MCN Phase 2: Heterogeneous Tribes — CTS Router Report

Phase 2 setup: T0=temp 0.1 (deterministic), T1=temp 0.5 (balanced), T2=temp 0.9 (creative). Router: CategoryThompsonSampling. Tasks: 1502 (8×250 stratified). Baseline: Phase 1C LinUCB + homogeneous tribes (2000 tasks).
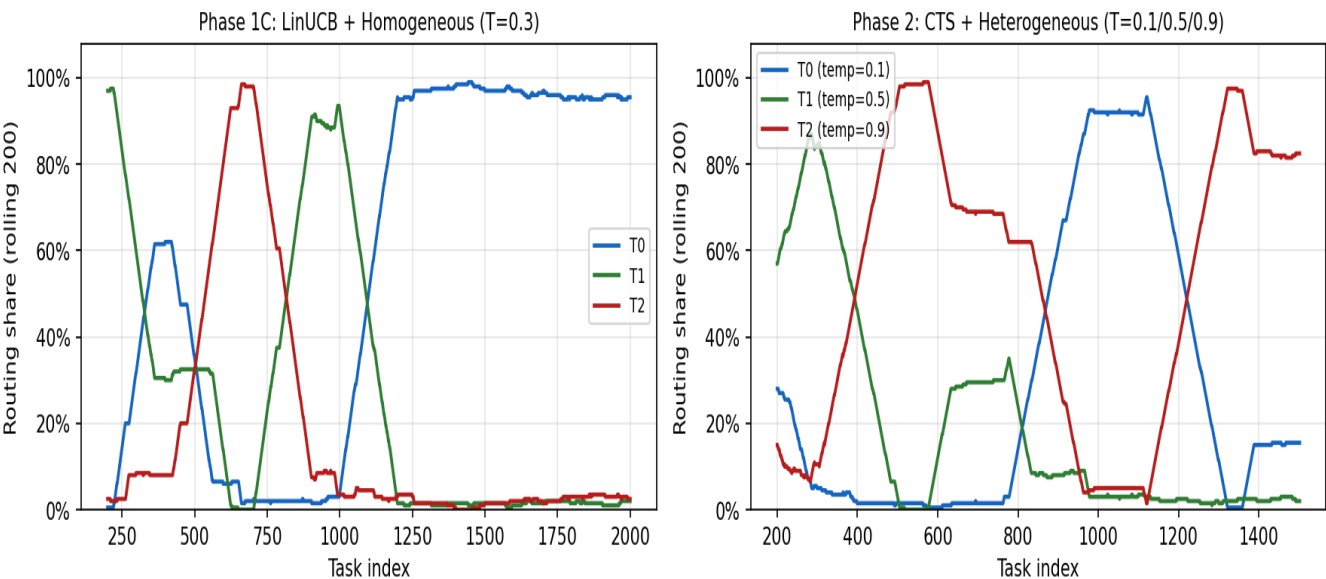
1. Overall Results

Experiment Comparison

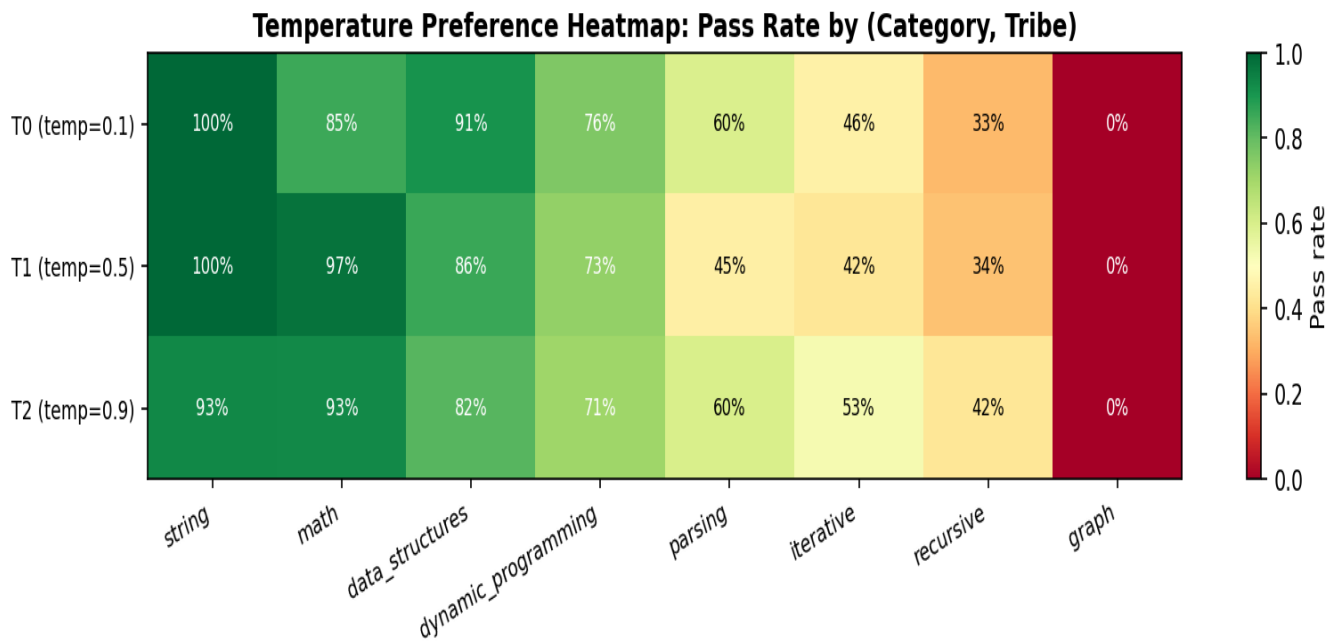| Metric	Phase 1C (LinUCB+Homo)	Phase 2 (CTS+Hetero)	Delta
Overall pass rate	60.7%	60.7%	-0.0 pp
Oracle (per-cat best)	65.8%	64.3%	-1.5 pp
Oracle gap	5.1%	3.7%	-1.4 pp

