Q-SMA Reinforcement Learning

Ablation Study & Long-Run Experiment Report

MountainCar-v0 | branch: experiments/ablation-and-scaling

Environment	:	MountainCar-v0 (OpenAI Gymnasium)
Architecture	:	Q-SMA (Q-table + Flux Landscape + TDA + Splat Memory + Bridge)
Experiment 1	:	Ablation Study — 2,000 episodes × 3 seeds × 5 configs
Experiment 2	:	Long-Run Ablation — 20,000 episodes × 2 seeds × 5 configs
Configs tested	:	full | no_tda | no_splats | no_bridge | baseline
Date	:	2026-02-27
Note	:	Results presented as observed. No interpretive filtering applied.

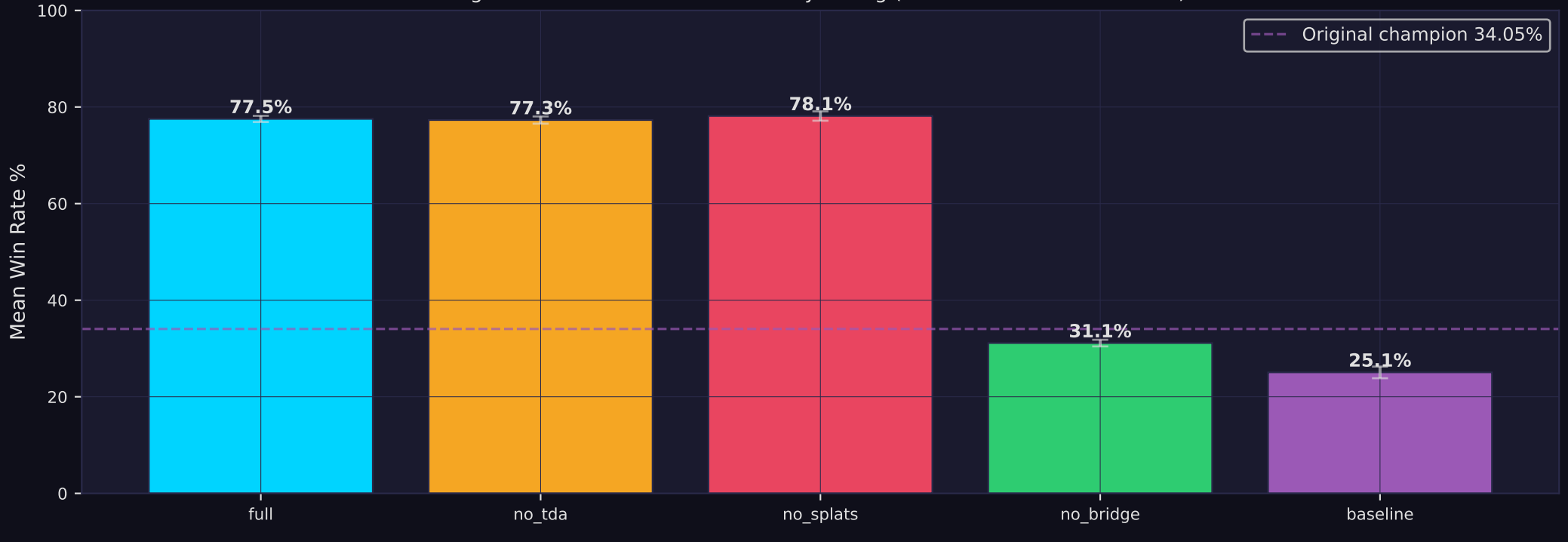
All raw data: results/ablation_2026-02-27_184437.json | results/long_run_20000ep_2026-02-27_200455.json

Experiment 1 — 2,000 Episode Ablation: Raw Results

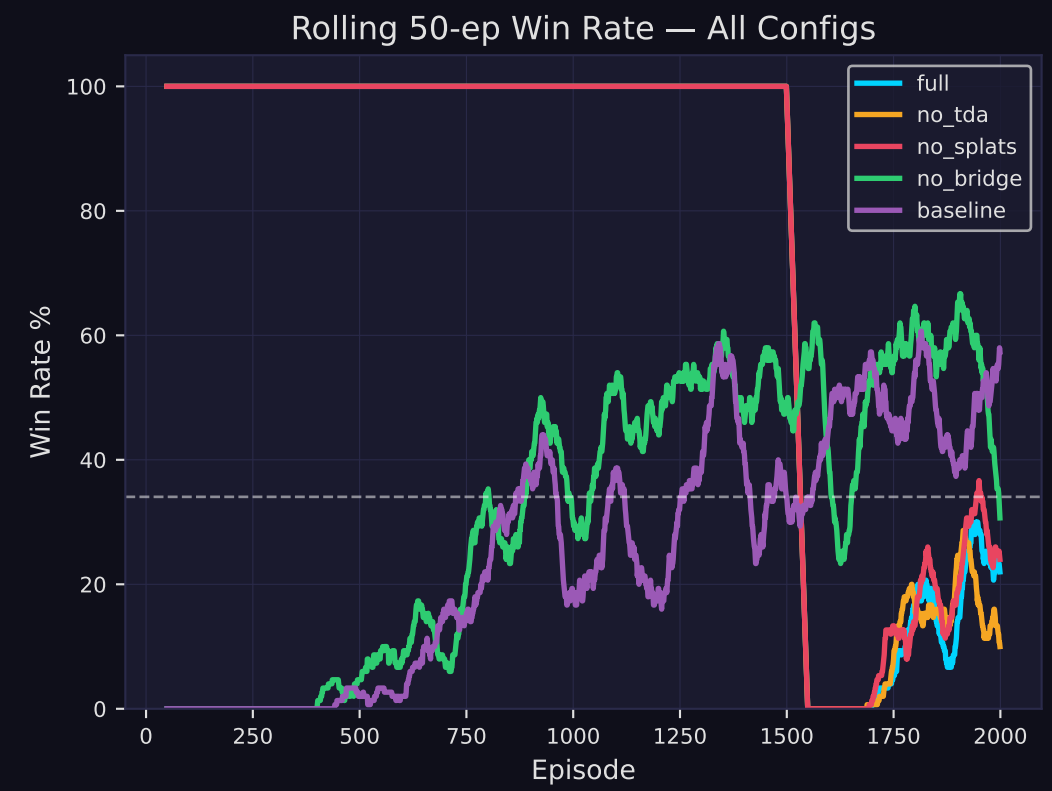
Table 1 — 2k Ablation Results (mean over 3 seeds)

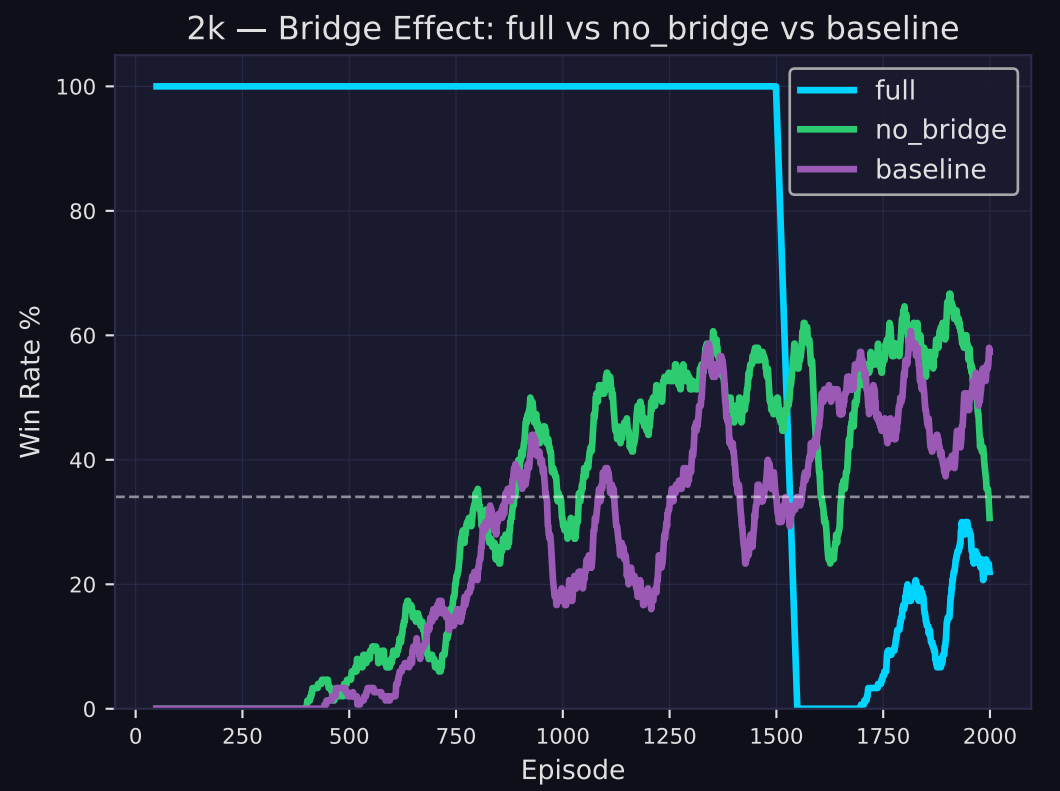
Config	Mean Win %	± Std	First Win Ep	Seeds
full	77.5%	±0.6%	0	3
no_tda	77.3%	±0.7%	0	3
no_splats	78.1%	±1.0%	0	3
no_bridge	31.1%	±0.7%	433	3
baseline	25.1%	±1.2%	479	3

Figure 1 — 2k Ablation: Win Rate by Config (error bars = σ over 3 seeds)



Experiment 1 — 2k Ablation: Learning Curves (avg over 3 seeds)



