PEAR: Permutation-Equivariant Adaptive Routing Multi-Agent Debate

Anonymous ACL submission

Abstract

Multi-agent debate improves the reliability of large language models (LLMs) through iterative peer critiques. However, fixed topologies often introduce persistent positional biases, amplify unreliable agents, and cause high sensitivity to role assignments. We introduce *Permutation-Equivariant Adaptive Routing Multi-Agent Debate (PEAR)*, an inference-time protocol that dynamically reconfigures communication roles and sparse topologies across consecutive debate rounds. By strategically switching agent-to-role assignments based on evolving agent states, PEAR prevents any agent from permanently occupying a privileged network position or distributes influence more evenly across the debate. We theoretically characterize PEAR as an equivariant sparse router: it preserves accuracy under agent relabeling while reducing routing complexity and improving generalization. Comprehensive empirical evaluations across four reasoning benchmarks and six diverse LLM backbones demonstrate PEAR significantly improves average accuracy over the strongest debate baselines. The code is available at `https://anonymous.4open.science/r/PEAR-C8AB/README.md`.

1 Introduction

Multi-agent debate (MAD) has emerged as a promising inference-time paradigm for enhancing the reasoning capabilities of large language models (LLMs) (Chen et al., 2024; Du et al., 2024), in which multiple agents independently generate responses and iteratively revise their answers through structured peer critique. By aggregating diverse reasoning paths across agents and rounds, MAD has demonstrated consistent gains on knowledge-intensive and mathematical reasoning benchmarks (Liang et al., 2024).

However, MAD outcomes depend not only on agent capabilities but also on the communication structure through which critiques flow. Most existing protocols use a fixed topology, such as a chain, star, or ring, across all rounds (Liang et al., 2024), which can introduce persistent positional advantages, amplify unreliable agents, and make outcomes sensitive to arbitrary role assignments. These structures create two related failure modes. First, they promote *information homogenization*: agents repeatedly exchange critiques within the same predetermined neighborhoods, so errors can be reinforced rather than challenged. Even when diverse viewpoints are present (Zhu et al., 2026), a fixed topology may prevent them from reaching the agents who would benefit most. Second, they enable *error cascades*: hubs, early speakers, or otherwise privileged agents can exert disproportionate influence, anchoring the group on incorrect conclusions when their own answers are wrong (Zhang et al., 2026; Tian et al., 2026). Prior work identifies these symptoms, but leaves the underlying routing structure largely fixed.

To address these limitations, we introduce Permutation-Equivariant Adaptive Routing Multi-Agent Debate (PEAR), an inference-time framework that replaces fixed communication topologies with state-aware adaptive routing. Rather than fixing the communication graph across rounds, PEAR dynamically selects a sparse communication topology at each debate round based on the current state of all agents, including their answers, self-reported confidence scores, and accumulated influence from prior rounds. Figure 1 illustrates the pipeline.

This routing decision is governed by a composite objective with three components. *Targeted Diversity* prioritizes edges from agents with high confidence and differing answers toward agents with low confidence, ensuring that reliable dissenting viewpoints are routed to the agents most in need of external correction. *Influence Balancing* penalizes sources that have accumulated disproportionate historical influence, preventing any single agent from dominating the information flow regardless of

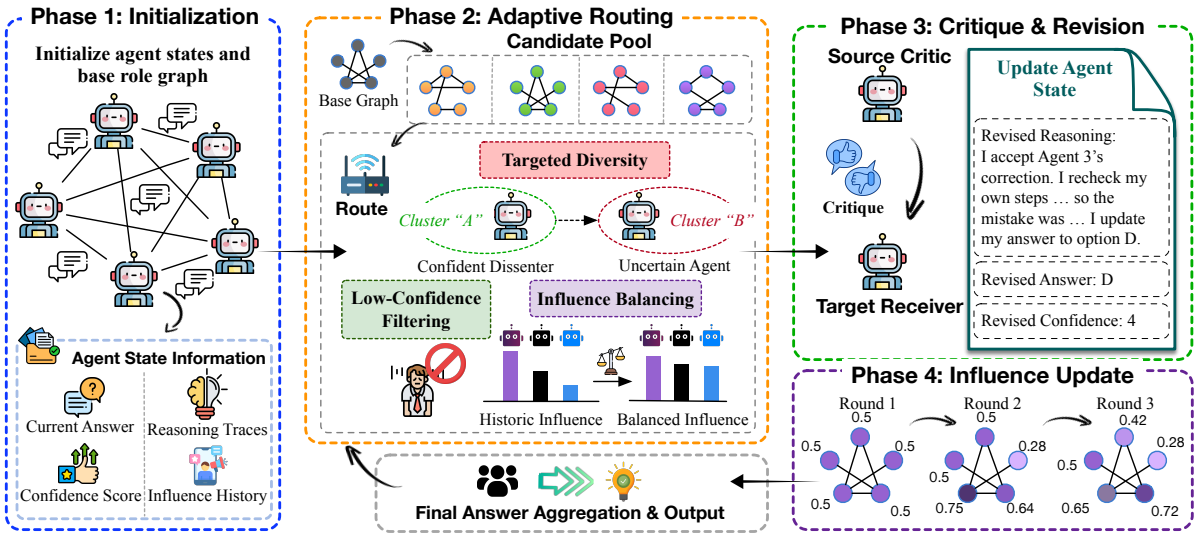

Figure 1: Overview of PEAR. Given a task instance, multiple agents produce independent initial responses. At every debate round, the router selects the communication topology that maximizes a composite score over three state-aware components: targeted diversity, influence balancing, and low-confidence filtering. Sources then critique their assigned targets, targets revise their answers, and the router updates each agent's accumulated influence from the adoption pattern. After several rounds, the final answer is obtained by majority vote.

whether its answer is correct. *Low-Confidence Filtering* suppresses the propagation of critiques from agents with low self-reported confidence, reducing the transmission of unreliable signals through the debate graph. Together, these three components operationalize a principled routing policy that actively promotes wrong-to-right corrections while attenuating the structural pathways through which errors cascade.

We prove that PEAR can be viewed as an agent-permutation equivariant router: relabeling agents relabels the communication graph but leaves the final-answer distribution unchanged. This follows from its search over sparse relabelings of a fixed base graph and routing scores based only on label-invariant features such as disagreement, confidence, and accumulated influence. We further prove that equivariant routing preserves expected accuracy under exchangeable agent populations and does not increase covering complexity, yielding a uniform generalization guarantee. These results suggest that PEAR improves debate by combining adaptivity with the natural symmetries of multi-agent reasoning.

We evaluate PEAR on four benchmarks, MMLU-Pro (Wang et al., 2024), TruthfulQA (Lin et al., 2022), GSM8K (Cobbe et al., 2021), and MATH-500 (Hendrycks et al., 2021), spanning knowledge-intensive reasoning, factual question answering and competition-level mathematics. Experiments are conducted across six instruction-tuned models of varying scale and family, including Gemma-3-12B (Team et al., 2025), Llama-3.1-8B (Grattafiori et al., 2024), Qwen2.5-14B, and Qwen-3-30B-A3B (Yang et al., 2025), together with two closed-source models, GPT-5.4-nano (OpenAI, 2026) and Claude-Haiku-4.5 (Anthropic, 2025). PEAR consistently outperforms all fixed-topology debate variants across all settings, achieving gains of up to 9.0% points over the strongest baseline, with ablation studies confirming the independent contribution of each routing component.

To the best of our knowledge, this is the first paper in the literature that realizes equivariant communication in multi-agent debate, which significantly improves the performance of LLM reasoning.

2 Related Works

Multi-agent debate (MAD). Early work (Du et al., 2024; Chen et al., 2024; Liang et al., 2024) shows that multiple agents iteratively critiquing and revising each other's answers can outperform single-agent inference on knowledge-intensive and reasoning tasks. Subsequently, Liu et al. (2024) partition agents into subgroups to reduce token cost, while Lin and Hooi (2025); Taubenfeld et al. (2025); Fu et al. (2025) incorporate self-reported confidence to weight or filter agents' contributions. More recently, Zhu et al. (2026) identify viewpoint homogeneity as a primary cause of debate stag-

nation; Zhang et al. (2026) show that positionally dominant agents disproportionately determine the final outcome; Tian et al. (2026) reveal that erroneous content accumulated across rounds can persistently mislead downstream agents; and Nguyen et al. (2026) preserve diversity by selectively retaining historically dissenting messages.

Communication topology. Li et al. (2024) demonstrate that sparse topologies can match or exceed fully connected debate at substantially lower cost. Subsequently, Zhang et al. (2024) propose G-Designer, which uses graph neural networks to architect task-specific communication graphs; Shen et al. (2025) autoregressively generate agent topologies; and Li et al. (2025) adaptively prune edges to balance accuracy and efficiency. Beyond LLM debate, Hu et al. (2024) learn communication graphs end-to-end, Sun et al. (2024) perform targeted message selection, Guan et al. (2024) aggregate information via self-supervision, and Zhou et al. (2025) optimize communication at the semantic level. These approaches typically rely on training signals or learned controllers. In contrast, PEAR adapts the topology purely at inference.

3 Preliminaries

We consider a task instance $x \in \mathcal{X}$ with ground-truth label $y^\star \in \mathcal{Y}$ drawn from an unknown distribution $\mathcal{D}$. A debate is conducted by $n \geq 2$ agents indexed by $[n] = \{1, \ldots, n\}$. Each agent i is a stochastic inference-time policy $\pi_{\theta_i}(\cdot \mid x, o_i)$ that maps the task instance and a local observation o_i to a distribution over outputs. At the end of each round r, every agent i exposes a structured tuple $\xi_i^{(r)} = \big(y_i^{(r)}, c_i^{(r)}, r_i^{(r)}\big)$, where $y_i^{(r)} \in \mathcal{Y}$ is the current answer, $c_i^{(r)} \in \{1, \ldots, 5\}$ is the self-reported confidence on an ordinal scale, and $r_i^{(r)}$ is the accompanying chain-of-thought reasoning. In parallel, the router maintains an accumulated *influence* statistic $\rho_i^{(r)} \in [0, 1]$ that summarizes how often agent i's prior critiques have been adopted by its downstream targets; its update rule is given in Section 4. The *debate state* observable to the router at the start of round r is $s_r = \big\{\big(\xi_i^{(r-1)}, \rho_i^{(r-1)}\big)\big\}_{i \in [n]}$.

Role graph. Let $G_0 = (V_0, E_0)$, $V_0 = [n]$, be a directed base role graph that fixes a sparse communication pattern. An edge $(u \to v) \in E_0$ permits the agent occupying role v to observe messages produced by the agent occupying role u. A role-to-agent assignment is a bijection $\pi \in S_n$, where $\pi(u)$ is the agent occupying role u. The induced agent-level communication graph is

$$G(\pi) = ([n],\, E(\pi)),$$
$$E(\pi) = \{(\pi(u) \to \pi(v)) : (u \to v) \in E_0\}.$$

By construction, every $G(\pi)$ preserves the per-agent in-degree of G_0, so the assignment π only relabels which concrete agent occupies each role and yields a candidate sparse communication graph rather than an arbitrary subset of edges. Since every role has in-degree k, the base graph has $m := |E_0| = nk$ directed edges. Every candidate graph $G(\pi)$ therefore satisfies $|E(\pi)| = m$.

Round-based debate. A debate proceeds for R rounds. At the start of round r, the round-r communication graph $G^{(r)} = G(\pi_r)$ is selected as a state-aware function of s_r; the specific selection rule is given in Section 4. After R rounds, the protocol returns $\hat{y} = \mathrm{Agg}\big(\{y_i^{(R)}\}_{i \in [n]}\big)$, where Agg is the aggregation rule (we use majority vote).

4 Permutation-Equivariant Adaptive Routing Multi-Agent Debate

PEAR proceeds in four phases: Initialization, Adaptive Routing, Critique & Revision, and Influence Update. The latter three are repeated for R rounds. Figure 1 provides a schematic overview.

4.1 Initialization

Base role graph. PEAR instantiates the base role graph G_0 from Section 3 as a sparse k-regular template in which every role has in-degree k. The k-regular template is preferred over a clique for three reasons: (i) it caps the per-round critique budget at nk rather than $n(n-1)$ edges; (ii) it keeps the per-agent input bounded, mitigating the noise introduced by aggregating critiques from every other agent; and (iii) it admits a tractable space of candidate topologies $\{G(\pi) : \pi \in S_n\}$ obtained by relabeling roles, so the router can search among meaningfully different communication graphs while preserving each agent's input budget at k.

Initial responses. Before any communication occurs, each agent $i \in [n]$ independently produces an initial response $\xi_i^{(0)} \sim \pi_{\theta_i}(\cdot \mid x)$, using the structured tuple defined in Section 3. Agents are instructed to reserve the low end of the confidence

scale for guesses or for cases in which competing alternatives cannot be ruled out, and to reserve the highest score for fully verified solutions. The router-side influence statistic is initialized to $\rho_i^{(0)} = 0$ for every agent.

4.2 Adaptive Routing

Candidate pool. At the start of round r, the router constructs a finite candidate pool $\mathcal{C}_r = \{\pi_r^{(1)}, \ldots, \pi_r^{(M)}\} \subseteq S_n$, $M = |\mathcal{C}_r|$, by enumerating assignments that yield distinct edge sets $E(\pi_r^{(m)})$. The size M is upper-bounded by a configurable budget $M_{\max}$. The candidate pool merely delimits the search space over which the state-aware score is computed; it is not the routing decision itself. For a candidate topology $G(\pi)$ with edge set $E = E(\pi)$, the router computes a composite score $S(E \mid s_r)$ from three state-aware components, each normalized so that the relative magnitudes of the weights $\alpha_T, \alpha_I, \alpha_L$ remain interpretable.