2. Router Convergence

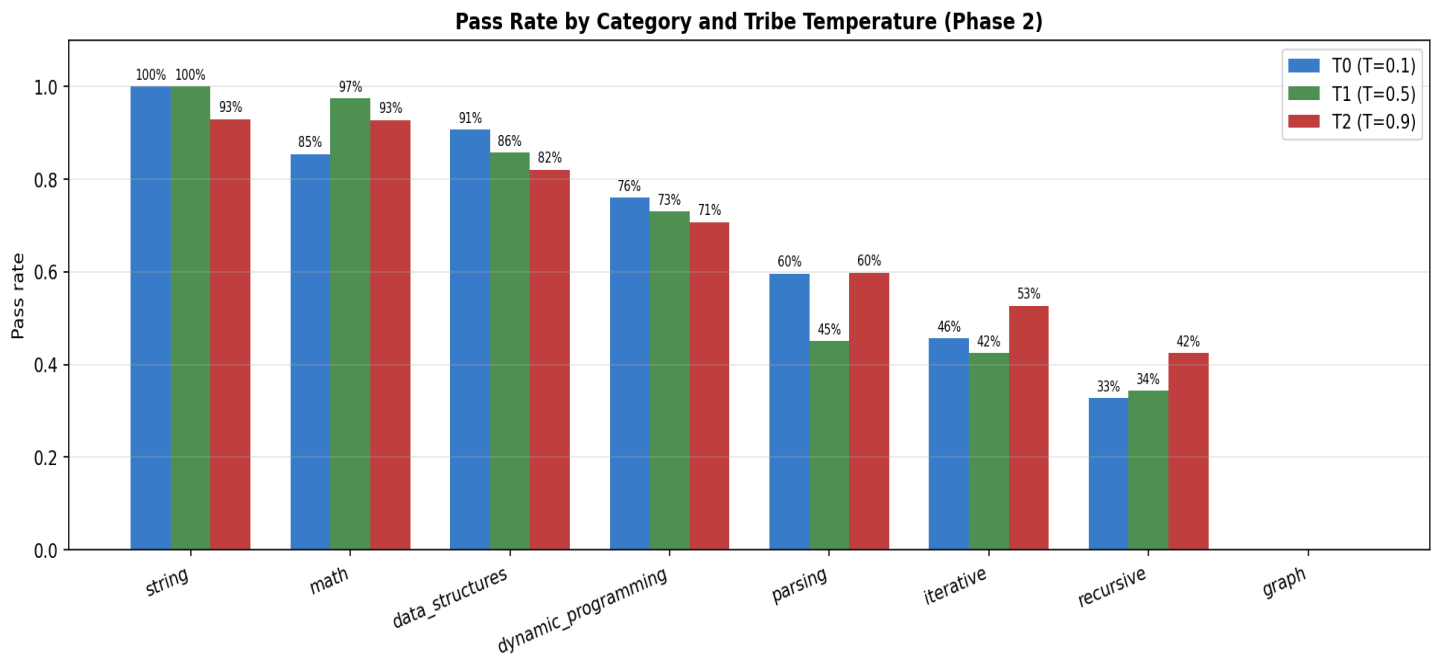Router Convergence: Homogeneous vs Heterogeneous Tribes