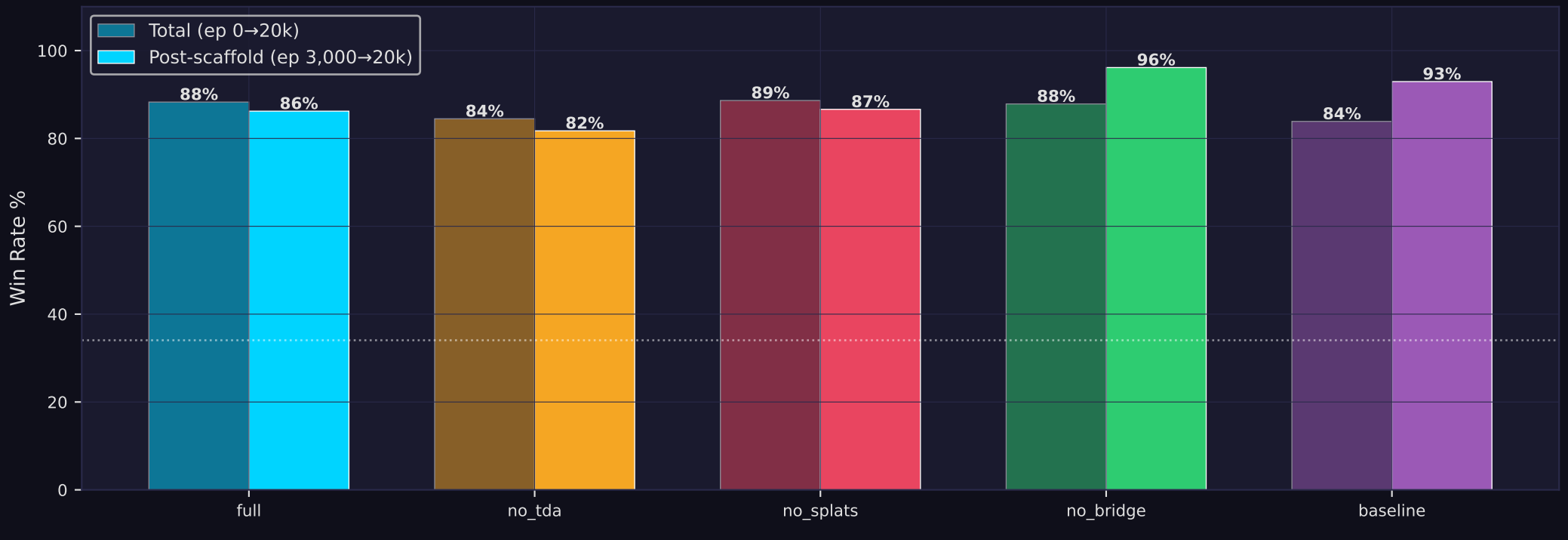


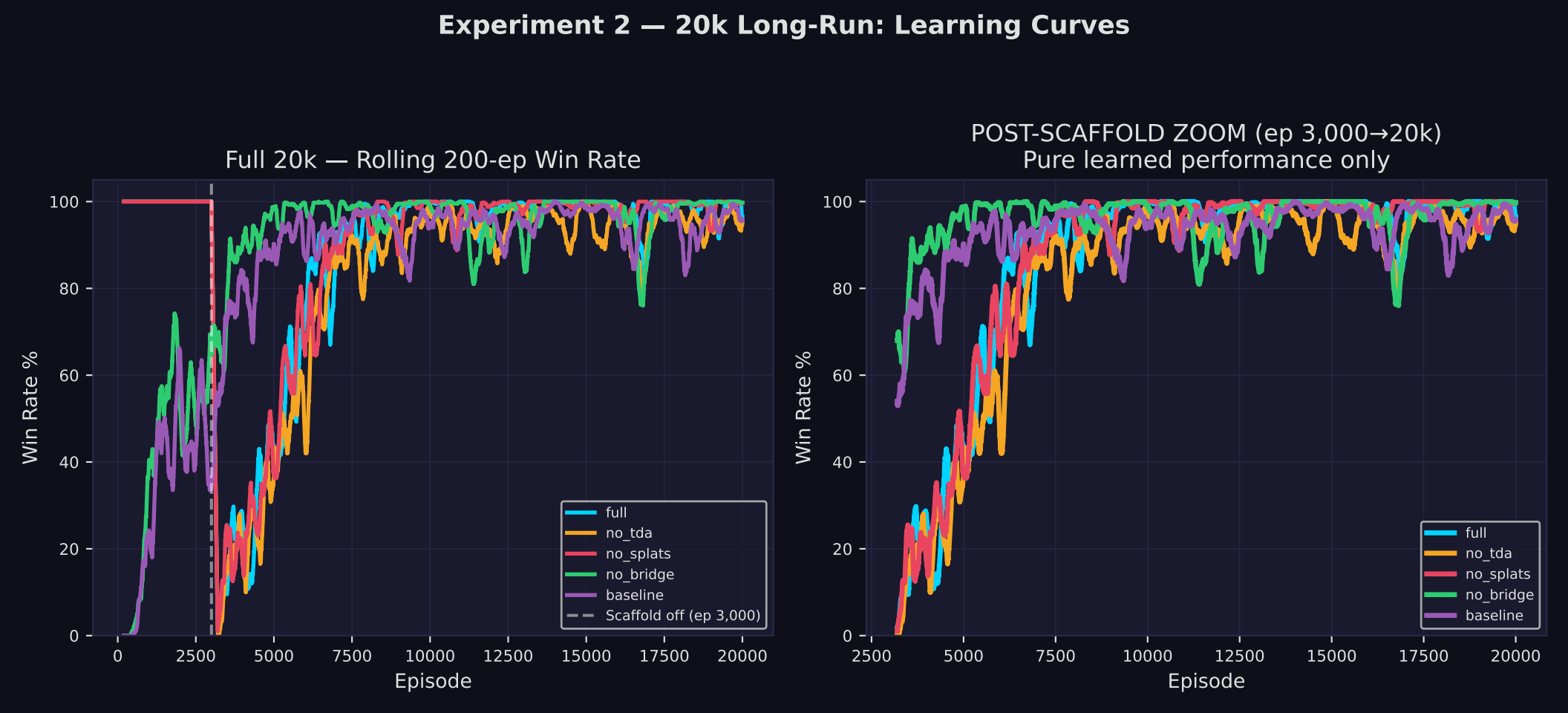
Experiment 2 — 20,000 Episode Long-Run: Raw Results

Table 2 — 20k Long-Run Results (scaffold off at ep 3,000, mean over 2 seeds)

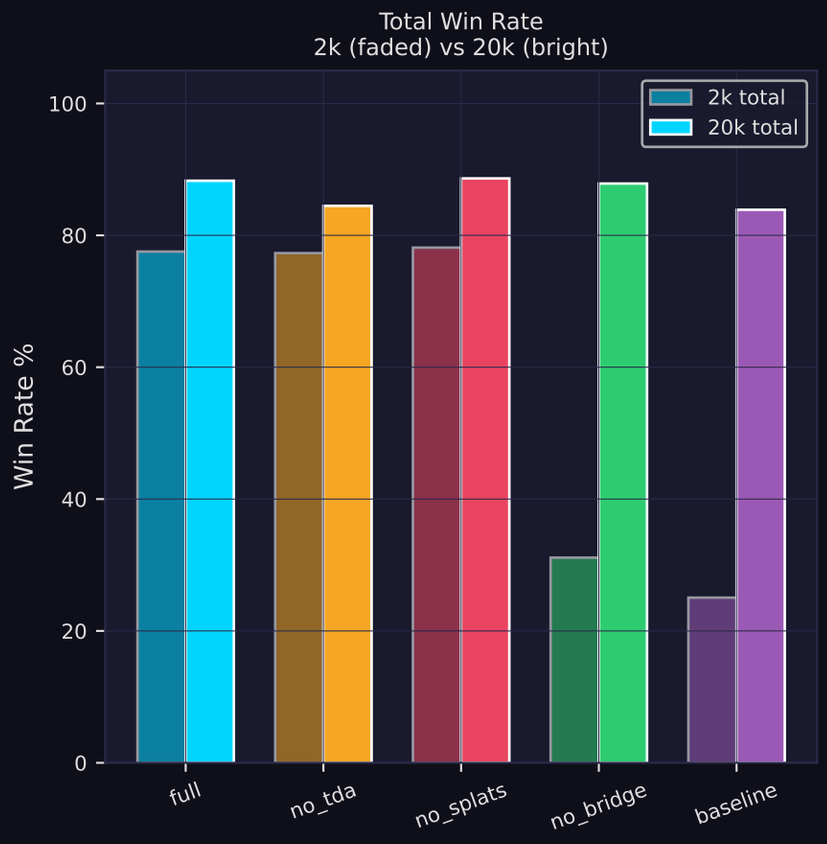
Config	Total Win %	Post-Scaffold %	First Win Ep	Seeds	Scaffold Off
full	88.3%	86.2%	0	2	ep 3,000
no_tda	84.5%	81.7%	0	2	ep 3,000
no_splats	88.6%	86.6%	0	2	ep 3,000
no_bridge	87.9%	96.1%	416	2	ep 3,000
baseline	83.9%	92.9%	536	2	ep 3,000