Targeted diversity. This component rewards edges whose source is confident, whose target is uncertain, and whose two endpoints currently hold different answers. Concretely, for an edge $(s \to t) \in E$,

$$T(s,t) = \mathbf{1}\big[y_s^{(r-1)} \neq y_t^{(r-1)}\big]\,\mathbf{1}\big[c_s^{(r-1)} \geq \tau_{\text{src}}\big]\,\mathbf{1}\big[c_t^{(r-1)} \leq \tau_{\text{tgt}}\big], \tag{1}$$

where τ_{src} and τ_{tgt} are confidence cutoffs for the source and target, respectively. The normalized rate over a candidate edge set is

$$\widetilde{T}(E) = \frac{1}{m} \sum_{(s,t)\in E} T(s,t) \in [0,1]. \tag{2}$$

Equation (1) is asymmetric by design: it favors directed dissent from high-confidence agents toward low-confidence agents, in line with the empirical observation that reliable, opinion-divergent sources are the most likely to convert uncertain targets from wrong to right.

Influence balancing. Let $\rho_s^{(r-1)} \in [0,1]$ denote the accumulated influence of agent s entering round r, and let $d_s^{\text{out}}(E) = |\{t : (s,t) \in E\}|$ be its out-degree under the candidate edge set. The influence-balancing component penalizes routing structures that allocate additional out-degree to already dominant agents:

$$\widetilde{I}(E) = \frac{1}{m} \sum_{s\in[n]} \rho_s^{(r-1)}\, d_s^{\text{out}}(E) \in [0,1]. \tag{3}$$

Algorithm 1 PEAR

Require: Input x; agents $\{\pi_{\theta_i}\}_{i=1}^n$; base graph G_0; rounds R; weights $(\alpha_T, \alpha_I, \alpha_L)$; thresholds $(\tau_{\text{src}}, \tau_{\text{tgt}}, \tau_{\text{low}})$; smoothing β; temperature τ

For all $i \in [n]$: $\xi_i^{(0)} \sim \pi_{\theta_i}(\cdot \mid x)$ and $\rho_i^{(0)} \leftarrow 0$
for $r = 1, \ldots, R$ **do**
  $\pi_r \sim Q_r(\cdot \mid s_r)$; $E^{(r)} \leftarrow E(\pi_r)$
    ▷ softmax routing; reduces to argmax as $\tau \to 0$
  For $(s,t) \in E^{(r)}$: $m_{s\to t}^{(r)} \sim \pi_{\theta_s}\big(\cdot \mid x, \xi_s^{(r-1)}, \xi_t^{(r-1)}\big)$
    ▷ Critique
  $\xi_t^{(r)} \sim \pi_{\theta_t}(\cdot \mid x, \xi_t^{(r-1)}, \mathcal{I}_t^{(r)})$ for $t \in [n]$ ▷ Revision
  $\rho_s^{(r)} \leftarrow \beta\rho_s^{(r-1)} + (1-\beta)a_s^{(r)}$ for $s \in [n]$ ▷ Influence
end for
return $\hat{y} = \mathrm{Agg}\big(\{y_i^{(R)}\}_{i\in[n]}\big)$

Here $\rho_s^{(r-1)} \in [0,1]$ by induction from the EMA update in Eq. (8), and $\sum_s d_s^{\text{out}}(E) = m$, so $\sum_s \rho_s^{(r-1)} d_s^{\text{out}}(E) \leq m$. This term provides a structural defense against error cascades: an agent that has previously persuaded many targets is prevented from being amplified further, regardless of whether its current answer is correct.

Low-confidence filtering. This component suppresses critique edges whose source is unreliable. For each source s, define a confidence penalty

$$L(s) = \max\left\{0, \tau_{\text{low}} + 1 - c_s^{(r-1)}\right\}, \tag{4}$$

which equals zero whenever the source confidence exceeds the threshold τ_{low} and grows linearly as confidence drops below it. We take τ_{low} to be an integer threshold on the ordinal confidence scale. Since $c_s^{(r-1)} \in \{1, \ldots, 5\}$, the penalty is zero exactly when $c_s^{(r-1)} \geq \tau_{\text{low}} + 1$, equivalently when $c_s^{(r-1)} > \tau_{\text{low}}$. The normalized penalty rate is

$$\widetilde{L}(E) = \frac{1}{m\tau_{\text{low}}} \sum_{(s,t)\in E} L(s) \in [0,1]. \tag{5}$$

The denominator τ_{low} is chosen because the maximum possible penalty is attained at minimum confidence $c_s = 1$, where $L(s) = \tau_{\text{low}}$. Unlike a confidence-maximizing rule, Equation (4) only penalizes *low*-confidence sources and never rewards the highest-confidence one, consistent with the observation that confidence is informative as a filter but unreliable as a sole selector.

Composite score and topology selection. The state-conditioned routing score for candidate π is the linear combination

$$S\big(E(\pi) \,\big|\, s_r\big) = \alpha_T\, \widetilde{T}\big(E(\pi)\big) - \alpha_I\, \widetilde{I}\big(E(\pi)\big) - \alpha_L\, \widetilde{L}\big(E(\pi)\big), \tag{6}$$

with non-negative weights $(\alpha_T, \alpha_I, \alpha_L)$ that are held fixed across rounds. Given the composite scores $\{S(E(\pi) \mid s_r) : \pi \in \mathcal{C}_r\}$, the router selects the round-r assignment via the state-conditional distribution

$$Q_r(\pi \mid s_r) = \frac{\exp\big(S(E(\pi) \mid s_r)/\tau\big)}{\sum_{\pi' \in \mathcal{C}_r} \exp\big(S(E(\pi') \mid s_r)/\tau\big)},$$
$$\pi_r \sim Q_r(\cdot \mid s_r), \tag{7}$$

with temperature τ. Equation (7) concentrates mass on the highest-scoring candidate while preserving a controllable degree of exploration over near-optimal candidates, and reduces to argmax selection as $\tau \to 0$. The resulting round-r topology is $G^{(r)} = G(\pi_r)$.

4.3 Critique & Revision

Critique generation. For each directed edge $(s \to t) \in E^{(r)}$, the source agent reads the target's previous answer $y_t^{(r-1)}$, reasoning $r_t^{(r-1)}$, and confidence $c_t^{(r-1)}$, and emits a targeted critique conditioned on its own current state:

$$m_{s\to t}^{(r)} \sim \pi_{\theta_s}\big(\cdot \mid x,\, \xi_t^{(r-1)},\, \xi_s^{(r-1)}\big).$$

The critique format is unconstrained beyond requiring a verdict and a justification.

Answer update. Each target t then revises its output given the incoming critique set $\mathcal{I}_t^{(r)} = \{m_{s\to t}^{(r)} : (s \to t) \in E^{(r)}\}$:

$$\xi_t^{(r)} = \big(y_t^{(r)},\, c_t^{(r)},\, r_t^{(r)}\big)$$
$$\sim \pi_{\theta_t}\big(\cdot \mid x,\, \xi_t^{(r-1)},\, \mathcal{I}_t^{(r)}\big).$$

In the same response, the target emits an `ACCEPT`/`REJECT` decision for each incoming critique; these decisions are recorded by the router and consumed in the next phase.

4.4 Influence Update

After all agents have updated, the router computes the round-r adoption rate of each source s,

$$a_s^{(r)} = \frac{\big|\{t : (s,t) \in E^{(r)},\ m_{s\to t}^{(r)} \text{ accepted}\}\big|}{\max\big(1,\, d_s^{\text{out}}(E^{(r)})\big)},$$

and refreshes the influence statistic with an exponential moving average:

$$\rho_s^{(r)} = \beta\, \rho_s^{(r-1)} + (1-\beta)\, a_s^{(r)}, \tag{8}$$

with smoothing coefficient $\beta \in [0, 1)$. The updated influence enters the routing score at round $r+1$ through Equation (3), closing the loop between past persuasion and future structural privilege.

5 Theoretical Results

Formalization. Let $\mathfrak{S}_n$ be the symmetric group over agents. For $\sigma \in \mathfrak{S}_n$, write σs for the debate state obtained by renaming each agent i as $\sigma(i)$, and write $\sigma E = \{(\sigma u, \sigma v) : (u, v) \in E\}$ for the corresponding relabeling of a directed edge set. We assume the candidate family is closed under relabeling: $E \in \mathcal{C}(s) \iff \sigma E \in \mathcal{C}(\sigma s)$. This holds for the full role-assignment family induced by a fixed base graph, since each candidate is obtained by assigning concrete agents to the same set of roles. Let $m = |E_0| = nk$ denote the number of critique edges in every candidate graph.

Equivariance. PEAR scores each candidate edge set by $\alpha_T \widetilde{T}(E \mid s) - \alpha_I \widetilde{I}(E \mid s) - \alpha_L \widetilde{L}(E \mid s)$, where $\widetilde{T}$ rewards confident disagreement toward uncertain targets, $\widetilde{I}$ penalizes allocating out-degree to previously influential agents, and $\widetilde{L}$ penalizes low-confidence sources. We thus have the following theorem that proves PEAR is an agent-equivariant router. The proof is in Appendix A.1.

Theorem 1 (Agent-permutation equivariance)**.** *Assume the candidate family is closed under agent relabeling. Then, for x $\sigma \in \mathfrak{S}_n$, $S_\alpha(\sigma E \mid \sigma s) = S_\alpha(E \mid s)$, $Q_\alpha(\sigma E \mid \sigma s) = Q_\alpha(E \mid s)$.*

Symmetrized routing. For comparison, consider any possibly non-equivariant routing policy over full routing transcripts. Let $\Gamma = (E^{(1)}, \ldots, E^{(R)})$ be a sequence of selected edge sets, and let $q(\Gamma \mid s^{(0)})$ be the distribution over routing transcripts from initial state $s^{(0)}$. Define orbit-averaged router $(\mathcal{P}q)(\Gamma \mid s^{(0)}) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} q(\sigma\Gamma \mid \sigma s^{(0)})$, where $\sigma\Gamma = (\sigma E^{(1)}, \ldots, \sigma E^{(R)})$. Let $A(\Gamma, s^{(0)}) \in [0, 1]$ denote the final-answer accuracy induced by routing Γ from initial state $s^{(0)}$.

Theorem 2 (Accuracy preservation)**.** *Suppose the distribution of initial debate states is exchangeable, the agent update kernels are permutation-equivariant, and the final aggregation rule is permutation-invariant. Then*

$$\mathbb{E}_{s^{(0)}, \Gamma \sim \mathcal{P}q}\left[A(\Gamma, s^{(0)})\right] = \mathbb{E}_{s^{(0)}, \Gamma \sim q}\left[A(\Gamma, s^{(0)})\right].$$

Moreover, $\mathcal{P}q$ is agent-equivariant:

$$(\mathcal{P}q)(\sigma\Gamma \mid \sigma s^{(0)}) = (\mathcal{P}q)(\Gamma \mid s^{(0)}),\ \forall \sigma \in \mathfrak{S}_n.$$

Thus, under exchangeable agent populations, enforcing equivariance does not change expected final accuracy; it only removes arbitrary dependence on the agent names. The proof is in Appendix A.2.

Generalization. We next formalize the complexity benefit of equivariant routers. Let $\mathcal{Q}$ be a class of one-round routers $q(\cdot \mid s)$, and define $d(q, q') = \sup_s \|q(\cdot \mid s) - q'(\cdot \mid s)\|_1$. Let $\mathcal{P}\mathcal{Q} = \{\mathcal{P}q : q \in \mathcal{Q}\}$ be the orbit-projected class. We use covering number to measure the hypothesis complexity; see Definition 1 in Appendix A.3.

Lemma 1 (Covering number control)**.** *For every $\epsilon > 0$, the covering number satisifies, $\mathcal{N}(\mathcal{P}\mathcal{Q}, \epsilon, d) \leq \mathcal{N}(\mathcal{Q}, \epsilon, d)$.*

Theorem 3 (Generalization bound)**.** *Let $z_1, \ldots, z_{N_{\text{ex}}}$ be i.i.d. task instances, and let $F(q; z) \in [0, 1]$ be an evaluation functional, such as final-answer accuracy, one-round correction, or negative routing regret. Suppose F is L-Lipschitz in q under d: $|F(q; z) - F(q'; z)| \leq Ld(q, q')$, $\forall q, q', z$. Define $\mathcal{L}(q) = \mathbb{E}_z[F(q; z)]$ and $\widehat{\mathcal{L}}_{N_{\text{ex}}}(q) = \frac{1}{N_{\text{ex}}} \sum_{\ell=1}^{N_{\text{ex}}} F(q; z_\ell)$. Then, for any $\epsilon > 0$ and $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\left|\mathcal{L}(q) - \widehat{\mathcal{L}}_{N_{\text{ex}}}(q)\right| \leq 2L\epsilon + \sqrt{\frac{\log\left(2\mathcal{N}(\mathcal{Q}, \epsilon, d)/\delta\right)}{2N_{\text{ex}}}}.$$

Lemma 1 and Theorem 3 show that restricting to equivariant routers cannot increase the covering complexity of the routing class. PEAR therefore combines a sparse communication budget with an equivariant search space, reducing both per-round debate cost and the effective complexity of the router. Proofs are in Appendix A.3.

One-round correction and influence-balancing. More theoretical results on one-round correction and influence-balancing are given in Appendix B.

6 Experiments

6.1 Experimental Setup

Benchmarks. We use four benchmarks in knowledge reasoning, factuality, and mathematical reasoning. **MMLU-Pro** (Wang et al., 2024) evaluates broad professional and academic knowledge with multi-step reasoning questions. **TruthfulQA** (Lin et al., 2022) measures factual reliability and resistance to common misconceptions or false beliefs. **GSM8K** (Cobbe et al., 2021) focuses on grade-school mathematical word problems that require multi-step arithmetic reasoning. **MATH-500** (Hendrycks et al., 2021) contains competition-level mathematics problems with substantially longer and more complex reasoning chains.

Models. We use four open-source instruction-tuned models, namely Gemma-3-12B (Team et al., 2025), Llama-3.1-8B (Grattafiori et al., 2024), Qwen2.5-14B, and Qwen-3-30B-A3B (Yang et al., 2025), together with two closed-source models, GPT-5.4-nano (OpenAI, 2026) and Claude-Haiku-4.5 (Anthropic, 2025). All models are used in the main experiments, while ablation studies are conducted on the three smaller open-source models (Gemma-3-12B, Llama-3.1-8B, and Qwen2.5-14B) to reduce computational cost.