Phase 1C LinUCB converges to T0 at ~task 1700 (55% share). Phase 2 CTS maintains per-category Beta posteriors — expect distributed routing if tribe temperatures create genuine category-level performance differences.
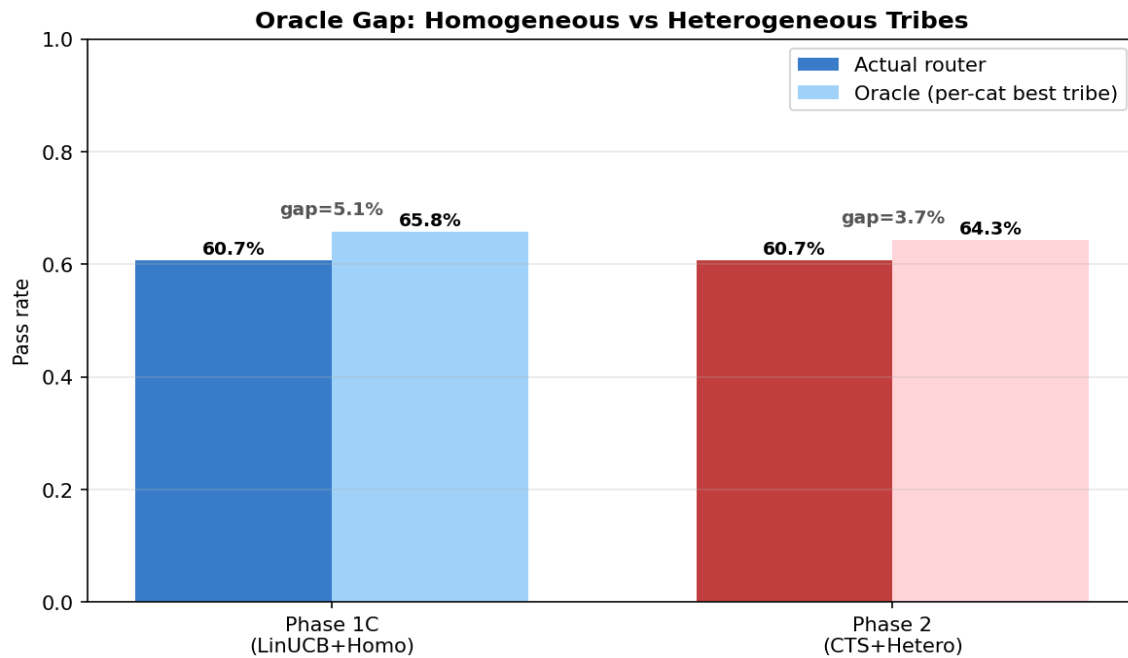
3. Temperature Preference by Category

Temperature Preference Heatmap: Pass Rate by (Category, Tribe)

Green cells indicate high pass rate for that (category, temperature) combination. If routing is learning correctly, CTS should route high-confidence categories to the temperature with the highest per-category pass rate.

Pass Rate by Category and Tribe Temperature (Phase 2)

4. Oracle Gap: Homogeneous vs Heterogeneous

Oracle Gap: Homogeneous vs Heterogeneous Tribes

A larger oracle gap in Phase 2 indicates that tribe diversity creates a real routing opportunity. A smaller gap means CTS is successfully exploiting it.

---

5. Per-Category Results (Phase 2)

Per-Category Pass Rate vs Oracle

Category	N	CTS Rate	Best Tribe	Oracle Rate	Gap
string	185	96%	T0 (T=0.1)	100%	3.8%
math	184	92%	T1 (T=0.5)	97%	5.6%
data_structures	178	85%	T0 (T=0.1)	91%	5.3%
dynamic_programming	186	73%	T0 (T=0.1)	76%	3.4%
parsing	199	57%	T2 (T=0.9)	60%	3.0%
iterative	190	48%	T2 (T=0.9)	53%	4.3%
recursive	189	38%	T2 (T=0.9)	42%	4.3%
graph	191	0%	T0 (T=0.1)	0%	0.0%

6. Per-Tribe Performance

Per-Tribe Routing Share and Pass Rate

Tribe	Temperature	Tasks	Share	Pass Rate
T0	temp=0.1	422	28%	60.4%
T1	temp=0.5	295	20%	59.3%
T2	temp=0.9	785	52%	61.3%

7. Interpretation

Hypothesis: Heterogeneous temperatures create a non-flat reward surface, allowing CTS to learn meaningful routing preferences per category.

Key question: Is the oracle gap larger in Phase 2 (more room to route) AND is the actual CTS pass rate higher (gap is being exploited)?

If both are true: tribe diversity is a prerequisite for effective routing. If oracle gap is larger but CTS rate unchanged: 2000 tasks is insufficient for CTS to learn. If neither: even temperature diversity does not create exploitable differences.