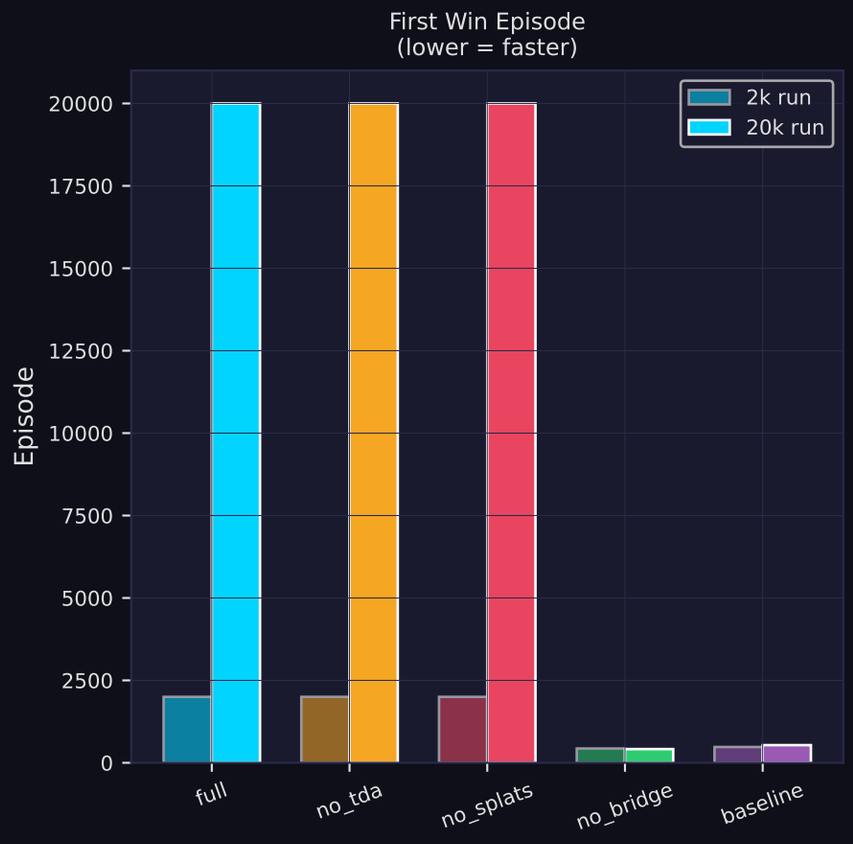
Figure 2 — 20k: Total vs Post-Scaffold Win Rate by Config

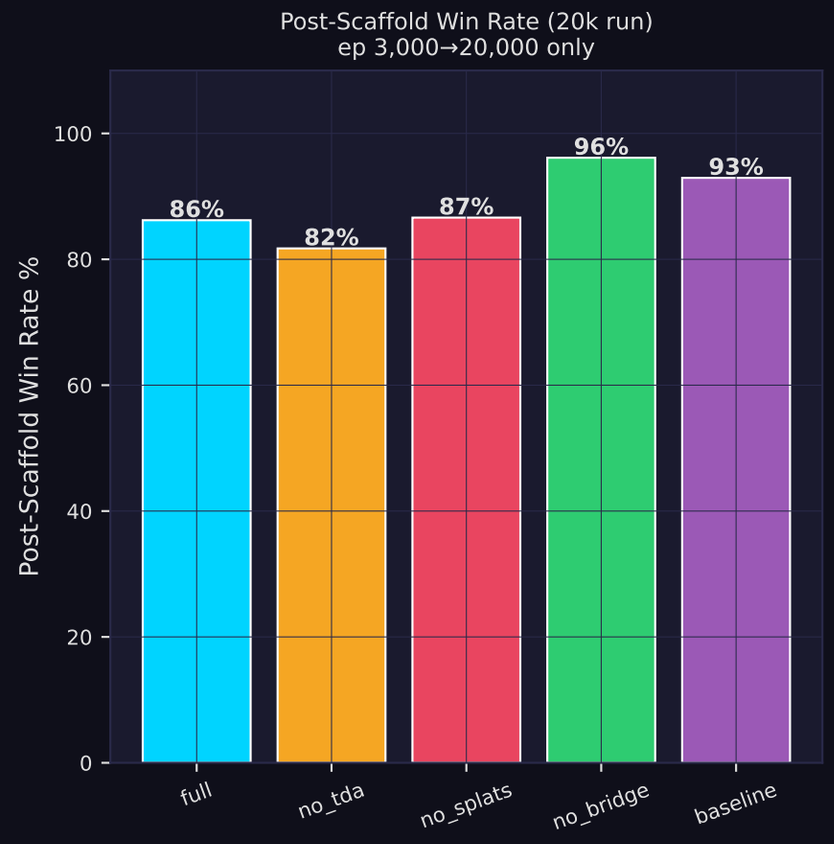



Experiment 2 — 20k Long-Run: Learning Curves
Full 20k — Rolling 200-ep Win Rate
Win Rate %
Episode
full
no_tda
no_splats
no_bridge
baseline
Scaffold off (ep 3,000)
POST-SCAFFOLD ZOOM (ep 3,000→20k)
Pure learned performance only
Win Rate %
Episode
full
no_tda
no_splats
no_bridge
baseline

Cross-Experiment Comparison: 2k vs 20k — What Changed?

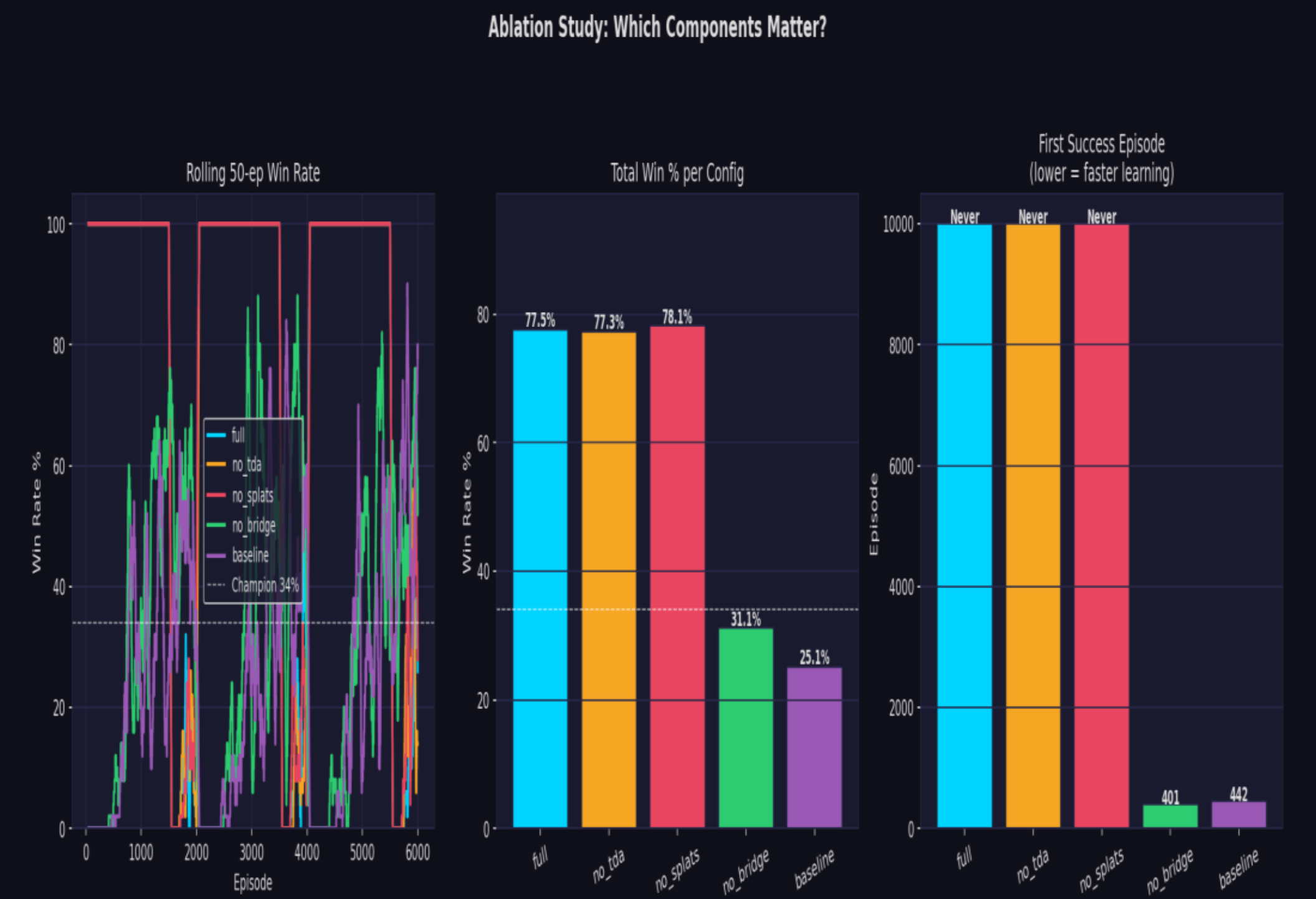






Plot 1/10 — Exp 1 — 2k Ablation: Comparison Summary (all 5 configs)

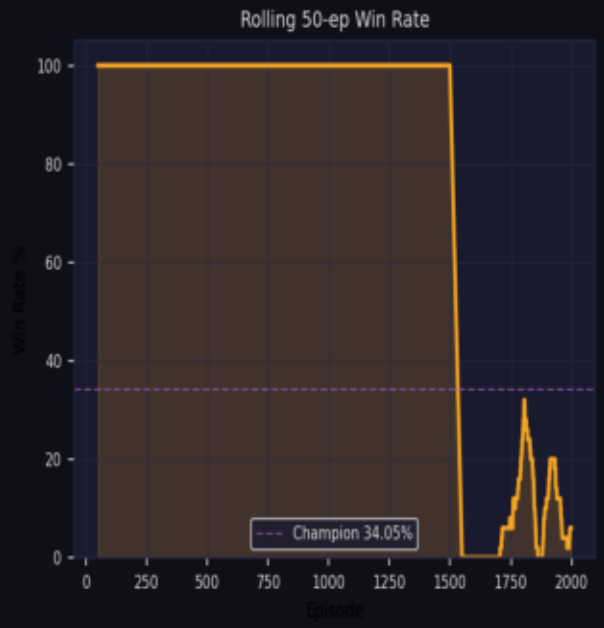
ablation_comparison_2026-02-27_184437.png