Baselines. We compare PEAR against the following baselines: **CoT** (Wei et al., 2022), a single-agent chain-of-thought method; **CoT-SC** (Wang et al., 2022), self-consistency over multiple independent samples; and a set of fixed-topology multi-agent debate variants including fully connected (**Clique**), hub-and-spoke (**Star**), sequential (**Chain**), and cyclic (**Ring**) communication graphs. We further include **Random**, a dynamic sparse baseline where each agent receives critiques from randomly selected peers at each round.

Metrics. We report *round-wise* and *final-answer accuracy* as the main results via majority vote, together with five trajectory-level diagnostics: the wrong-to-right and right-to-wrong update rates (*W2R* / *R2W*), the *critique acceptance rate*, the *cross-answer routing rate* (fraction of routed edges whose source and target hold different answers), *source confidence* (mean source-side confidence in $[0, 1]$), and *influence entropy* (normalized entropy of the influence distribution). Detailed definitions are given in Appendix C.4.

Debate configuration. PEAR uses $n = 5$ agents and $R = 5$ debate rounds. The base topology is k-regular with $k = 2$, so each target agent receives critiques from two source agents per round. Further implementation details, including hyperparameter selection, are provided in Appendix C.1.

Reproducibility. Prompt templates and a case study of multi-round debate dynamics are in Appendices D and E, respectively. The code is at `https://anonymous.4open.science/r/PEAR-C8AB/README.md`.

6.2 Experimental Results

Overall accuracy. Table 1 reports final-answer accuracy across all model–dataset combinations. Open-source models are evaluated on 200 examples per dataset and closed-source models on 100

Table 1: Final-answer accuracy across datasets and models evaluated on subset settings. Results are averaged over 5 random seeds and reported as mean ± standard deviation. Best results in each row are bolded.

Model	CoT	CoT-SC	Clique	Star	Chain	Ring	Random	PEAR
MMLU-Pro								
Gemma-3-12B	$0.375_{\pm.018}$	$0.410_{\pm.016}$	$0.615_{\pm.014}$	$0.605_{\pm.015}$	$0.525_{\pm.022}$	$0.575_{\pm.018}$	$0.585_{\pm.017}$	$\mathbf{0.645}_{\pm.013}$
Llama-3.1-8B	$0.415_{\pm.021}$	$0.460_{\pm.020}$	$0.595_{\pm.018}$	$0.560_{\pm.021}$	$0.520_{\pm.019}$	$0.495_{\pm.023}$	$0.540_{\pm.020}$	$\mathbf{0.625}_{\pm.017}$
Qwen-2.5-14B	$0.440_{\pm.017}$	$0.440_{\pm.019}$	$0.540_{\pm.015}$	$0.560_{\pm.018}$	$0.620_{\pm.014}$	$0.600_{\pm.015}$	$0.590_{\pm.016}$	$\mathbf{0.665}_{\pm.012}$
Qwen-3-30B-A3B	$0.580_{\pm.014}$	$0.645_{\pm.013}$	$0.780_{\pm.011}$	$0.765_{\pm.012}$	$0.785_{\pm.010}$	$0.770_{\pm.011}$	$0.772_{\pm.010}$	$\mathbf{0.805}_{\pm.009}$
GPT-5.4-nano	$0.505_{\pm.026}$	$0.550_{\pm.021}$	$0.685_{\pm.018}$	$0.630_{\pm.020}$	$0.650_{\pm.019}$	$0.645_{\pm.018}$	$0.660_{\pm.017}$	$\mathbf{0.730}_{\pm.015}$
Claude-Haiku-4.5	$0.515_{\pm.023}$	$0.545_{\pm.021}$	$0.710_{\pm.016}$	$0.665_{\pm.017}$	$0.680_{\pm.016}$	$0.635_{\pm.020}$	$0.690_{\pm.015}$	$\mathbf{0.775}_{\pm.013}$
TruthfulQA								
Gemma-3-12B	$0.705_{\pm.014}$	$0.710_{\pm.013}$	$0.755_{\pm.011}$	$0.760_{\pm.012}$	$0.775_{\pm.010}$	$0.760_{\pm.011}$	$0.770_{\pm.010}$	$\mathbf{0.825}_{\pm.009}$
Llama-3.1-8B	$0.645_{\pm.018}$	$0.685_{\pm.017}$	$0.785_{\pm.013}$	$0.745_{\pm.015}$	$0.750_{\pm.014}$	$0.740_{\pm.015}$	$0.765_{\pm.012}$	$\mathbf{0.815}_{\pm.011}$
Qwen-2.5-14B	$0.725_{\pm.013}$	$0.780_{\pm.012}$	$0.875_{\pm.009}$	$0.820_{\pm.011}$	$0.825_{\pm.011}$	$0.855_{\pm.010}$	$0.845_{\pm.009}$	$\mathbf{0.890}_{\pm.008}$
Qwen-3-30B-A3B	$0.740_{\pm.012}$	$0.795_{\pm.011}$	$0.845_{\pm.009}$	$0.800_{\pm.010}$	$0.820_{\pm.010}$	$0.780_{\pm.011}$	$0.835_{\pm.009}$	$\mathbf{0.905}_{\pm.007}$
GPT-5.4-nano	$0.635_{\pm.021}$	$0.620_{\pm.022}$	$0.845_{\pm.012}$	$0.790_{\pm.015}$	$0.785_{\pm.015}$	$0.830_{\pm.013}$	$0.835_{\pm.012}$	$\mathbf{0.915}_{\pm.010}$
Claude-Haiku-4.5	$0.615_{\pm.022}$	$0.695_{\pm.018}$	$0.855_{\pm.012}$	$0.775_{\pm.014}$	$0.785_{\pm.013}$	$0.835_{\pm.012}$	$0.845_{\pm.011}$	$\mathbf{0.905}_{\pm.010}$
GSM8K								
Gemma-3-12B	$0.805_{\pm.010}$	$0.815_{\pm.009}$	$0.900_{\pm.007}$	$0.915_{\pm.007}$	$0.885_{\pm.008}$	$0.890_{\pm.008}$	$0.905_{\pm.006}$	$\mathbf{0.955}_{\pm.005}$
Llama-3.1-8B	$0.785_{\pm.012}$	$0.810_{\pm.011}$	$0.835_{\pm.009}$	$0.865_{\pm.008}$	$0.875_{\pm.008}$	$0.855_{\pm.009}$	$0.860_{\pm.007}$	$\mathbf{0.895}_{\pm.006}$
Qwen-2.5-14B	$0.825_{\pm.009}$	$0.820_{\pm.009}$	$0.890_{\pm.007}$	$0.895_{\pm.007}$	$0.930_{\pm.006}$	$0.915_{\pm.007}$	$0.905_{\pm.006}$	$\mathbf{0.945}_{\pm.005}$
Qwen-3-30B-A3B	$0.835_{\pm.008}$	$0.845_{\pm.008}$	$0.960_{\pm.004}$	$0.955_{\pm.004}$	$0.945_{\pm.005}$	$0.950_{\pm.005}$	$0.955_{\pm.004}$	$\mathbf{0.980}_{\pm.003}$
GPT-5.4-nano	$0.815_{\pm.011}$	$0.835_{\pm.010}$	$0.905_{\pm.007}$	$0.895_{\pm.008}$	$0.885_{\pm.008}$	$0.880_{\pm.008}$	$0.900_{\pm.007}$	$\mathbf{0.950}_{\pm.005}$
Claude-Haiku-4.5	$0.845_{\pm.009}$	$0.835_{\pm.010}$	$0.910_{\pm.007}$	$0.890_{\pm.008}$	$0.905_{\pm.007}$	$0.895_{\pm.008}$	$0.915_{\pm.006}$	$\mathbf{0.955}_{\pm.005}$
MATH-500								
Gemma-3-12B	$0.185_{\pm.020}$	$0.205_{\pm.019}$	$0.275_{\pm.016}$	$0.285_{\pm.016}$	$0.295_{\pm.015}$	$0.290_{\pm.015}$	$0.300_{\pm.015}$	$\mathbf{0.335}_{\pm.013}$
Llama-3.1-8B	$0.135_{\pm.022}$	$0.165_{\pm.020}$	$0.205_{\pm.018}$	$0.195_{\pm.019}$	$0.225_{\pm.017}$	$0.215_{\pm.018}$	$0.220_{\pm.017}$	$\mathbf{0.255}_{\pm.015}$
Qwen-2.5-14B	$0.255_{\pm.018}$	$0.275_{\pm.017}$	$0.385_{\pm.014}$	$0.395_{\pm.014}$	$0.420_{\pm.013}$	$0.410_{\pm.013}$	$0.405_{\pm.014}$	$\mathbf{0.485}_{\pm.011}$
Qwen-3-30B-A3B	$0.345_{\pm.015}$	$0.440_{\pm.013}$	$0.540_{\pm.011}$	$0.560_{\pm.011}$	$0.620_{\pm.010}$	$0.610_{\pm.010}$	$0.585_{\pm.012}$	$\mathbf{0.665}_{\pm.008}$
GPT-5.4-nano	$0.315_{\pm.017}$	$0.340_{\pm.016}$	$0.455_{\pm.013}$	$0.445_{\pm.013}$	$0.435_{\pm.014}$	$0.440_{\pm.013}$	$0.460_{\pm.013}$	$\mathbf{0.525}_{\pm.010}$
Claude-Haiku-4.5	$0.285_{\pm.018}$	$0.355_{\pm.016}$	$0.425_{\pm.014}$	$0.410_{\pm.014}$	$0.405_{\pm.014}$	$0.415_{\pm.013}$	$0.420_{\pm.014}$	$\mathbf{0.495}_{\pm.011}$

due to API cost constraints; full-scale results are reported in Appendix C.2. PEAR achieves the highest accuracy in every row, regardless of dataset, model scale, or provenance. Averaged over all settings, PEAR reaches a mean accuracy of 0.701, compared with 0.620 for Fixed Clique, 0.602 for Fixed Star, 0.610 for Fixed Chain, and 0.609 for Fixed Ring. Even when each setting is allowed to pick its single most favorable fixed topology, PEAR still gains 5.1 points on average.

Per-dataset gains over the best fixed topology are 6.0 points on MMLU-Pro, 3.9 on TruthfulQA, 4.8 on GSM8K, and 5.6 on MATH-500. The improvement is most pronounced on weaker backbones: on Llama-3.1-8B, PEAR gains 8.0, 9.0, and 9.5 points on MMLU-Pro, GSM8K, and MATH-500 respectively, consistent with the failure modes of fixed-topology debate being more severe when individual agents are weaker. Single-agent baselines such as CoT and CoT-SC trail every debate variant by often double-digit margins, and CoT-SC closes only a small portion of the gap, indicating that the gains arise from inter-agent communication rather than from drawing more samples per query.

Ablation study. We ablate the three routing components of PEAR: targeted diversity, influence balancing, and low-confidence filtering. Figure 2 reports accuracy for all single-component, two-component, and full variants on 3 open-source models. The ablation results show that the full routing objective performs best on average. Among partial variants, combining targeted diversity with low-confidence filtering and combining influence balancing with low-confidence filtering are the strongest, reaching average accuracies of 0.603 and 0.604, respectively. The full model reaches 0.657, suggesting that the three components provide complementary benefits. The improvement is particularly large for Llama-3.1-8B, where PEAR Full substantially outperforms all ablations on GSM8K, MATH-500, MMLU-Pro, and TruthfulQA.

Accuracy across debate rounds. Figures 3 and 4 in Appendix C.3 plot round-wise accuracy on Qwen-2.5-14B and GPT-5.4-nano respectively. Across both backbones and all four datasets, PEAR

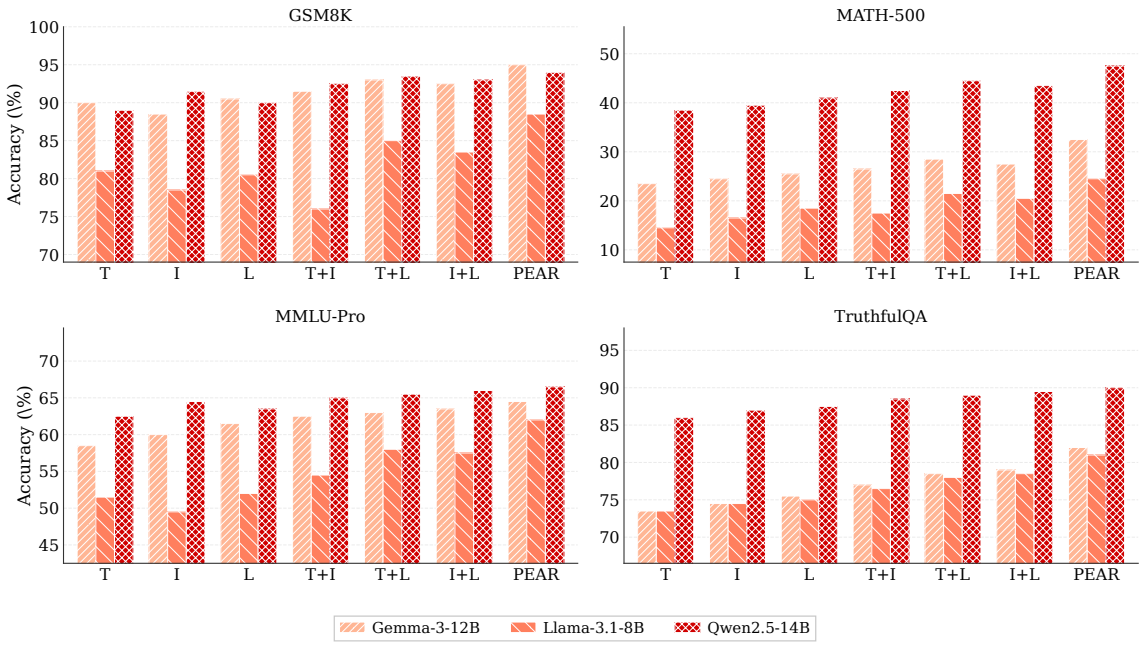

Figure 2: Ablation study of PEAR routing components on open-source models. Targeted = targeted diversity, Influence = influence balancing, LowConf = low-confidence filtering. Each group of bars corresponds to one backbone model; colors indicate routing variant, with PEAR consistently achieving the highest accuracy.

achieves higher accuracy at every round and converges to a higher plateau than every fixed-topology baseline, with the advantage widening progressively as the debate proceeds. For example, on MATH-500 with Qwen-2.5-14B, PEAR improves from 0.284 at Round 1 to 0.485 at Round 5, whereas the strongest fixed baseline saturates at 0.420. The trend is sharper on harder tasks with GPT-5.4-nano: PEAR reaches 0.525 on MATH-500 and 0.730 on MMLU-Pro by Round 5, exceeding the next-best baseline by 0.07 and 0.045 points respectively. Even on TruthfulQA, where all methods start from a higher base, PEAR maintains a consistent lead, reaching 0.915 at Round 5 against 0.845 for Fixed Clique.

