75,724$. With $\alpha_{2i} = 0.430$, the relative-rank penalty is $\alpha_{2i}\Delta_{\bar{x}} \approx \$32,574$. Since $\Delta_S \geq \alpha_{2i}\Delta_{\bar{x}}$, the high-pay firm is optimal. **Answer: A**