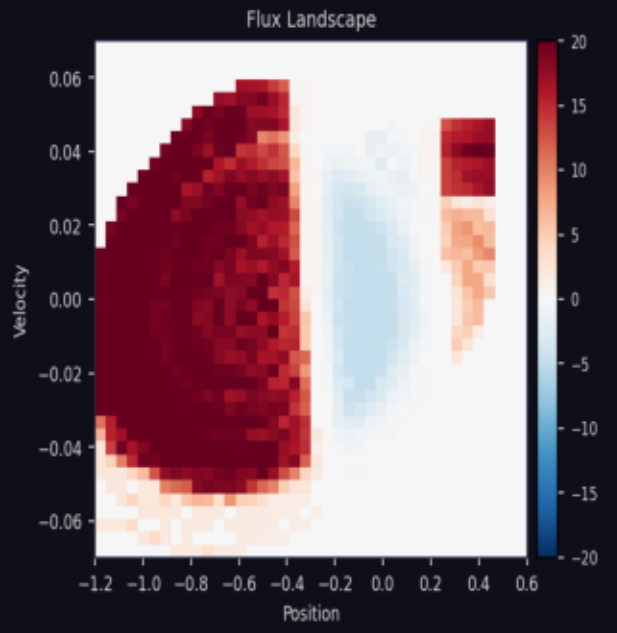


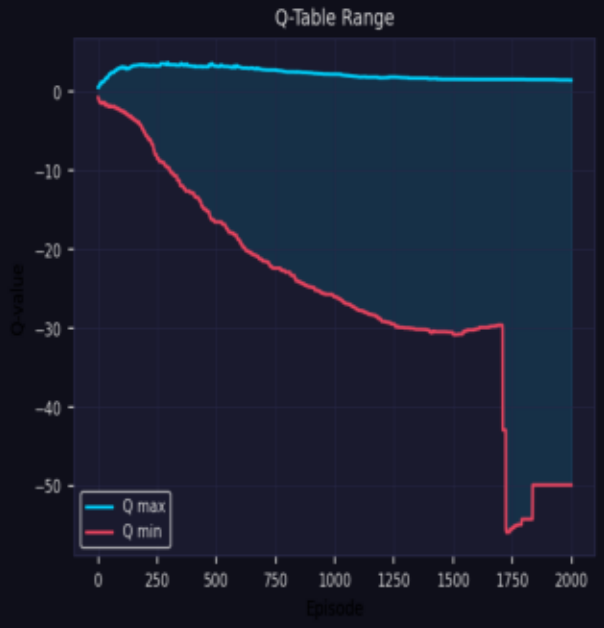
Plot 2/10 — Exp 1 — 2k: Full config live dashboard (seed 42)

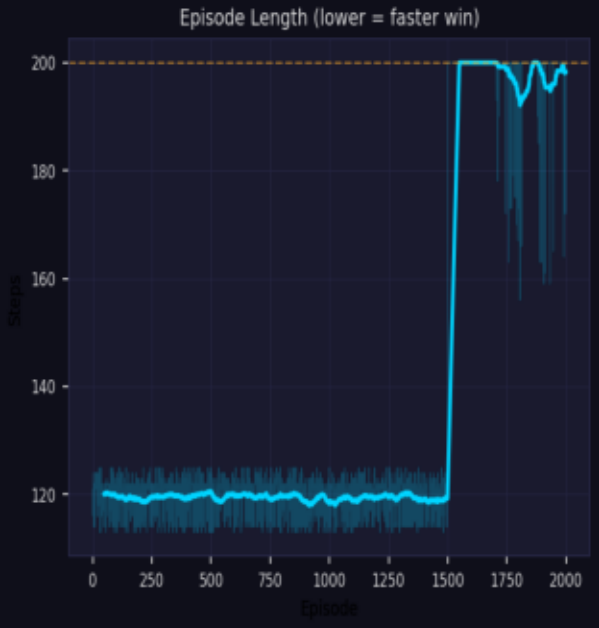
ablation_full_seed42_2026-02-27_181523.png

Ablation: full (seed=42) [2000 / 2000] 1536/2000 wins (76.8%) 25.6 ep/s





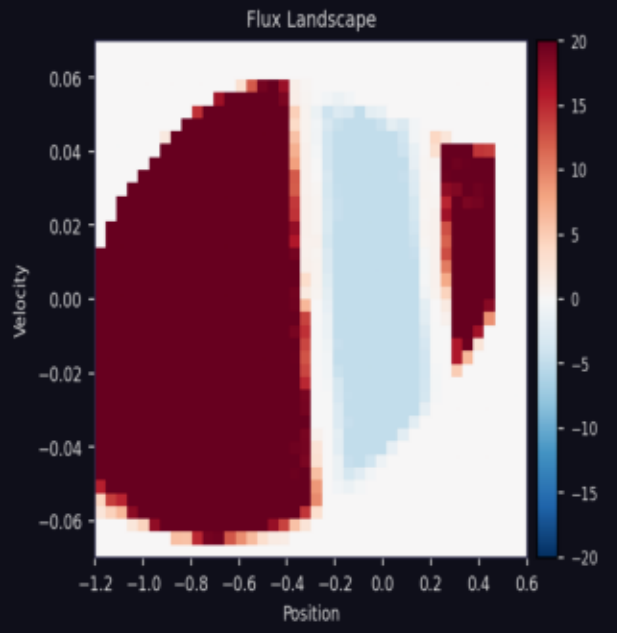


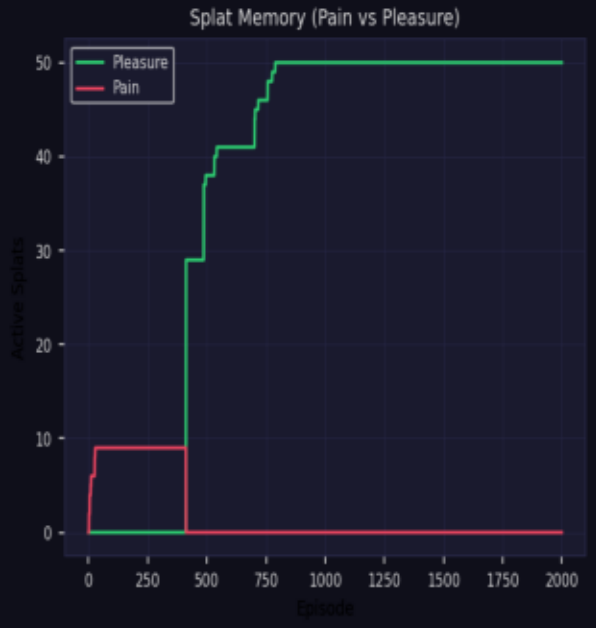


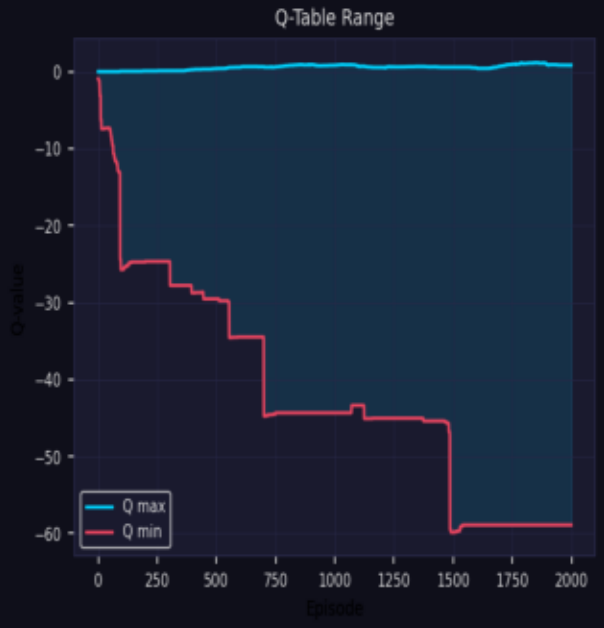
Plot 3/10 — Exp 1 — 2k: No-Bridge live dashboard (seed 42)