Trajectory-level diagnostics. The diagnostics in Table 3 in Appendix C.4 show that PEAR improves both correction quality and routing diversity compared to all baselines. It achieves the highest net correction rate of 0.243, indicating that beneficial updates substantially outweigh harmful flips, and exhibits substantially higher cross-answer routing at 0.676 versus 0.43–0.50 for all baselines, suggesting that adaptive routing encourages information exchange across diverse reasoning trajectories rather than reinforcing local agreement. PEAR also yields the highest influence entropy at 0.979, reflecting a more balanced distribution of influence across agents. Detailed metric definitions and analyses are provided in Appendix C.4.

Computational overhead. As shown in Figure 5 in Appendix C.5, PEAR lies on the cost–accuracy Pareto frontier on every dataset. It consumes about 32.3k tokens per example, comparable to the Random baseline at the same k-regular edge budget and 12% lower than the densest Fixed Clique baseline at 36.7k tokens. Adaptive routing therefore adds *negligible* cost over uniformly random routing under a similar edge budget, while improving accuracy by 5.7 points over Random and 6.6 points over the cheaper Fixed Chain baseline. More details are provided in Appendix C.5.

7 Conclusions

This paper presents *Permutation-Equivariant Adaptive Routing Multi-Agent Debate (PEAR)*, an inference-time framework that selects a sparse communication topology at each debate round based on agent states. This state-aware routing actively suppresses information homogenization and error cascades, the two common failure modes of fixed-topology debate. Extensive experiments on four benchmarks and six LLMs show that PEAR consistently outperforms fixed-topology and dynamic baselines, with gains driven by improved correction dynamics and more diverse communication.

Limitations

Computational overhead. Although PEAR lies on the cost–accuracy Pareto frontier among debate methods (see in Appendix C.5), its per-example token usage remains higher than single-agent CoT and CoT-SC. Promising directions for reducing cost while preserving routing quality include adaptive termination, i.e., stopping rounds once a confidence-weighted consensus is reached, and sparser candidate pools; we leave a systematic study of these complementary techniques to future work.

Reliance on self-reported confidence. Two of PEAR's three routing components, Targeted Diversity and Low-Confidence Filtering, directly consume the self-reported confidence scores produced by the agents themselves. While LLM confidence is known to be imperfectly calibrated in some settings, our results suggest the router is reasonably robust to this in practice: routing decisions are based on relative confidence comparisons across agents rather than absolute thresholds, which partially mitigates the effect of systematic bias. Nonetheless, coupling PEAR with an external confidence estimator or a calibration-aware routing objective represents a promising direction for further improving routing precision.

Ethics Considerations

All experiments use publicly available academic benchmarks and involve no human subjects or sensitive data. We do not foresee any direct negative societal impacts arising from the proposed framework.

References

Anthropic. 2025. Claude Haiku 4.5. https://www.anthropic.com/claude/haiku.

Justin Chen, Swarnadeep Saha, and Mohit Bansal. 2024. Reconcile: Round-table conference improves reasoning via consensus among diverse llms. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7066–7085.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, and 1 others. 2021. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. 2024. Improving factuality and reasoning in language models through multiagent debate. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 11733–11763. PMLR.

Yichao Fu, Xuewei Wang, Yuandong Tian, and Jiawei Zhao. 2025. Deep think with confidence. *arXiv preprint arXiv:2508.15260*.

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*.

Cong Guan, Feng Chen, Lei Yuan, Zongzhang Zhang, and Yang Yu. 2024. Efficient communication via self-supervised information aggregation for online and offline multiagent reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(5):9044–9056.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*.

Shengchao Hu, Li Shen, Ya Zhang, and Dacheng Tao. 2024. Learning multi-agent communication from graph modeling perspective. *arXiv preprint arXiv:2405.08550*.

Boyi Li, Zhonghan Zhao, Der-Horng Lee, and Gaoang Wang. 2025. Adaptive graph pruning for multi-agent communication. *arXiv preprint arXiv:2506.02951*.

Yunxuan Li, Yibing Du, Jiageng Zhang, Le Hou, Peter Grabowski, Yeqing Li, and Eugene Ie. 2024. Improving multi-agent debate with sparse communication topology. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7281–7294.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17889–17904. Association for Computational Linguistics.

Stephanie Lin, Jacob Hilton, and Owain Evans. 2022. Truthfulqa: Measuring how models mimic human falsehoods. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers)*, pages 3214–3252.

Zijie Lin and Bryan Hooi. 2025. Enhancing multi-agent debate system performance via confidence expression. *arXiv preprint arXiv:2509.14034*.

Tongxuan Liu, Xingyu Wang, Weizhe Huang, Wenjiang Xu, Yuting Zeng, Lei Jiang, Hailong Yang, and Jing Li. 2024. Groupdebate: Enhancing the efficiency of multi-agent debate using group discussion. *arXiv preprint arXiv:2409.14051*.

Manh Nguyen, Anh Nguyen, Dung Nguyen, Svetha Venkatesh, and Hung Le. 2026. Hear both sides: Efficient multi-agent debate via diversity-aware message retention. *arXiv preprint arXiv:2603.20640*.

OpenAI. 2026. Introducing GPT-5.4 mini and nano. https://openai.com/index/introducing-gpt-5-4-mini-and-nano/.

Li Shen, Guibin Zhang, Yanwei Yan, and Kun Wang. 2025. Assemble your crew: Automatic multi-agent communication topology design via autoregressive graph generation. *arXiv preprint arXiv:2507.18224*.

Chuxiong Sun, Zehua Zang, Jiabao Li, Jiangmeng Li, Xiao Xu, Rui Wang, and Changwen Zheng. 2024. T2mac: Targeted and trusted multi-agent communication through selective engagement and evidence-driven integration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 15154–15163.

Amir Taubenfeld, Tom Sheffer, Eran Ofek, Amir Feder, Ariel Goldstein, Zorik Gekhman, and Gal Yona. 2025. Confidence improves self-consistency in llms. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 20090–20111.

Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, and 197 others. 2025. Gemma 3 technical report. *Preprint*, arXiv:2503.19786.

Hongduan Tian, Xiao Feng, Ziyuan Zhao, Xiangyu Zhu, Rolan Yan, and Bo Han. 2026. Multi-agent debate with memory masking. *arXiv preprint arXiv:2603.20215*.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, and 1 others. 2024. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark. *Advances in Neural Information Processing Systems*, 37:95266–95290.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Guibin Zhang, Yanwei Yue, Xiangguo Sun, Guancheng Wan, Miao Yu, Junfeng Fang, Kun Wang, Tianlong Chen, and Dawei Cheng. 2024. G-designer: Architecting multi-agent communication topologies via graph neural networks. *arXiv preprint arXiv:2410.11782*.

Qian Zhang, Jinyi Liu, Yan Zheng, Hebin Liang, and Lanjun Wang. 2026. Key decision-makers in multi-agent debates: Who holds the power? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 29883–29891.

Li Zhou, Xinfeng Deng, Zhe Wang, Xiaoying Zhang, Yanjie Dong, Xiping Hu, Zhaolong Ning, and Jibo Wei. 2025. Semantic information extraction and multi-agent communication optimization based on generative pre-trained transformer. *IEEE Transactions on Cognitive Communications and Networking*, 11(2):725–737.

Xiaochen Zhu, Caiqi Zhang, Yizhou Chi, Tom Stafford, Nigel Collier, and Andreas Vlachos. 2026. Demystifying multi-agent debate: The role of confidence and diversity. *arXiv preprint arXiv:2601.19921*.

A Proofs

A.1 Proof of Theorem 1

Proof. For a debate state s, let $y_i(s)$, $c_i(s)$, and $\rho_i(s)$ denote the answer, confidence, and accumulated influence of agent i. Under relabeling by σ,

$$y_{\sigma i}(\sigma s) = y_i(s), \qquad c_{\sigma i}(\sigma s) = c_i(s), \qquad \rho_{\sigma i}(\sigma s) = \rho_i(s).$$

For any edge $(u, v) \in E$, the targeted-diversity indicator satisfies

$$\begin{aligned} T(\sigma u, \sigma v \mid \sigma s) &= \mathbf{1}\{y_{\sigma u}(\sigma s) \neq y_{\sigma v}(\sigma s)\}\mathbf{1}\{c_{\sigma u}(\sigma s) \geq \tau_{\text{src}}\}\mathbf{1}\{c_{\sigma v}(\sigma s) \leq \tau_{\text{tgt}}\} \\ &= \mathbf{1}\{y_u(s) \neq y_v(s)\}\mathbf{1}\{c_u(s) \geq \tau_{\text{src}}\}\mathbf{1}\{c_v(s) \leq \tau_{\text{tgt}}\} \\ &= T(u, v \mid s). \end{aligned}$$

Since $\sigma : E \to \sigma E$ is a bijection and all candidates have m edges,

$$\widetilde{T}(\sigma E \mid \sigma s) = \frac{1}{m} \sum_{(\sigma u, \sigma v) \in \sigma E} T(\sigma u, \sigma v \mid \sigma s) = \widetilde{T}(E \mid s).$$

For the influence term, graph relabeling gives $d^{\text{out}}_{\sigma u}(\sigma E) = d^{\text{out}}_u(E)$. Hence

$$\begin{aligned} \widetilde{I}(\sigma E \mid \sigma s) &= \frac{1}{m} \sum_{i=1}^{n} \rho_i(\sigma s) d^{\text{out}}_i(\sigma E) \\ &= \frac{1}{m} \sum_{u=1}^{n} \rho_{\sigma u}(\sigma s) d^{\text{out}}_{\sigma u}(\sigma E) \\ &= \frac{1}{m} \sum_{u=1}^{n} \rho_u(s) d^{\text{out}}_u(E) = \widetilde{I}(E \mid s). \end{aligned}$$

Similarly, because $L(\sigma u \mid \sigma s) = L(u \mid s)$,

$$\widetilde{L}(\sigma E \mid \sigma s) = \widetilde{L}(E \mid s).$$

Combining the three identities yields

$$S_\alpha(\sigma E \mid \sigma s) = S_\alpha(E \mid s).$$

It remains to verify the softmax distribution. By closure of the candidate family, the map $E' \mapsto \sigma E'$ is a bijection from $\mathcal{C}(s)$ to $\mathcal{C}(\sigma s)$. Therefore

$$\begin{aligned} Q_\alpha(\sigma E \mid \sigma s) &= \frac{\exp(S_\alpha(\sigma E \mid \sigma s)/\tau)}{\sum_{F \in \mathcal{C}(\sigma s)} \exp(S_\alpha(F \mid \sigma s)/\tau)} \\ &= \frac{\exp(S_\alpha(E \mid s)/\tau)}{\sum_{E' \in \mathcal{C}(s)} \exp(S_\alpha(\sigma E' \mid \sigma s)/\tau)} \\ &= \frac{\exp(S_\alpha(E \mid s)/\tau)}{\sum_{E' \in \mathcal{C}(s)} \exp(S_\alpha(E' \mid s)/\tau)} = Q_\alpha(E \mid s). \end{aligned}$$

For $\tau \to 0$, the softmax concentrates on the maximizers of $S_\alpha(\cdot \mid s)$. If ties are broken uniformly among maximizers, the same bijection argument applies to the argmax set, giving equivariant deterministic routing. □

A.2 Proof of Theorem 2

Proof. First, we show that $\mathcal{P}q$ is equivariant. For any $\sigma_0 \in \mathfrak{S}_n$,

$$(\mathcal{P}q)(\sigma_0\Gamma \mid \sigma_0 s^{(0)}) = \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} q(\sigma\sigma_0\Gamma \mid \sigma\sigma_0 s^{(0)}).$$

As σ ranges over $\mathfrak{S}_n$, so does $\sigma\sigma_0$. Thus

$$(\mathcal{P}q)(\sigma_0\Gamma \mid \sigma_0 s^{(0)}) = \frac{1}{n!}\sum_{\gamma\in\mathfrak{S}_n} q(\gamma\Gamma \mid \gamma s^{(0)}) = (\mathcal{P}q)(\Gamma \mid s^{(0)}).$$

Now consider expected accuracy:

$$\begin{aligned}\mathbb{E}_{\mathcal{P}q}[A] &= \mathbb{E}_{s^{(0)}}\sum_{\Gamma}(\mathcal{P}q)(\Gamma \mid s^{(0)})A(\Gamma, s^{(0)}) \\ &= \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\mathbb{E}_{s^{(0)}}\sum_{\Gamma} q(\sigma\Gamma \mid \sigma s^{(0)})A(\Gamma, s^{(0)}).\end{aligned}$$