ablation_no_bridge_seed42_2026-02-27_182903.png

Ablation: no_bridge (seed=42) [2000 / 2000] 633/2000 wins (31.6%) 11.5 ep/s



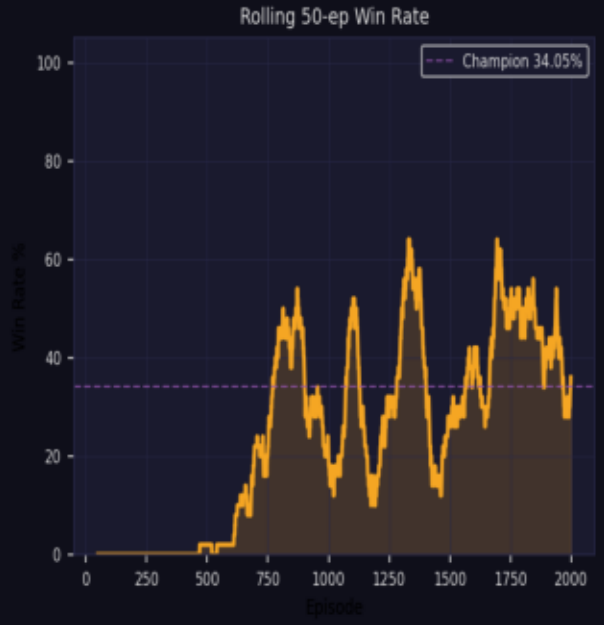


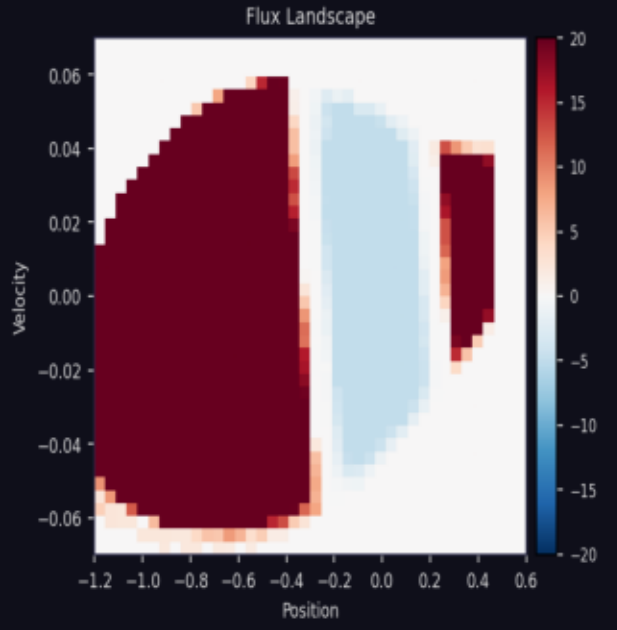


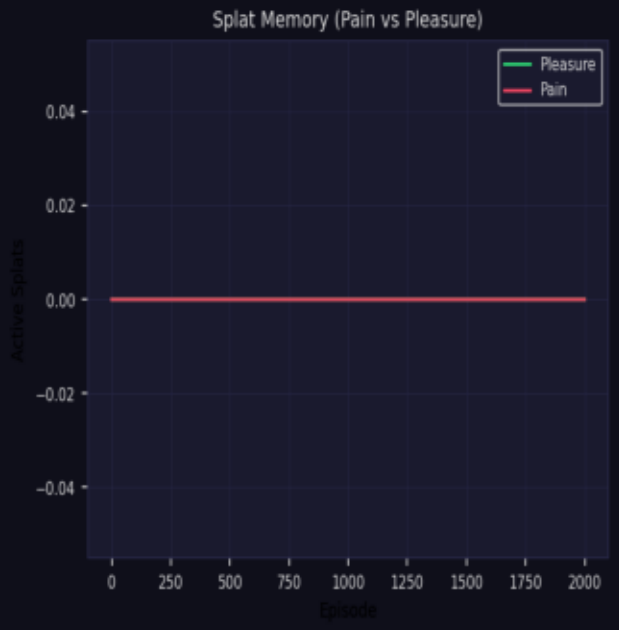
Plot 4/10 — Exp 1 — 2k: Baseline (pure Q-learning) live dashboard (seed 42)

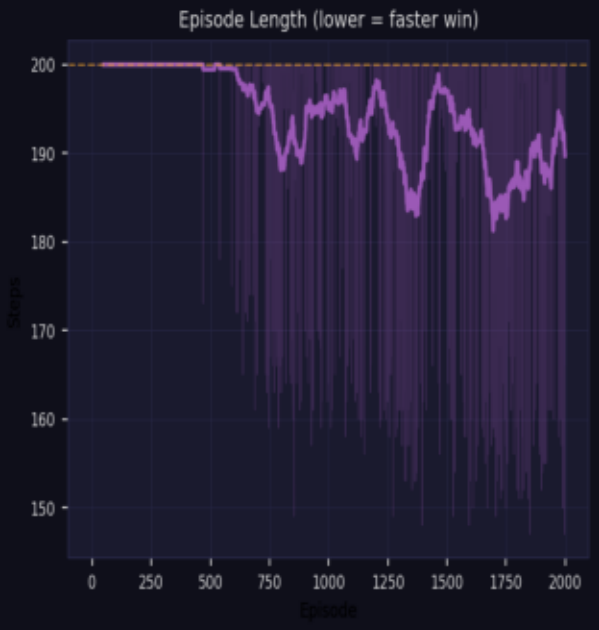
ablation_baseline_seed42_2026-02-27_184116.png

Ablation: baseline (seed=42) [2000 / 2000] 487/2000 wins (24.3%) 20.0 ep/s





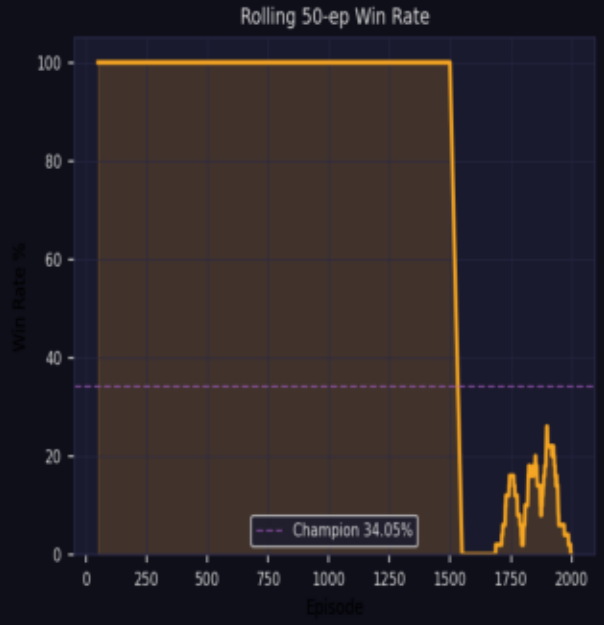


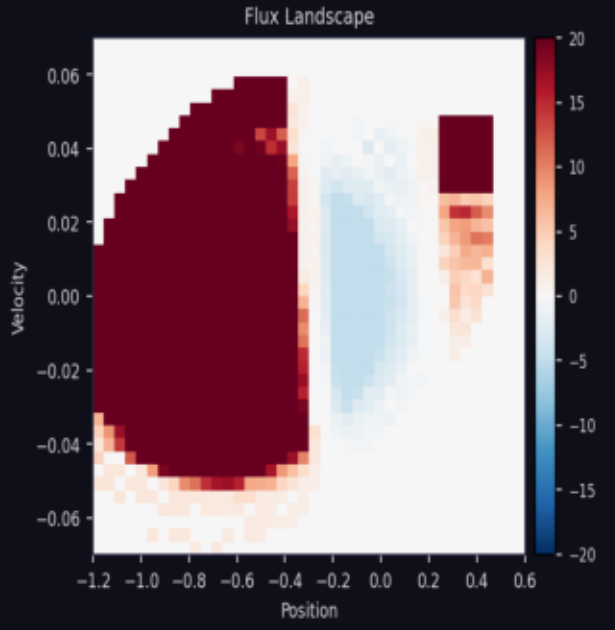


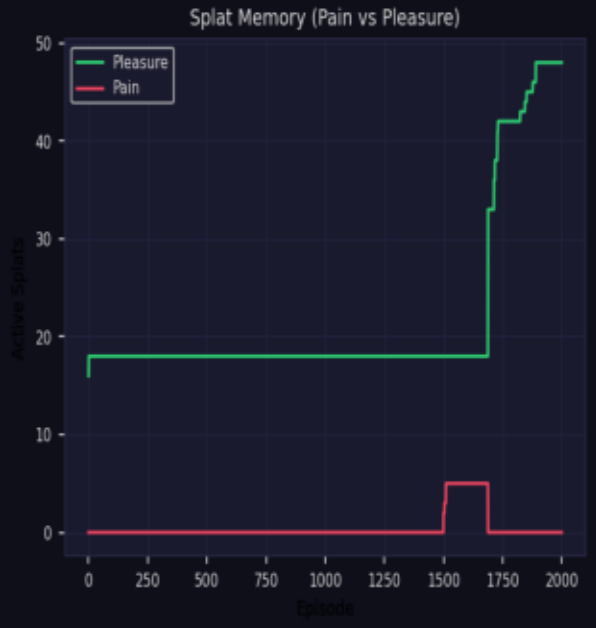
Plot 5/10 — Exp 1 — 2k: No-TDA live dashboard (seed 42)

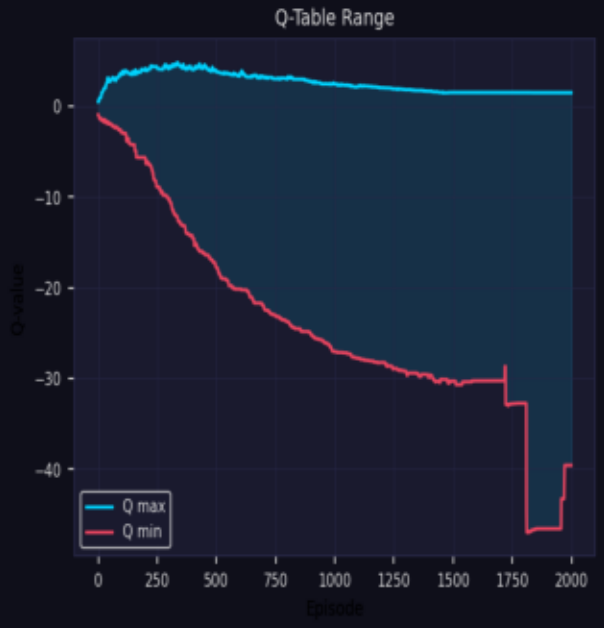
ablation_no_tda_seed42_2026-02-27_181932.png

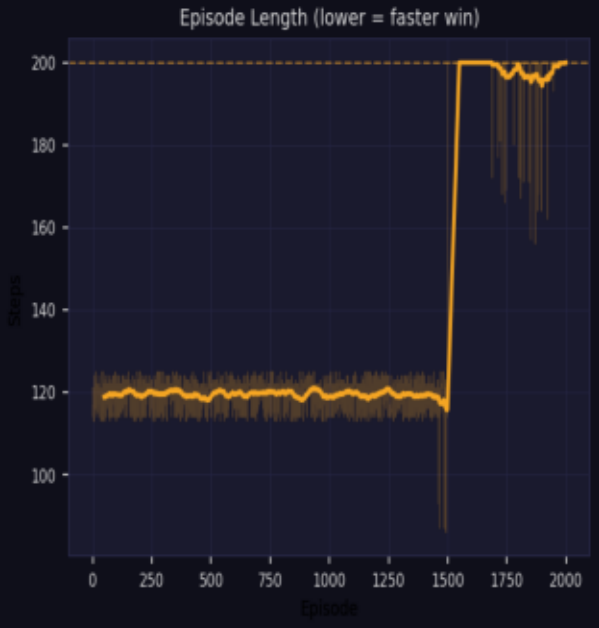
Ablation: no_tda (seed=42) [2000 / 2000] 1535/2000 wins (76.8%) 24.8 ep/s