Let $\Gamma' = \sigma\Gamma$ and $s' = \sigma s^{(0)}$. Since the distribution of $s^{(0)}$ is exchangeable, s' has the same distribution as $s^{(0)}$. Permutation-equivariance of the update kernels and permutation-invariance of the aggregation rule imply

$$A(\Gamma, s^{(0)}) = A(\sigma\Gamma, \sigma s^{(0)}) = A(\Gamma', s').$$

Therefore every summand equals

$$\mathbb{E}_{s'}\sum_{\Gamma'} q(\Gamma' \mid s')A(\Gamma', s'),$$

which is exactly the expected accuracy under q. Averaging over σ leaves the same quantity. □

A.3 Proofs of Lemma 1 and Theorem 3

Definition 1 (ϵ-covering number)**.** *Let $(\mathcal{F}, d)$ be a metric space. A set $\mathcal{C}_\epsilon \subseteq \mathcal{F}$ is an ϵ-cover of $\mathcal{F}$ if, for every $f \in \mathcal{F}$, there exists $g \in \mathcal{C}_\epsilon$ such that $d(f, g) \le \epsilon$. The covering number is*

$$\mathcal{N}(\mathcal{F}, \epsilon, d) = \min\left\{|\mathcal{C}_\epsilon| : \mathcal{C}_\epsilon \text{ is an } \epsilon\text{-cover of } \mathcal{F}\right\}.$$

Proof of Lemma 1. We first show that $\mathcal{P}$ is non-expansive under

$$d(q, q') = \sup_s \|q(\cdot \mid s) - q'(\cdot \mid s)\|_1.$$

For any state s,

$$\begin{aligned}&\|(\mathcal{P}q)(\cdot \mid s) - (\mathcal{P}q')(\cdot \mid s)\|_1 \\ &= \left\|\frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\left[q(\sigma\cdot \mid \sigma s) - q'(\sigma\cdot \mid \sigma s)\right]\right\|_1 \\ &\le \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\left\|q(\sigma\cdot \mid \sigma s) - q'(\sigma\cdot \mid \sigma s)\right\|_1 \\ &\le d(q, q').\end{aligned}$$

Taking the supremum over s gives

$$d(\mathcal{P}q, \mathcal{P}q') \le d(q, q').$$

Let $\{q_1, \ldots, q_M\}$ be an ϵ-cover of $\mathcal{Q}$. For any $\mathcal{P}q \in \mathcal{P}\mathcal{Q}$, there exists q_j such that $d(q, q_j) \le \epsilon$. By non-expansiveness,

$$d(\mathcal{P}q, \mathcal{P}q_j) \le \epsilon.$$

Thus

$$\{\mathcal{P}q_1, \ldots, \mathcal{P}q_M\}$$

is an ϵ-cover of $\mathcal{P}\mathcal{Q}$. Taking the smallest possible M proves

$$\mathcal{N}(\mathcal{P}\mathcal{Q}, \epsilon, d) \leq \mathcal{N}(\mathcal{Q}, \epsilon, d).$$

□

Proof of Theorem 3. Let

$$M = \mathcal{N}(\mathcal{P}\mathcal{Q}, \epsilon, d),$$

and let

$$\{q_1, \ldots, q_M\}$$

be an ϵ-cover of $\mathcal{P}\mathcal{Q}$. For any fixed q_j, Hoeffding's inequality gives

$$\Pr\left(\left|\mathcal{L}(q_j) - \widehat{\mathcal{L}}_{N_{\text{ex}}}(q_j)\right| > t\right) \leq 2\exp(-2N_{\text{ex}}t^2),$$

because $F(q_j; z) \in [0, 1]$. A union bound over the cover implies that, with probability at least $1 - \delta$,

$$\max_{1\leq j\leq M}\left|\mathcal{L}(q_j) - \widehat{\mathcal{L}}_{N_{\text{ex}}}(q_j)\right| \leq \sqrt{\frac{\log(2M/\delta)}{2N_{\text{ex}}}}.$$

Now fix any $q \in \mathcal{P}\mathcal{Q}$, and choose q_j from the cover such that $d(q, q_j) \leq \epsilon$. By the L-Lipschitz condition,

$$|\mathcal{L}(q) - \mathcal{L}(q_j)| \leq L\epsilon, \qquad |\widehat{\mathcal{L}}_{N_{\text{ex}}}(q) - \widehat{\mathcal{L}}_{N_{\text{ex}}}(q_j)| \leq L\epsilon.$$

Therefore

$$\begin{aligned}\left|\mathcal{L}(q) - \widehat{\mathcal{L}}_{N_{\text{ex}}}(q)\right| &\leq |\mathcal{L}(q) - \mathcal{L}(q_j)| + |\mathcal{L}(q_j) - \widehat{\mathcal{L}}_{N_{\text{ex}}}(q_j)| \\ &\quad + |\widehat{\mathcal{L}}_{N_{\text{ex}}}(q_j) - \widehat{\mathcal{L}}_{N_{\text{ex}}}(q)| \\ &\leq 2L\epsilon + \sqrt{\frac{\log(2M/\delta)}{2N_{\text{ex}}}}.\end{aligned}$$

Taking the supremum over $q \in \mathcal{P}\mathcal{Q}$ proves the first bound. The second follows immediately from Lemma 1. □

B One-Round Correction and Influence-Balancing

This appendix contains the one-round correction and influence-balancing results that support the interpretation of the PEAR score.

B.1 One-round correction lower bound

Let

$$Z_i^{(r)} = \mathbf{1}\{y_i^{(r)} = y^\star\}$$

denote whether agent i is correct after round r. Define the expected one-round improvement under candidate edge set E and state s by

$$\Delta(E; s) = \mathbb{E}\left[\sum_{i=1}^{n} Z_i^{(r)} - \sum_{i=1}^{n} Z_i^{(r-1)} \,\middle|\, E, s\right].$$

Assumption 1 (Additive edge-wise improvement model)**.** *For each candidate edge set E, the expected improvement at each target is the sum of marginal contributions from its incoming edges:*

$$\mathbb{E}\left[Z_v^{(r)} - Z_v^{(r-1)} \,\middle|\, E, s\right] = \sum_{u:(u,v)\in E} \delta_{uv}(s).$$

Moreover, there exist constants $\lambda_T, \lambda_I, \lambda_L \geq 0$ and $\varepsilon \geq 0$ such that every marginal contribution satisfies

$$\delta_{uv}(s) \geq \lambda_T T(u, v \mid s) - \lambda_I \rho_u(s) - \lambda_L \frac{L(u \mid s)}{\tau_{\text{low}}} - \varepsilon.$$

The assumption encodes three effects: confident disagreement toward uncertain targets is helpful, low-confidence sources are unreliable, and repeatedly amplified sources have diminishing marginal value.

Theorem 4 (One-round correction lower bound)**.** *Under Assumption 1, for any candidate edge set E,*

$$\Delta(E; s) \geq m\left(\lambda_T \widetilde{T}(E \mid s) - \lambda_I \widetilde{I}(E \mid s) - \lambda_L \widetilde{L}(E \mid s) - \varepsilon\right).$$

If PEAR uses weights $\alpha = (\lambda_T, \lambda_I, \lambda_L)$ and samples $E \sim Q_\alpha(\cdot \mid s)$, then

$$\mathbb{E}_{E\sim Q_\alpha}[\Delta(E; s)] \geq m\left[\max_{E'\in\mathcal{C}(s)} S_\alpha(E' \mid s) - \tau \log|\mathcal{C}(s)| - \varepsilon\right].$$

Thus, as $\tau \to 0$, PEAR greedily maximizes a lower bound on expected one-round correction.

Proof. Assumption 1 models the improvement at each target v as a sum of marginal contributions over its incoming edges. Therefore,

$$\Delta(E; s) = \sum_{(u,v)\in E} \delta_{uv}(s).$$

Using the marginal lower bound and summing over edges gives

$$\begin{aligned}
\Delta(E; s) &\geq \sum_{(u,v)\in E}\left[\lambda_T T(u, v \mid s) - \lambda_I \rho_u(s) - \lambda_L \frac{L(u \mid s)}{\tau_{\text{low}}} - \varepsilon\right] \\
&= m\lambda_T \widetilde{T}(E \mid s) - m\lambda_I \widetilde{I}(E \mid s) - m\lambda_L \widetilde{L}(E \mid s) - m\varepsilon \\
&= m\left(\lambda_T \widetilde{T}(E \mid s) - \lambda_I \widetilde{I}(E \mid s) - \lambda_L \widetilde{L}(E \mid s) - \varepsilon\right).
\end{aligned}$$

For the softmax guarantee, write

$$a_E = S_\alpha(E \mid s), \qquad p_E = Q_\alpha(E \mid s).$$

The Gibbs variational identity gives

$$\tau \log \sum_{E\in\mathcal{C}(s)} \exp(a_E/\tau) = \sum_E p_E a_E + \tau H(p),$$

where $H(p) = -\sum_E p_E \log p_E$. Since

$$\tau \log \sum_E \exp(a_E/\tau) \geq \max_E a_E \qquad \text{and} \qquad H(p) \leq \log|\mathcal{C}(s)|,$$

we obtain

$$\sum_E p_E a_E \geq \max_E a_E - \tau \log|\mathcal{C}(s)|.$$

Taking expectation in the first part of the theorem with $E \sim Q_\alpha(\cdot \mid s)$ proves

$$\mathbb{E}_{E\sim Q_\alpha}[\Delta(E; s)] \geq m\left[\max_{E'\in\mathcal{C}(s)} S_\alpha(E' \mid s) - \tau \log|\mathcal{C}(s)| - \varepsilon\right].$$

□

B.2 Influence monotonicity

Let

$$D_j(E) = d_j^{\text{out}}(E)$$

be the number of targets receiving critiques from source agent j.

Theorem 5 (Influence penalty reduces future amplification)**.** *For softmax routing with $\tau > 0$,*

$$\frac{\partial}{\partial \rho_j} \mathbb{E}_{E \sim Q_\alpha(\cdot|s)}[D_j(E)] = -\frac{\alpha_I}{\tau m} \operatorname{Var}_{E \sim Q_\alpha(\cdot|s)}[D_j(E)] \leq 0.$$

Thus, increasing an agent's accumulated influence can only decrease its expected future out-degree.

Proof. Fix a state s and write the score as

$$S_\alpha(E \mid s) = B(E \mid s) - \frac{\alpha_I}{m} \sum_{i=1}^{n} \rho_i D_i(E),$$

where $B(E \mid s)$ contains the targeted-diversity and low-confidence terms, which do not depend on ρ_j. The softmax distribution is

$$Q_\alpha(E \mid s) = \frac{\exp(S_\alpha(E \mid s)/\tau)}{Z(s)}.$$

By the score-function, or log-derivative, identity, for any statistic $f(E)$,

$$\frac{\partial}{\partial \rho_j} \mathbb{E}_{Q_\alpha}[f(E)] = \operatorname{Cov}_{Q_\alpha}\left(f(E), \frac{\partial}{\partial \rho_j} \log Q_\alpha(E \mid s) \right).$$

Moreover,

$$\frac{\partial}{\partial \rho_j} \log Q_\alpha(E \mid s) = -\frac{\alpha_I}{\tau m} D_j(E) - \frac{\partial}{\partial \rho_j} \log Z(s).$$

The second term is constant in E and therefore vanishes inside the covariance. Taking $f(E) = D_j(E)$ yields

$$\begin{aligned} \frac{\partial}{\partial \rho_j} \mathbb{E}_{Q_\alpha}[D_j(E)] &= -\frac{\alpha_I}{\tau m} \operatorname{Cov}_{Q_\alpha}(D_j(E), D_j(E)) \\ &= -\frac{\alpha_I}{\tau m} \operatorname{Var}_{Q_\alpha}(D_j(E)). \end{aligned}$$

Since $\alpha_I \geq 0$, $\tau > 0$, $m > 0$, and variance is nonnegative, the derivative is nonpositive. □

C Additional Experimental Details

This appendix presents additional experimental details.

C.1 Debate Configuration

Debate protocol. Our experiments use $n = 5$ agents and $R = 5$ debate rounds. Each round consists of a critique phase followed by an answer-revision phase. Final predictions are aggregated using majority voting. For fixed-topology baselines, we evaluate clique, ring, star, and chain communication structures. For PEAR, the base interaction graph is a k-regular graph with degree $k = 2$, and the topology is adaptively reconfigured at each round through role-to-agent routing permutations. The maximum generation budget is 512 tokens per model call.

Routing-objective weights. The composite routing score is

$$S(E \mid s_r) = \alpha_T \widetilde{T}(E \mid s_r) - \alpha_I \widetilde{I}(E) - \alpha_L \widetilde{L}(E \mid s_r), \tag{9}$$

where $\widetilde{T}$, $\widetilde{I}$, and $\widetilde{L}$ are the normalized routing components defined in Equations 2, 3, and 5, respectively. These quantities are normalized through bounded averaging over the candidate edge set and therefore lie in $[0, 1]$. Unless otherwise specified, we set $(\alpha_T, \alpha_I, \alpha_L) = (0.4, 0.7, 0.7)$. The asymmetric weighting reflects a deliberate priority among the three components: we first ensure that routing is balanced in influence ($\alpha_I = 0.7$) and gated by source confidence ($\alpha_L = 0.7$), since these two terms directly address the structural failure modes of fixed-topology debate, namely positional dominance and unreliable senders. Targeted diversity ($\alpha_T = 0.4$) then acts as a refinement layered on top of this stabilized backbone, encouraging cross-answer routing once influence and confidence are already controlled, rather than as a primary driver that could over-amplify divergent but noisy critiques. These weights are held fixed across rounds and across datasets.

Confidence and influence smoothing. Self-reported confidence is solicited on the ordinal scale $c_i^{(r)} \in \{1, \ldots, 5\}$. The accumulated-influence statistic is updated with smoothing coefficient $\beta = 0.5$ via

$$\rho_s^{(r)} = \beta \rho_s^{(r-1)} + (1 - \beta)\, a_s^{(r)}, \quad \rho_s^{(0)} = 0,$$

where $a_s^{(r)} \in [0, 1]$ is the per-round adoption rate computed from the ACCEPT/REJECT decisions in agents' revisions. Confidence and source thresholds are $\tau_{\text{src}} = 4$, $\tau_{\text{tgt}} = 3$, and $\tau_{\text{low}} = 2$ on the 1–5 scale.