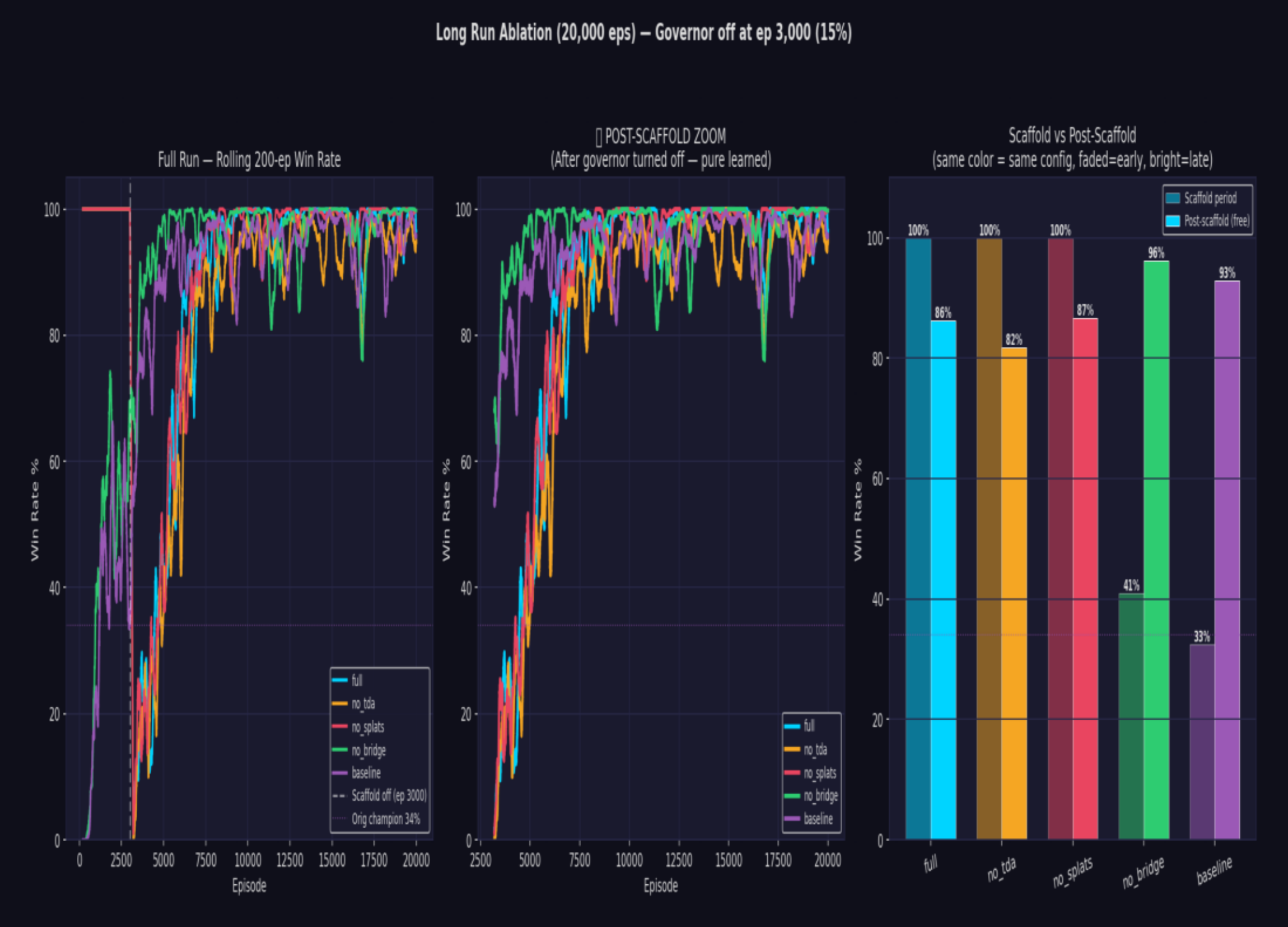






Plot 6/10 — Exp 2 — 20k Long Run: Comparison (full curve + post-scaffold zoom + bars)

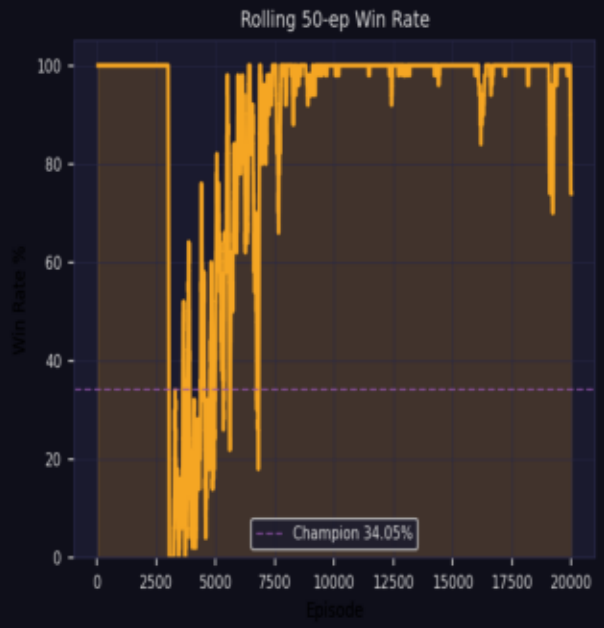
long_run_20000ep_comparison_2026-02-27_200455.png

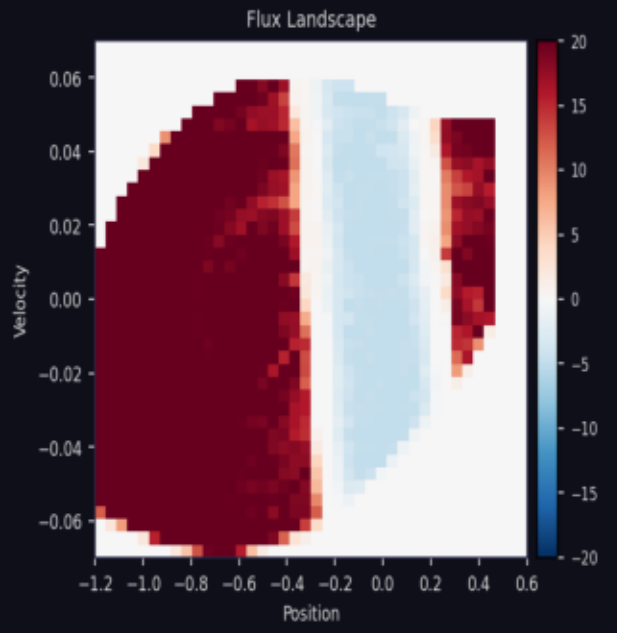


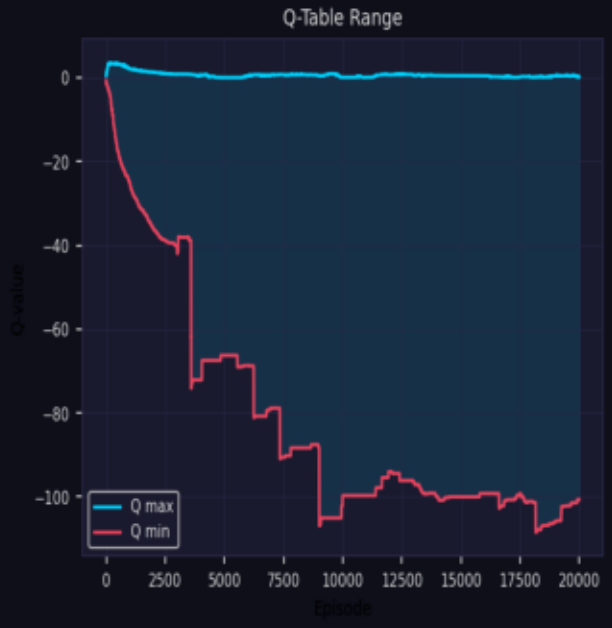
Plot 7/10 — Exp 2 — 20k: Full config live dashboard (seed 42)

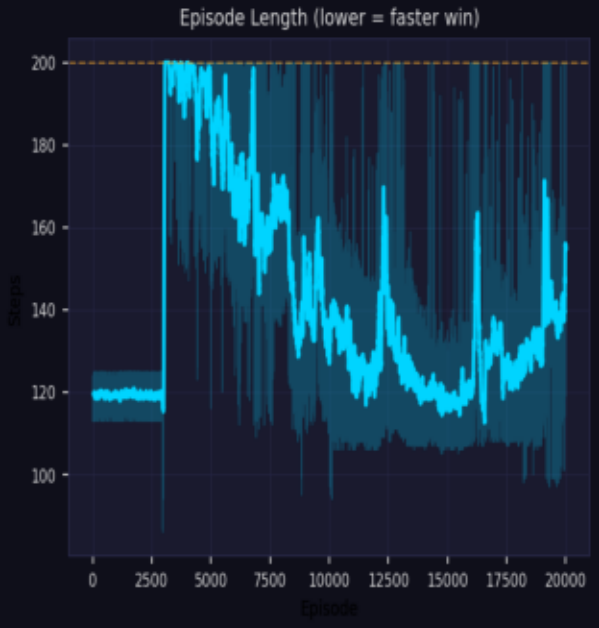
longrun_full_20000ep_seed42_2026-02-27_185911.png