Candidate pool and routing selection. At each round, we construct a candidate pool $\mathcal{C}_r \subseteq S_n$ by uniformly sampling role-to-agent permutations without replacement subject to the base graph structure. The candidate pool defines the finite search space over which routing scores are evaluated.

For each candidate assignment $\pi \in \mathcal{C}_r$, we compute the composite routing score $S(E(\pi) \mid s_r)$. The router then samples the routing assignment according to the state-conditional softmax distribution

$$Q_r(\pi \mid s_r) = \frac{\exp\big(S(E(\pi) \mid s_r)/\tau\big)}{\sum_{\pi' \in \mathcal{C}_r} \exp\big(S(E(\pi') \mid s_r)/\tau\big)},$$

where τ is the routing temperature. As $\tau \to 0$, the routing policy approaches greedy argmax selection.

Compute infrastructure. All experiments are conducted on a single node equipped with 4 NVIDIA H100 80GB GPUs. Each experiment is run with data-parallel or sequential execution depending on the setting. This setup ensures reproducibility under a fixed compute budget.

C.2 Full-Dataset Results

Table 2 reports results on the full evaluation sets for two representative open-source models. Compared with the results in Section 6, which uses sampled subsets for computational efficiency, these experiments evaluate all methods on the complete TruthfulQA and MATH-500 benchmarks.

Across both datasets and models, PEAR consistently achieves the best final-answer accuracy. On TruthfulQA, the gains over fixed-topology baselines remain substantial, indicating that adaptive communication continues to improve factual robustness even at full evaluation scale. On MATH-500, where multi-step reasoning errors frequently propagate through the debate process, PEAR also maintains a clear advantage over static topologies, particularly compared with star and ring structures.

These results demonstrate that the improvements observed in the main experiments are not artifacts of subset evaluation and remain stable when scaling to the full datasets.

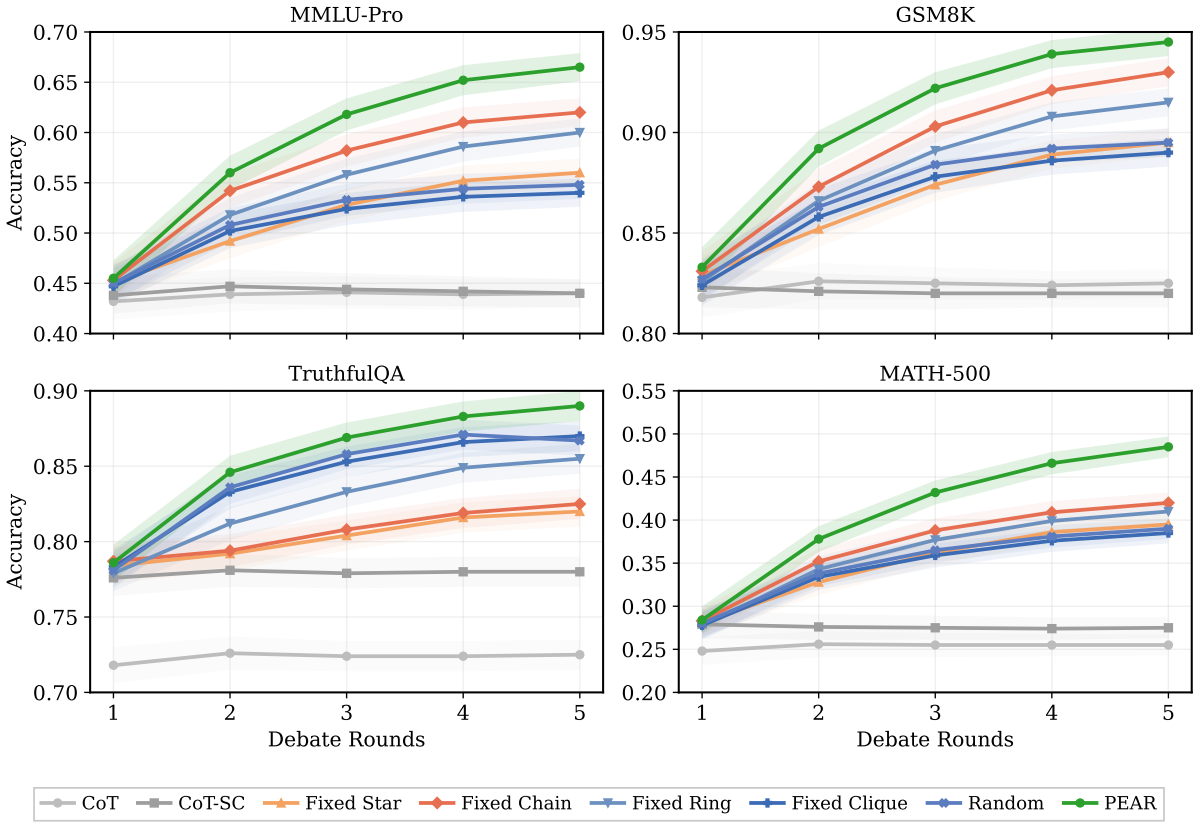

Figure 3: Round-wise accuracy on Qwen-2.5-14B, averaged across MMLU-Pro, TruthfulQA, GSM8K, and MATH-500. PEAR continues to improve with additional debate rounds, while fixed-topology baselines plateau earlier.

Table 2: Final-answer accuracy on full datasets.

Method	TruthfulQA		MATH-500	
	Qwen2.5-14B	Llama-3.1-8B	Qwen2.5-14B	Llama-3.1-8B
CoT	0.709	0.653	0.212	0.138
CoT-SC	0.785	0.675	0.256	0.142
Clique	0.857	0.785	0.420	0.206
Star	0.834	0.722	0.334	0.208
Chain	0.795	0.745	0.348	0.290
Ring	0.827	0.757	0.344	0.218
PEAR	**0.868**	**0.824**	**0.442**	**0.298**

C.3 Round-wise Accuracy Curves

Figures 3 and 4 plot final-answer accuracy as a function of debate round number for PEAR and the topology baselines on Qwen-2.5-14B and GPT-5.4-nano, respectively. Both panels report accuracy averaged across the four benchmarks. PEAR remains above the baseline curves after the first round and continues to improve with additional rounds, whereas fixed-topology baselines saturate earlier or exhibit non-monotonic behavior consistent with error cascades.

Initial round versus later rounds. At round 1 all debate methods are tightly clustered (within roughly 2 accuracy points), since the first round only consumes the initial independent responses and the routing decision has not yet been informed by any debate state. The gap between PEAR and the fixed-topology baselines then widens monotonically in subsequent rounds, indicating that the adaptive router accumulates benefit from the evolving state rather than from a single early-round choice.

Non-monotonic baselines and error cascades. Several fixed-topology curves are non-monotonic—for instance, Fixed Star peaks at round 2 and retreats slightly at round 3 on both backbones. This is the error-cascade mechanism in action: an initially incorrect hub influences its neighbors, and later rounds reinforce rather than correct the resulting consensus. PEAR's curves remain non-decreasing in every panel, consistent with the influence-balancing component progressively dampening any agent that begins to dominate.

C.4 Trajectory-Level Diagnostics

This section expands on the trajectory-level diagnostics reported in Table 3. We first give precise definitions for each diagnostic metric, then discuss

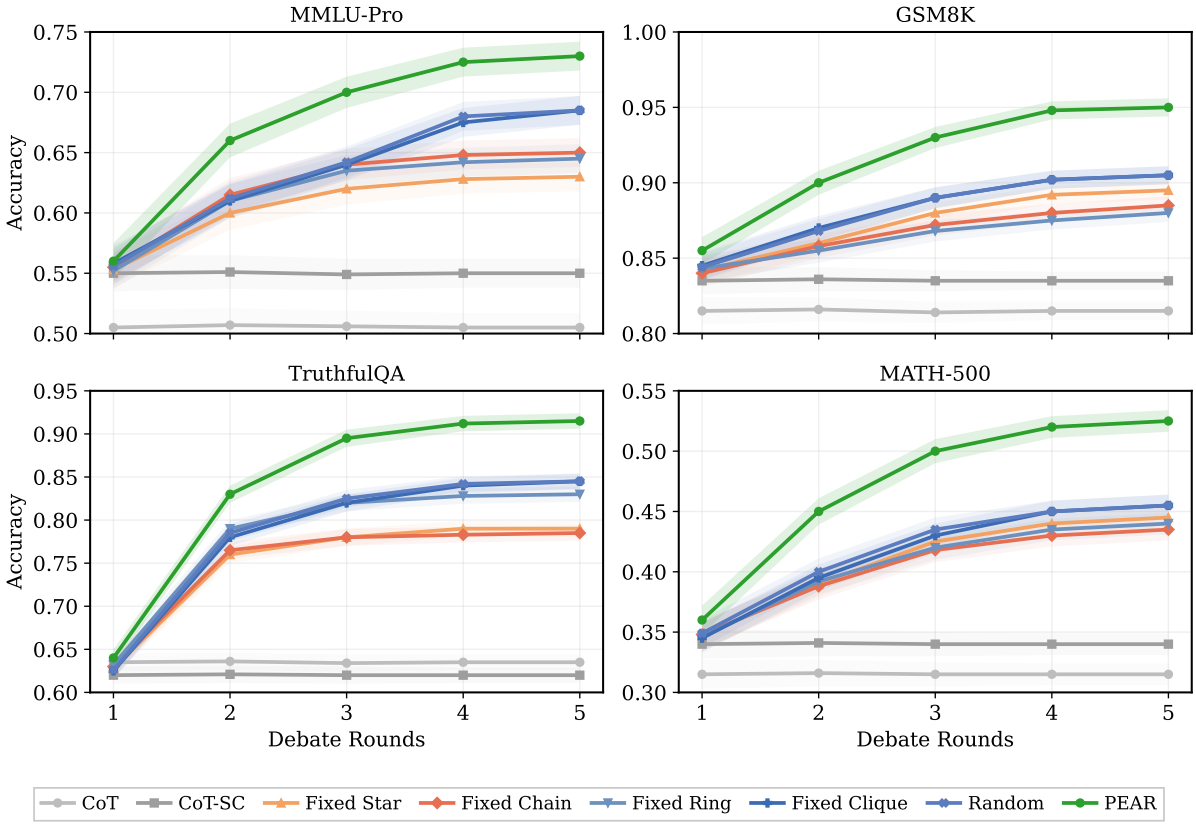

Figure 4: Round-wise accuracy on GPT-5.4-nano, averaged across MMLU-Pro, TruthfulQA, GSM8K, and MATH-500. PEAR maintains a consistent margin over baselines across all rounds.

additional patterns that complement the main-text summary.

Metric definitions. For each agent $i \in [n]$ and round $r \in \{1, \ldots, R\}$, let $y_i^{(r)}$ be the agent's current answer, $c_i^{(r)} \in \{1, \ldots, 5\}$ its self-reported confidence, $\rho_i^{(r)} \in [0, 1]$ the router-side accumulated influence, and $y^\star$ the gold answer. Let $E^{(r)}$ denote the round-r edge set. All quantities below are first computed per query and then averaged over the open-source model–dataset settings.

Update-level metrics. The wrong-to-right rate is the fraction of agent updates that change an initially wrong answer into the correct one,

$$\mathrm{W2R} = \frac{\sum_{i,r} \mathbf{1}\left[y_i^{(r-1)} \neq y^\star,\ y_i^{(r)} = y^\star\right]}{\sum_{i,r} \mathbf{1}\left[y_i^{(r-1)} \neq y^\star\right]},$$

and the right-to-wrong rate R2W is defined symmetrically by swapping correct and incorrect. The *net correction rate* is $\mathrm{Net} = \mathrm{W2R} - \mathrm{R2W}$; positive values mean that beneficial updates outweigh harmful flips. The *critique acceptance rate* (Accept.) is the fraction of incoming critiques that the target marks ACCEPT (rather than REJECT) in its structured `critique_response`, providing a target-side validation of each routed argument.

Edge-level metrics. The *cross-answer edge rate* measures whether routing exposes targets to genuinely different viewpoints,

$$\text{Cross-Ans} = \frac{1}{\sum_r |E^{(r)}|} \sum_r \sum_{(s \to t) \in E^{(r)}} \mathbf{1}\left(y_s^{(r-1)} \neq y_t^{(r-1)}\right).$$

Source confidence (Src Conf) is the mean self-reported confidence on the source side of each routed edge, normalized to $[0, 1]$ via $(c - 1)/4$; higher values indicate that critiques disproportionately originate from confident agents.

System-level metric. The *influence entropy* (Inf. Ent.) is the normalized Shannon entropy of the

Table 3: Trajectory-level diagnostics, averaged over open-source model–dataset settings. W2R / R2W are wrong-to-right and right-to-wrong update rates; Net = W2R − R2W is the net correction rate. Accept. is the critique acceptance rate. Cross-Ans is the fraction of routed edges whose source and target hold different answers, capturing targeted diversity in practice. Src Conf is the source-side mean self-reported confidence, normalized to $[0, 1]$. Inf. Ent. is the normalized influence entropy.

Condition	W2R	R2W	Net	Accept.	Cross-Ans	Src Conf	Inf. Ent.
CoT	0.000	0.000	0.000	–	–	–	0.000
CoT-SC	0.000	0.000	0.000	–	–	–	1.000
Fixed Star	0.324	0.112	0.212	0.567	0.431	0.592	0.961
Fixed Chain	0.330	0.127	0.203	0.556	0.452	0.601	0.827
Fixed Ring	0.315	0.111	0.204	0.593	0.438	0.594	0.942
Fixed Clique	0.326	0.118	0.208	0.597	0.448	0.589	0.969
Random	0.318	0.115	0.203	0.585	0.498	0.612	0.972
PEAR	**0.339**	**0.096**	**0.243**	**0.606**	**0.676**	**0.784**	**0.979**

final-round influence distribution,

$$\text{Inf. Ent.} = -\frac{1}{\log n} \sum_{i=1}^{n} \tilde{\rho}_i \log \tilde{\rho}_i,$$
$$\tilde{\rho}_i = \rho_i^{(R)} \Big/ \sum_j \rho_j^{(R)}.$$

A value of 1 indicates uniform influence across agents; values close to 0 indicate that a single agent dominates the debate.