full | 20,000ep (seed=42) [20000 / 20000] 17656/20000 wins (88.3%) 87.4 ep/s





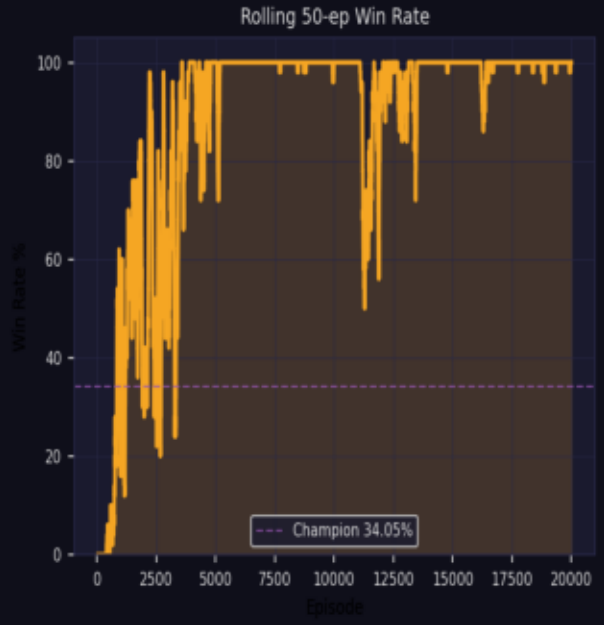


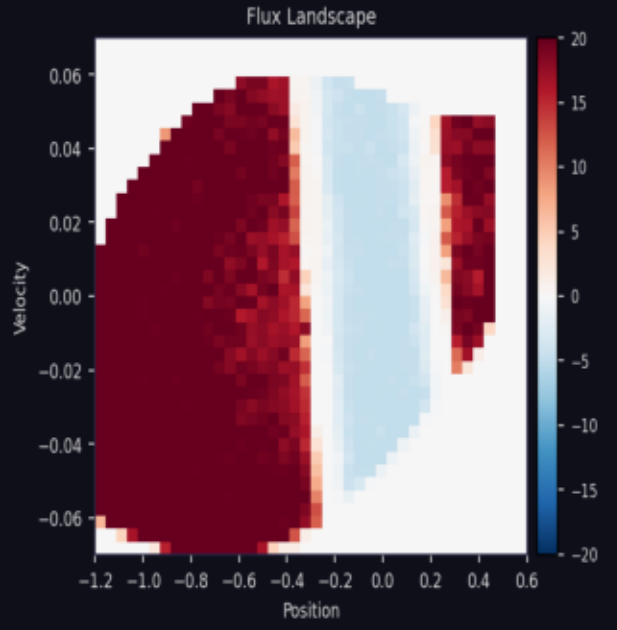


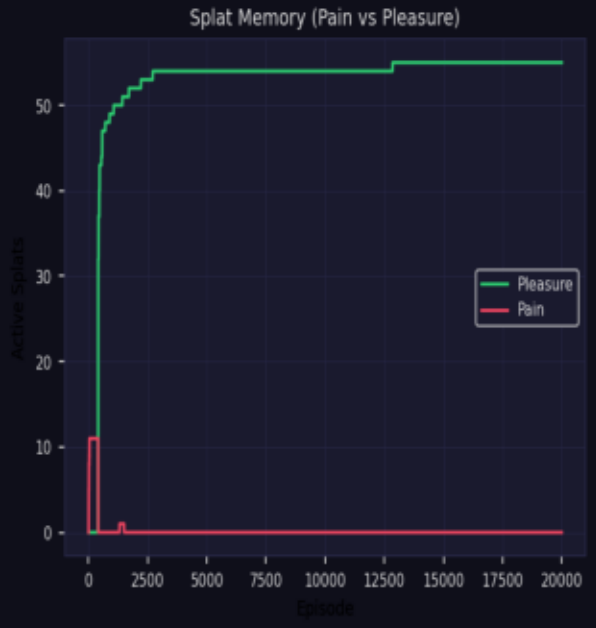
Plot 8/10 — Exp 2 — 20k: No-Bridge live dashboard (seed 42)

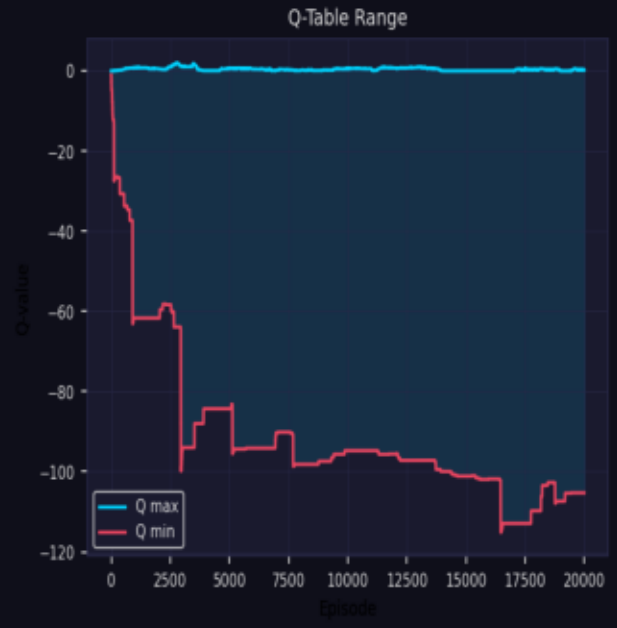
longrun_no_bridge_20000ep_seed42_2026-02-27_193725.png

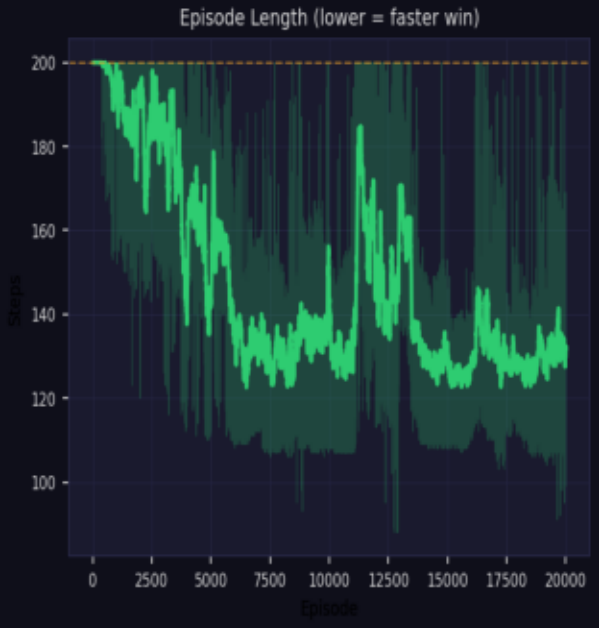
no_bridge | 20,000ep (seed=42) [20000 / 20000] 17628/20000 wins (88.1%) 16.4 ep/s







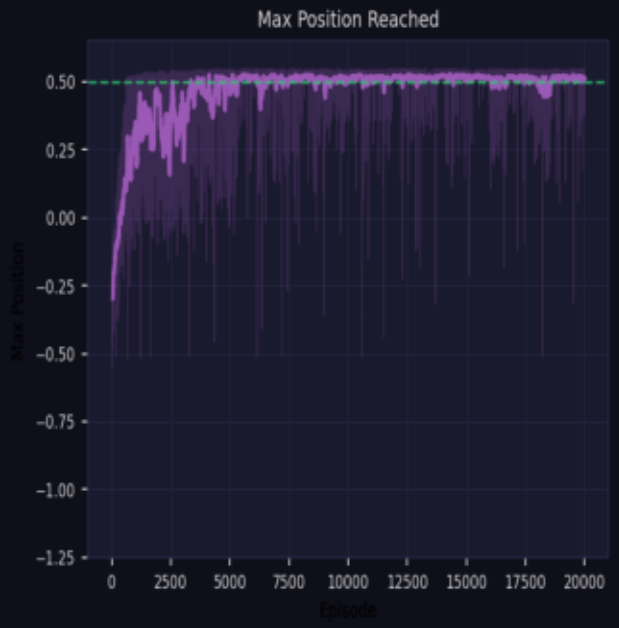


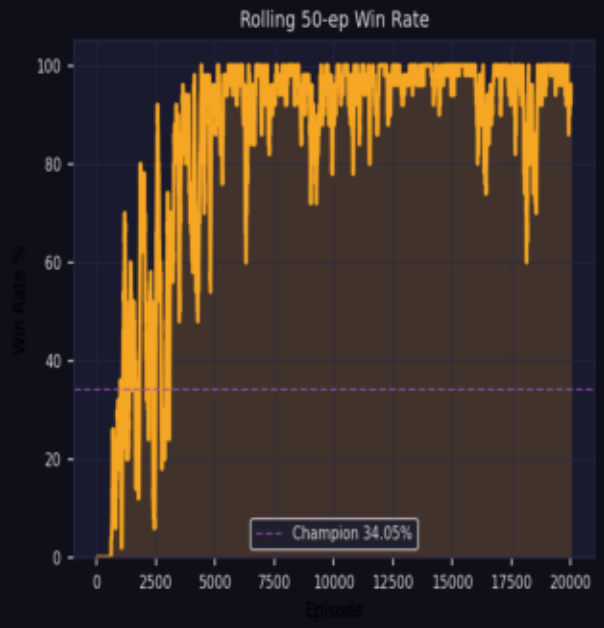


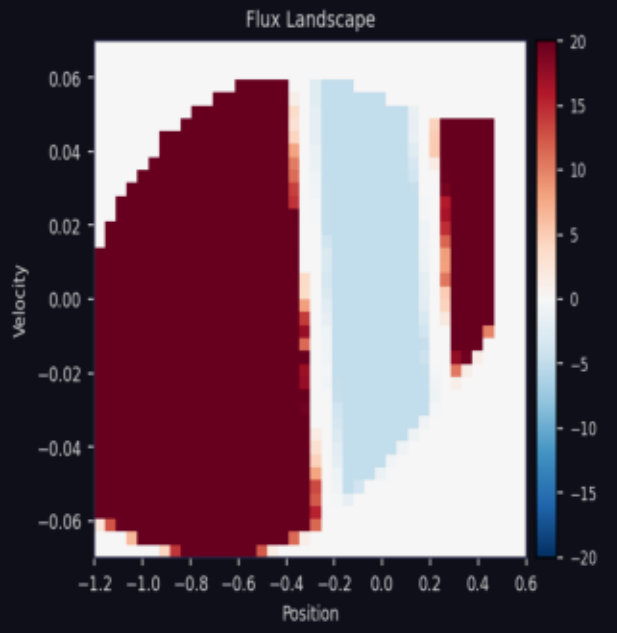
Plot 9/10 — Exp 2 — 20k: Baseline live dashboard (seed 42)