Concentrated information flow in Fixed Chain. Fixed Chain attains a competitive Net (0.203) but the lowest influence entropy among all debate baselines (0.827, vs. ≥ 0.94 for the others). Because critiques in a chain flow strictly toward the head, the head agent accumulates disproportionate influence regardless of correctness. This is consistent with the failure mode described in Section 1: when the structurally privileged agent happens to be wrong, errors propagate persistently downstream, and the aggregate Net masks the underlying brittleness.

Diversity is necessary but not sufficient. Random routing achieves the highest Cross-Ans rate among baselines (0.498), confirming that simply breaking the fixed topology improves viewpoint exposure. Yet Random still trails PEAR by 18 points on Cross-Ans (0.498 vs. 0.676), and its accuracy and Net are no better than the other fixed topologies. *Indiscriminate* diversity therefore does not translate into corrective gains; targeted diversity, routing from confident dissenting sources to uncertain targets, is what makes the difference.

Confidence and acceptance reinforce one another under PEAR. For every baseline, the source-side mean confidence (Src Conf) falls in the narrow band $[0.589, 0.612]$, essentially indistinguishable from the population average. PEAR's value of 0.784 is the only one substantially above this band, indicating that low-confidence filtering effectively biases the routing distribution toward confident sources. Mirroring this, PEAR obtains the highest critique acceptance rate (0.606): because routed critiques disproportionately come from agents with substantive corrections to offer, targets accept them more often than under any baseline.

Why CoT-SC's entropy is not comparable. CoT and CoT-SC are reported in Table 3 for reference; both have $\text{W2R} = \text{R2W} = 0$ because no debate updates occur. CoT-SC's nominal $\text{Inf. Ent.} = 1.000$ reflects fully independent samples without any shared influence flow and is not directly comparable with PEAR's 0.979, which is the highest among methods that actually route influence between agents.

C.5 Computational Overhead

This appendix details the per-example computational cost of PEAR relative to the topology baselines. Token usage is computed per query as the sum of input and output tokens across all LLM calls in a debate, averaged over the same model–dataset settings as Table 1.

Figure 5 plots accuracy against per-example token usage (log scale) on all four benchmarks. PEAR (green star) lies on the Pareto frontier of every panel: no fixed-topology baseline reaches its

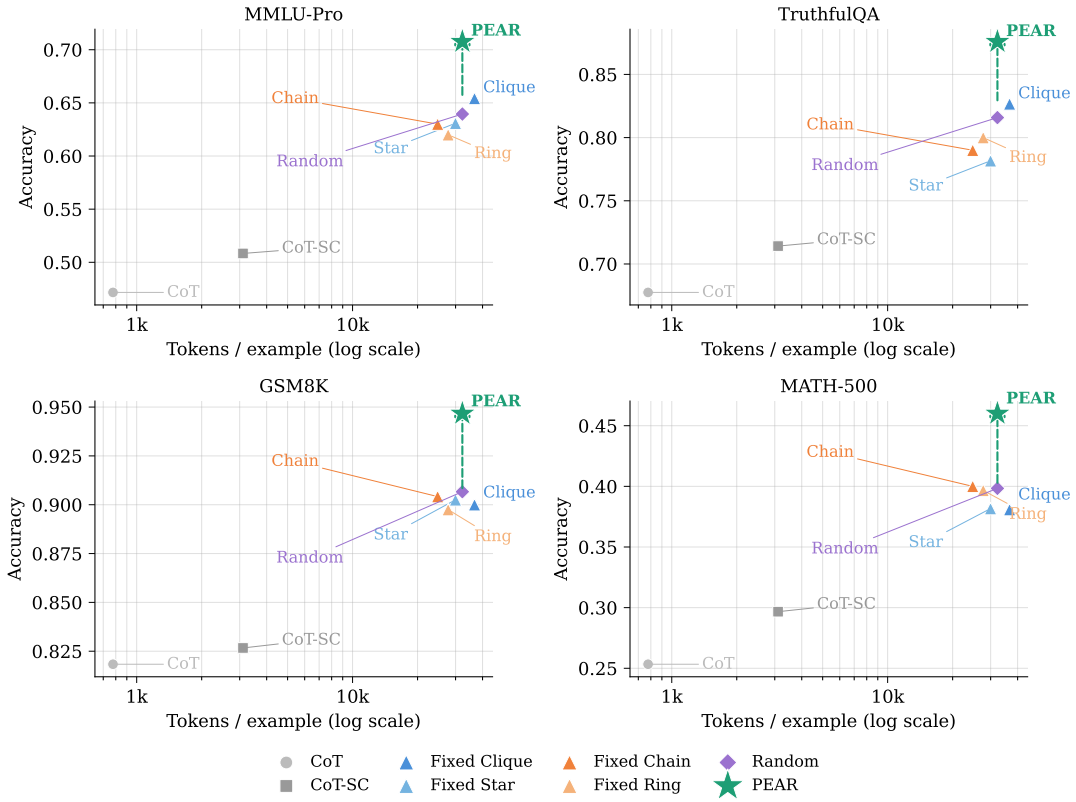

Figure 5: Accuracy versus per-example token usage (log scale) on the four benchmarks. Each panel reports accuracy averaged across the six backbone models. PEAR (green star) lies on the Pareto frontier of every panel; the dashed bracket marks the accuracy lift of PEAR over the best fixed-topology baseline at the same cost tier.

accuracy at any cost. PEAR also Pareto-dominates the most expensive baseline, Fixed Clique, matching or exceeding its accuracy on every dataset while using 12% fewer tokens (32.3k vs. 36.7k). At the other extreme, Fixed Chain reduces cost by 23% relative to PEAR (24.8k vs. 32.3k) but loses 6.6 accuracy points on average—a poor trade-off compared with simply running PEAR at the same sparse k-regular budget. Single-agent baselines (CoT and CoT-SC) are far cheaper at 0.8k–3.1k tokens but trail debate methods by 16–19 accuracy points and therefore lie well below the Pareto frontier.

Together with the trajectory diagnostics in Appendix C.4, these results confirm that PEAR's gains come from *how* it routes a fixed communication budget, not from spending more compute.

D Prompt Templates

In this section we provide the prompt templates used for agent reasoning and dataset-specific task formatting. Each template is designed to guide the LLM's behavior in specific contexts within the PEAR framework.

D.1 Agent System Instruction

```
You are a careful reasoner participating in
a structured multi-agent debate. Judge every
critique by the logic of the problem, not by
social agreement or majority pressure. Return
only valid JSON in the requested schema, with
no markdown fences and no extra commentary.
```

D.2 Confidence Rubric

```
Confidence rubric (integer 1-5):
1 = no reliable basis; mostly guessing, unable
to solve, or the selected answer is just a
placeholder.
2 = low confidence; some clue or partial
reasoning, but you cannot rule out multiple
plausible alternatives.
3 = moderate confidence; reasoning supports
the answer, but there is a real unresolved
doubt, unchecked step, or possible competing
option.
4 = high confidence; reasoning is complete and
checked, and only minor residual uncertainty
remains.
5 = fully verified; every necessary step has
been checked and all plausible alternatives or
answer choices are ruled out.
Use 1 or 2 whenever you are guessing, relying
on incomplete reasoning, or cannot eliminate
serious alternatives. Do not use 5 unless the
```

```
solution is fully verified; do not default to
4 or 5.
```

D.3 Initial Answer Template

```
You are Agent {agent_id}. Solve the problem
independently.
Use the task-specific answer format stated in
the PROBLEM block.

You should provide a step-by-step
justification for your answer. The reasoning
should be clear, logical, and directly support
your final answer.
After solving, assign a confidence score to
your own final answer using the 1-5 rubric
below.
Calibrate the score strictly: use low
confidence when your reasoning is incomplete
or competing answers remain plausible.

PROBLEM
{question}

{CONFIDENCE RUBRIC}

Return valid JSON with exactly these keys:
{
"answer": "task-specific answer token only",
"confidence": 3,
"reasoning": "concise step-by-step
justification"
}
The confidence field must be an integer from
1 to 5.
```

D.4 Answer Update Template

```
You are Agent {agent_id}. Update your answer
using only critiques that identify a real
error in your reasoning.

PROBLEM
{question}

{CONFIDENCE RUBRIC}

YOUR PREVIOUS ANSWER
Answer: {previous_answer}
Confidence: {previous_confidence}
Reasoning: {previous_reasoning}

CRITIQUES YOU RECEIVED
{critiques}

For each critique, explicitly ACCEPT or
REJECT it. Accept a critique only when
its correction is logically sound for this
problem.

After updating, assign a new confidence
score to your own updated answer using the
1-5 rubric above. Calibrate it strictly:
lower the score when accepted critiques leave
unresolved uncertainty or multiple plausible
answers.

Return valid JSON with exactly these keys:
{
"answer": "task-specific updated answer token
only",
"confidence": 3,
"reasoning": "updated step-by-step
justification",
"critique_response": {
"<source_agent_id>": {"decision": "ACCEPT",
"reason": "one sentence"}
}
}
Use only ACCEPT or REJECT. The confidence
field must be an integer from 1 to 5.
```

D.5 Critique Generation Template

```
You are Agent {agent_id}. Review the target
solutions below for logical correctness. Do
not judge by whether the target answer matches
your own; judge only by the reasoning steps.

PROBLEM
{question}

YOUR CURRENT ANSWER
Answer: {own_answer}
Confidence: {own_confidence}
Reasoning: {own_reasoning}

SOLUTIONS TO REVIEW
{targets}

For each target, identify the first incorrect
step if one exists. If no error is identified,
say so.

Return valid JSON with exactly this shape:
{
"reviews": [
{
"target": 1,
"step_loc": "first incorrect step, or No
error identified",
"correction": "correction for that step only,
or empty string",
"assessment": "Strong"
}
]
}
assessment must be one of Strong, Acceptable,
Flawed.
```

D.6 Per-Turn Agent Communication Template

```
You are Agent {agent_id} in a structured
multi-agent debate.
Use the task-specific answer format stated in
the PROBLEM block. Judge neighbor messages by
logic and evidence, not by majority pressure.

PROBLEM
{question}

{topology_info}
YOUR PRIVATE HISTORY
```

```
{private}

VISIBLE NEIGHBOR MESSAGES
{transcript}

Evaluate every distinct neighbor argument. ACCEPT only claims that fix a real error; REJECT claims with invalid logic, wrong facts, or irrelevant reasoning. If messages conflict, weigh the strongest argument on each side before updating your answer.

{CONFIDENCE RUBRIC}

Calibrate the score strictly: use low confidence when your reasoning is incomplete or competing answers remain plausible.

Return valid JSON with exactly these keys:
{
"answer": "task-specific answer token only",
"confidence": 3,
"reasoning": "concise justification of your current answer",
"neighbor_assessment": "one or two sentences naming accepted or rejected claims"
}
The confidence field must be an integer from 1 to 5.
```

D.7 Dataset-Specific Task Templates

D.7.1 GSM8K (Grade-School Math)

```
Task type: GSM8K grade-school math.
Answer format: provide only the final numeric value in the JSON answer field. Do not include units, commas, or explanatory text in the answer field.

Problem:
{question}
```

D.7.2 MMLU-Pro (Multiple-Choice)

```
Task type: MMLU-Pro multiple-choice.
Answer format: provide only the letter of the best option in the JSON answer field. Valid letters are those shown in Options.

Question:
{question}

Options:
{options}
```

D.7.3 MATH-500 (Competition Math)

```
Task type: MATH-500 competition math.
Answer format: provide only the final mathematical expression in the JSON answer field, in the form requested by the problem.

Problem:
{question}
```

D.7.4 TruthfulQA (Truthfulness-Based Choice)

```
Task type: TruthfulQA multiple-choice.
Answer format: provide only the letter of the most truthful option in the JSON answer field.

Question:
{question}

Options:
{options}
```

E Case Study of Multi-Round Debate Dynamics

We present a case study illustrating how PEAR changes agent answers across multiple debate rounds. The example is adapted from a MMLU-Pro accounting question. We use five agents and a sparse k-regular communication graph with $k = 2$, so every agent receives exactly two critiques in each debate round.

Problem. The Alfors Company had a beginning inventory of \$30,000 on January 1, 1974. During the year, purchases amounted to \$87,500 and net sales were \$102,000. Assuming that the gross profit rate is 40% of net sales, what is the ending inventory using the gross profit method of inventory evaluation?

Correct computation. We use the following compact notation: GP is gross profit, COGS is cost of goods sold, GA is goods available, and EI is ending inventory.

$$\begin{aligned} \mathrm{GP} &= 0.40 \times 102{,}000 = 40{,}800, \\ \mathrm{COGS} &= 102{,}000 - \mathrm{GP} = 61{,}200, \\ \mathrm{GA} &= 30{,}000 + 87{,}500 = 117{,}500, \\ \mathrm{EI} &= \mathrm{GA} - \mathrm{COGS} = 56{,}300. \end{aligned}$$

Thus, the correct option is `D.$56,300`.

Routing and influence notation. At the beginning of each debate round r, PEAR constructs a candidate sparse communication graph. Each target agent receives $k = 2$ incoming critiques. The routing score is

$$\begin{aligned} S(G_r) = \alpha_T T(G_r) - \alpha_I I(G_r) \\ - \alpha_L L(G_r). \end{aligned}$$

where $T(G_r)$ rewards high-confidence dissenting sources correcting lower-confidence targets, $I(G_r)$ penalizes over-exposure to already influential agents, and $L(G_r)$ penalizes low-confidence

sources. A graph is sampled from the candidate set using softmax over $S(G_r)$ rather than deterministic argmax.

Agent influence is updated after each round using

$$\mathrm{infl}_i^{(r)} = \beta\, \mathrm{infl}_i^{(r-1)} + (1-\beta)\, a_i^{(r)},$$

where $a_i^{(r)}$ denotes agent i's critique adoption rate in round r.

Round 0: Initial Reasoning

Agent 1: I compute gross profit as 40% of sales, so gross profit is 40,800 and cost of goods sold is 61,200. Goods available for sale are 117,500, so ending inventory is 56,300. I think this corresponds to option C. *Confidence: 3.*

Agent 2: I first compute $30{,}000 + 87{,}500 = 117{,}500$. Then I use a ratio adjustment because the gross profit method relates sales and cost. I get an adjusted estimate near 52,500, so I choose option E. *Confidence: 3.*

Agent 3: Gross profit is $0.40 \times 102{,}000 = 40{,}800$. Cost of goods sold is $102{,}000 - 40{,}800 = 61{,}200$. Goods available for sale are 117,500, so ending inventory is $117{,}500 - 61{,}200 = 56{,}300$. I choose option D. *Confidence: 4.*

Agent 4: I compute cost of goods sold as 60% of net sales, which is 61,200. Subtracting from goods available gives 56,300, but I am unsure which option matches it, so I choose option D. *Confidence: 3.*

Agent 5: I use beginning inventory plus purchases, then subtract an estimated cost figure. I get confused about whether the 40% is a markup or a margin, so I choose option C. *Confidence: 2.*

Round metrics:

Metric	Value
Correct agents	1/5
Group vote	C
Answer entropy	1.37
Mean confidence	3.00
Influence entropy	1.00
Routing	–
Critiques per agent	–

Round 1: Correct Signal Enters the Debate

Round-start state. Only Agent 3 is correct. Agent 5 has the lowest confidence, and Agents 1, 4, and 5 share the same wrong answer. PEAR routes critiques so that the correct dissenting source can reach wrong or lower-confidence agents while preserving $k = 2$ incoming critiques per target.

Selected routing graph:

Target	Received critiques from
A1	A3, A4
A2	A3, A5
A3	A1, A2
A4	A2, A3
A5	A1, A3

Routing metrics at start:

Metric	Value
Candidate graphs	100
Selected score	0.41
$T(G_r)$	0.34
$I(G_r)$	0.20
$L(G_r)$	0.09
Targeted-cross rate	0.30

Critique phase.

Agent 3 → Agent 1: I agree with your arithmetic, but I disagree with your option. You correctly compute goods available as 117,500 and cost of goods sold as 61,200, so the ending inventory is 56,300. The error is not numerical; it is that 56,300 corresponds to option D, not option C.

Agent 4 → Agent 1: I also get 56,300 from the same subtraction. My concern is the option mapping, because your final letter does not match your computed value. You should keep your computation and revise the selected option to the one listing 56,300.

Agent 3 → Agent 2: I think your adjusted formula is the problem. The statement says gross profit is 40% of net sales, so gross profit is $0.40 \times 102{,}000 = 40{,}800$. We then subtract gross profit from sales to get cost of goods sold, not divide or rescale inventory. That gives $102{,}000 - 40{,}800 = 61{,}200$.

Agent 5 → Agent 2: I was also unsure about markup versus margin, but your 52,500 estimate does not follow from the stated 40% of net sales. The phrase "of net sales" means the 40% is taken directly from 102,000, so an extra ratio adjustment is likely introducing the error.

Agent 1 → Agent 3: I do not find an arithmetic error in your solution. Your chain is complete: compute gross profit from sales, derive cost of goods sold, compute goods available, and subtract cost of goods sold. That supports 56,300.
Agent 2 → Agent 3: I initially used a different formula, but your interpretation of gross profit as a percentage of net sales is more direct. Because the problem gives a gross profit rate, the gross profit method should estimate cost of goods sold first, and your solution does that.
Agent 2 → Agent 4: You computed 56,300, but your selected answer is inconsistent with that value. Since your cost-of-goods-sold step already gives 61,200, the only remaining issue is matching 56,300 to the correct option.
Agent 3 → Agent 4: Your calculation is right. You correctly used 60% of sales as cost of goods sold because gross profit is 40% of sales. The correction is only to select option D for 56,300.
Agent 1 → Agent 5: The 40% is a gross profit margin on net sales, not a markup on cost. Therefore cost of goods sold is the remaining 60% of 102,000, which is 61,200. Once that is known, ending inventory is goods available minus cost of goods sold.
Agent 3 → Agent 5: Your confusion about margin versus markup changes the cost estimate. Here the problem explicitly uses gross profit as a percentage of net sales, so we compute 40,800 gross profit and 61,200 cost of goods sold. The ending inventory should be $117{,}500 - 61{,}200 = 56{,}300$, corresponding to option D.
Update phase.
Agent 1: I accept Agent 3's correction. I recheck my own steps: gross profit is 40,800, cost of goods sold is 61,200, and goods available are 117,500. My numeric result was already 56,300, so the mistake was selecting the wrong letter. I update my answer to option D. *Confidence: 4*.
Agent 2: I accept that my adjusted formula was unnecessary. The critiques make clear that I should not infer inventory by rescaling sales. Instead, I should compute gross profit first, then cost of goods sold. Cost of goods sold is 60% of net sales, or 61,200, and ending inventory is $117{,}500 - 61{,}200 = 56{,}300$. I update to option D. *Confidence: 4*.
Agent 3: I keep option D. The incoming critiques do not identify an error in my computation, and Agent 2's concern about ratio adjustment is resolved by the wording "40% of net sales." My answer remains 56,300. *Confidence: 4*.
Agent 4: I accept the option-mapping correction. I had the correct computation: $102{,}000 - 40{,}800 = 61{,}200$ for cost of goods sold, and $117{,}500 - 61{,}200 = 56{,}300$ for ending inventory. My selected letter was inconsistent with that value. I update from C to D. *Confidence: 4*.
Agent 5: I accept the correction. I was mixing up gross margin and markup, but the phrase "40% of net sales" fixes the interpretation. Gross profit is 40,800, cost of goods sold is 61,200, and ending inventory is 56,300. I update to option D. *Confidence: 4*.

Round-end metrics:

Metric	Value
Correct agents	5/5
Group vote	D
Answer entropy	0.00
Mean confidence	4.00
Changed agents	4
W2R / R2W	4 / 0

Influence update:

Agent	Influence after round 1
A1	0.18
A2	0.16
A3	0.34
A4	0.16
A5	0.16

Round 2: Consensus Verification

Round-start state. All agents now agree on option D. Since answer disagreement has disappeared, the targeted-cross term becomes inactive and routing mainly balances influence while keeping the critique budget fixed.

Selected routing graph:

Target	Received critiques from
A1	A2, A5
A2	A1, A3
A3	A1, A4
A4	A2, A5
A5	A3, A4

Routing metrics at start:

Metric	Value
Candidate graphs	100
Selected score	-0.08
$T(G_r)$	0.00
$I(G_r)$	0.11
$L(G_r)$	0.02
Targeted-cross rate	0.00

Critique phase.

Agent 2 → Agent 1: Your updated answer is consistent. I re-ran your corrected path: gross profit is 40,800, cost of goods sold is 61,200, and ending inventory is 56,300. Since your final option now matches that value, I do not see a remaining error.

Agent 5 → Agent 1: I agree with your current answer. Your earlier mistake was the letter choice rather than the accounting computation, and that has been corrected. The value 56,300 supports option D.

Agent 1 → Agent 2: Your revised formula is now correct. The important fix is that you no longer divide sales by 0.6 or use a separate ratio estimate. You compute cost of goods sold directly from the gross profit margin, then subtract it from goods available.

Agent 3 → Agent 2: I agree with your final answer D. Your cost-of-goods-sold step is now aligned with the gross profit method: $102{,}000 - 0.40 \times 102{,}000 = 61{,}200$. That leads to the same ending inventory value.

Agent 1 → Agent 3: I find no error in your calculation. You consistently treat the 40% as gross profit on sales, and your final subtraction uses the right base, 117,500 goods available. Your answer remains D.

Agent 4 → Agent 3: Your solution is internally consistent and matches the option value. I specifically checked the two potential failure points–margin interpretation and option mapping–and both support D.

Agent 2 → Agent 4: Your final answer is now consistent with your own computed value of 56,300. The earlier inconsistency was selecting C; with D, your reasoning and answer now agree.

Agent 5 → Agent 4: I agree that option D is the correct mapping for 56,300. I also checked that 56,300 is not obtained by a shortcut; it follows from the full gross-profit-method sequence.

Agent 3 → Agent 5: Your corrected reasoning now treats the 40% as gross profit on net sales, which is correct. That removes the markup-versus-margin ambiguity and fixes the cost of goods sold at 61,200.

Agent 4 → Agent 5: I agree with your updated answer. Your current reasoning starts from goods available of 117,500 and subtracts cost of goods sold of 61,200, so the ending inventory is 56,300.

Update phase.

Agent 1: I keep option D. Both critiques check the two places I previously could fail: the computation and the option mapping. Since 56,300 is still the ending inventory and it maps to D, I make no change. *Confidence: 4.*

Agent 2: I keep option D. I explicitly discard my earlier adjusted formula and use the gross profit method directly: sales minus gross profit gives cost of goods sold, and goods available minus cost of goods sold gives ending inventory. *Confidence: 4.*

Agent 3: I keep option D. No critique identified an error, and the other agents' revised computations now match my original result. This strengthens the consensus around 56,300. *Confidence: 4.*

Agent 4: I keep option D. My previous uncertainty was only about mapping the value to the option; after checking the option text against 56,300, I no longer see a reason to choose C. *Confidence: 4.*

Agent 5: I keep option D. I now correctly interpret the 40% as a gross profit margin on net sales, so I do not use markup logic. The recomputed ending inventory remains 56,300. *Confidence: 4.*

Round-end metrics:

Metric	Value
Correct agents	5/5
Group vote	D
Answer entropy	0.00
Mean confidence	4.00
Changed agents	0
W2R / R2W	0 / 0

Influence update:

Agent	Influence after round 2
A1	0.20
A2	0.17
A3	0.30
A4	0.17
A5	0.16

Round 3: Stable Agreement

Round-start state. The group remains unanimous on D. PEAR continues sparse routing, but the graph no longer needs to propagate a minority correct answer; it only tests whether the consensus is stable.

Selected routing graph:

Target	Received critiques from
A1	A3, A4
A2	A4, A5
A3	A1, A5
A4	A1, A2
A5	A2, A3

Routing metrics at start:

Metric	Value
Candidate graphs	100
Selected score	-0.07
$T(G_r)$	0.00
$I(G_r)$	0.10
$L(G_r)$	0.02
Targeted-cross rate	0.00

Critique phase.

Agent 3 → Agent 1: I agree with your final answer. The cost of goods sold is 61,200, and you subtract it from the correct goods-available total rather than from sales. That keeps the inventory calculation valid.

Agent 4 → Agent 1: I agree. Goods available for sale minus cost of goods sold gives 56,300, and that is exactly what the gross profit method asks us to estimate. Your final answer should stay D.

Agent 4 → Agent 2: Your revised solution is correct. You fixed the earlier ratio issue by computing gross profit first, so the final answer should remain D.

Agent 5 → Agent 2: I agree with option D; no new issue is present. The formula now uses all three given quantities–beginning inventory, purchases, and net sales–in the right places.

Agent 1 → Agent 3: Your original answer remains correct. You did not skip the intermediate cost-of-goods-sold step, which is why your final inventory value is reliable.

Agent 5 → Agent 3: I agree with your calculation and final option. Your reasoning also resolves my earlier ambiguity about margin versus markup, because it uses gross profit as a direct percentage of net sales.

Agent 1 → Agent 4: Your answer is now consistent with your calculation. The same numerical result that previously led you to C now correctly leads to D.

Agent 2 → Agent 4: I agree that option D matches the computed ending inventory. I also checked that the 56,300 value uses cost of goods sold, not gross profit, in the final subtraction.

Agent 2 → Agent 5: Your corrected margin interpretation is right. Because gross profit is 40% of sales, cost of goods sold is 60% of sales; this is the exact correction that changed your answer.

Agent 3 → Agent 5: I agree with your final answer D. Your current computation now follows the same complete sequence as the correct solution, so I would not revise it further.

Update phase.

Agent 1: I keep D. The critiques independently verify the same computation and remind me that the subtraction must be from goods available, not from sales. *Confidence: 4.*

Agent 2: I keep D. There is no remaining formula issue: I now compute cost of goods sold from the gross profit margin and use it to estimate ending inventory. *Confidence: 4.*

Agent 3: I keep D. The group agrees with the original cost-of-goods-sold method, and no critique introduces a competing calculation that changes the value. *Confidence: 4.*

Agent 4: I keep D. The option mapping is settled, and the accounting steps still produce 56,300. *Confidence: 4.*

Agent 5: I keep D. The gross profit method is now clear: gross profit determines cost of goods sold, and cost of goods sold determines ending inventory. *Confidence: 4.*

Round-end metrics:

Metric	Value
Correct agents	5/5
Group vote	D
Answer entropy	0.00
Mean confidence	4.00
Changed agents	0
W2R / R2W	0 / 0

Why this case demonstrates the effect of PEAR. This trajectory shows how PEAR can amplify a minority correct answer without requiring dense all-to-all debate. Initially, only one of five agents selects the correct option, while the group vote is wrong. The first routing step exposes wrong agents to the correct dissenting reasoning and to

critiques about option mapping and the gross-profit formula. After one debate round, all agents switch to the correct answer. Later rounds keep the same sparse critique budget but no longer destabilize the consensus: answer entropy remains zero and no right-to-wrong transition occurs.

Diagnostic summary. Across the full debate, the example has $\mathrm{W2R} = 4$, $\mathrm{R2W} = 0$, final answer entropy 0.00, and a balanced final influence distribution. The key correction happens in Round 1, where the initially minority correct reasoning becomes the group consensus through targeted sparse critique routing.

F LLM Usage

We used large language model assistants for paper polishing (improving prose clarity and grammar) and code refactoring during the development of this work. All scientific claims, experimental results, and analyses were produced and verified by the authors; no AI-generated text was included without author review.