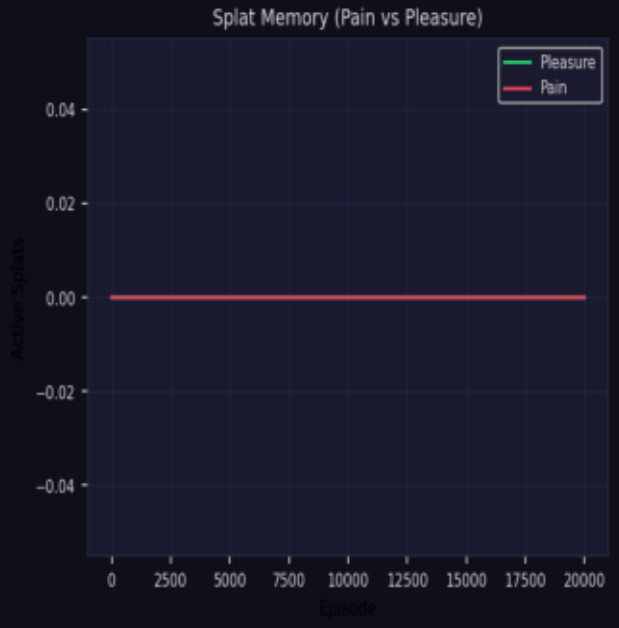
longrun_baseline_20000ep_seed42_2026-02-27_200138.png

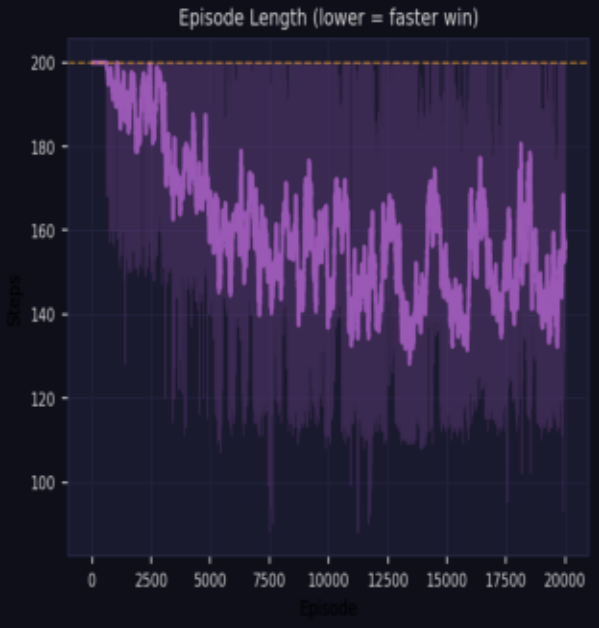
baseline | 20,000ep (seed=42) [20000 / 20000] 16772/20000 wins (83.9%) 103.4 ep/s







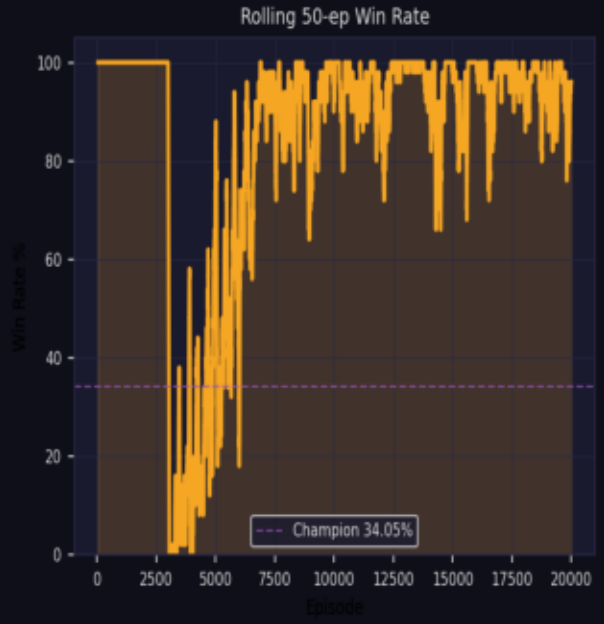


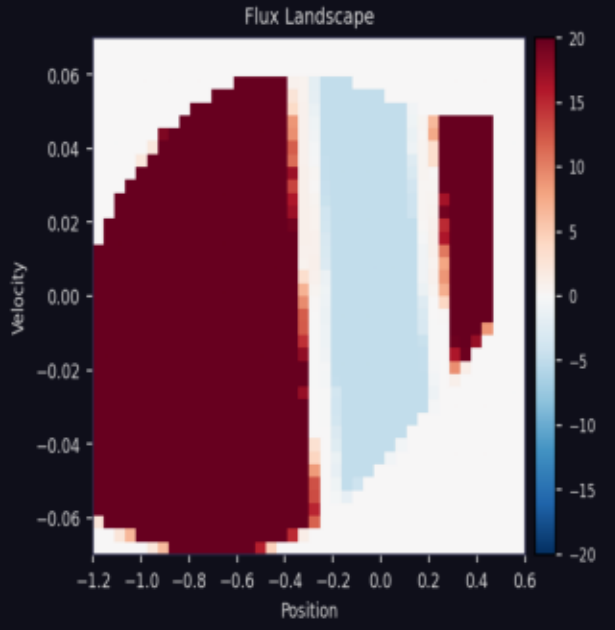


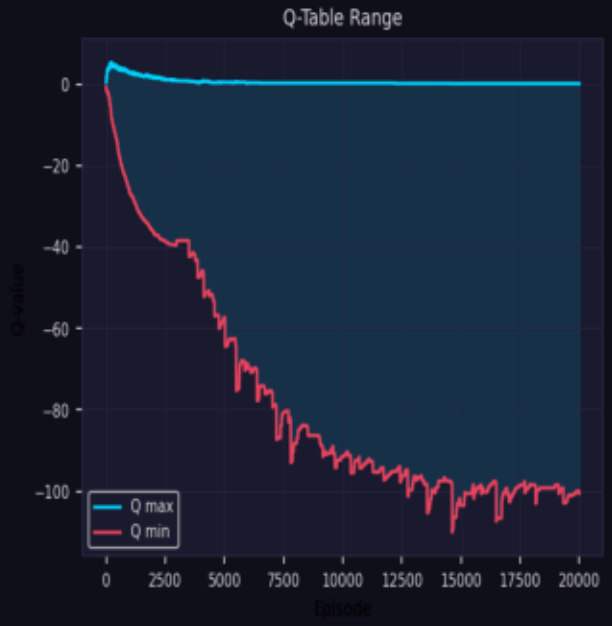
Plot 10/10 — Exp 2 — 20k: No-TDA live dashboard (seed 42)

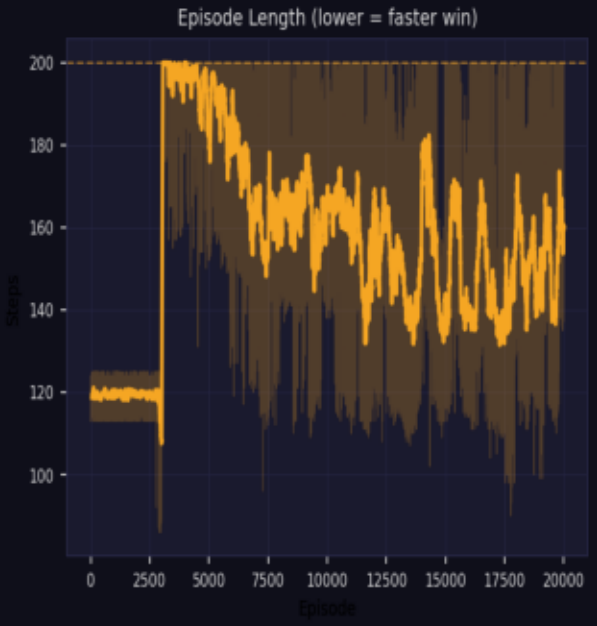
longrun_no_tda_20000ep_seed42_2026-02-27_190631.png

no_tda | 20,000ep (seed=42) [20000 / 20000] 16890/20000 wins (84.5%) 95.6 ep/s









Factual Summary — Observed Metrics (no interpretation)

Metric	2k Ablation	20k Long-Run
Highest total win rate	no_splats 78.1%	no_splats 88.6%
Lowest total win rate	baseline 25.1%	baseline 83.9%
Highest post-scaffold win rate	N/A (2k no scaffold split)	no_bridge 96.1%
Lowest post-scaffold win rate	N/A	no_tda 81.7%
Configs with first win at ep 0	full, no_tda, no_splats (bridge present)	full, no_tda, no_splats (bridge present)
Configs without ep-0 win	no_bridge (ep 433 avg), baseline (ep 479 avg)	no_bridge (ep 416 avg), baseline (ep 536 avg)
Gap: full vs baseline (total)	52.5pp	4.4pp
Gap: full vs no_bridge (total)	46.4pp	0.4pp
Seeds used	3	2
Total episodes per config	2,000	20,000
Governor scaffold off at	fixed ep 1,500	ep 3,000 (15%)

Notes:

- 'Post-scaffold' = win rate for episodes AFTER governor turned off (pure learned behavior).
- All win rates averaged over available seeds. Standard deviation available in JSON files.
- 'First win ep' = earliest episode in which position ≥ 0.5 was observed, averaged over seeds.
- Configs with bridge (full, no_tda, no_splats) begin with physics-seeded Q-table → ep-0 wins expected by design.
- Environment: MountainCar-v0, reward = −1/step, win = position ≥ 0.5, timeout = 500 steps.
- Raw JSON data: ablation_2026-02-27_184437.json | long_run_20000ep_2026-02-27_200455.json
- Generated: 2026-02-27 20:10