mcfeedback — Iteration 6: Reward Annealing

experiment-006.mjs · N = 10 seeds · Seeds: 42, 137, 271, 314, 500, 618, 777, 888, 999, 1234 · 1000 training episodes · Frozen-weight evaluation · Random chance = 50%

Change: smooth linear → squared reward blend over training. Base: experiment-004.

```
rewardAnnealStart  350  — blend begins at episode 350
rewardAnnealEnd    700  — blend completes at episode 700; pure squared thereafter
rewardExponent     2.0  — target exponent
flagStrengthGain / flagDecayRate / flagStrengthThreshold carried over from experiment-004.
```

Reward schedule across 1000 episodes

ep 1–349 linear **ep 350–700** blend **ep 701–1000** squared

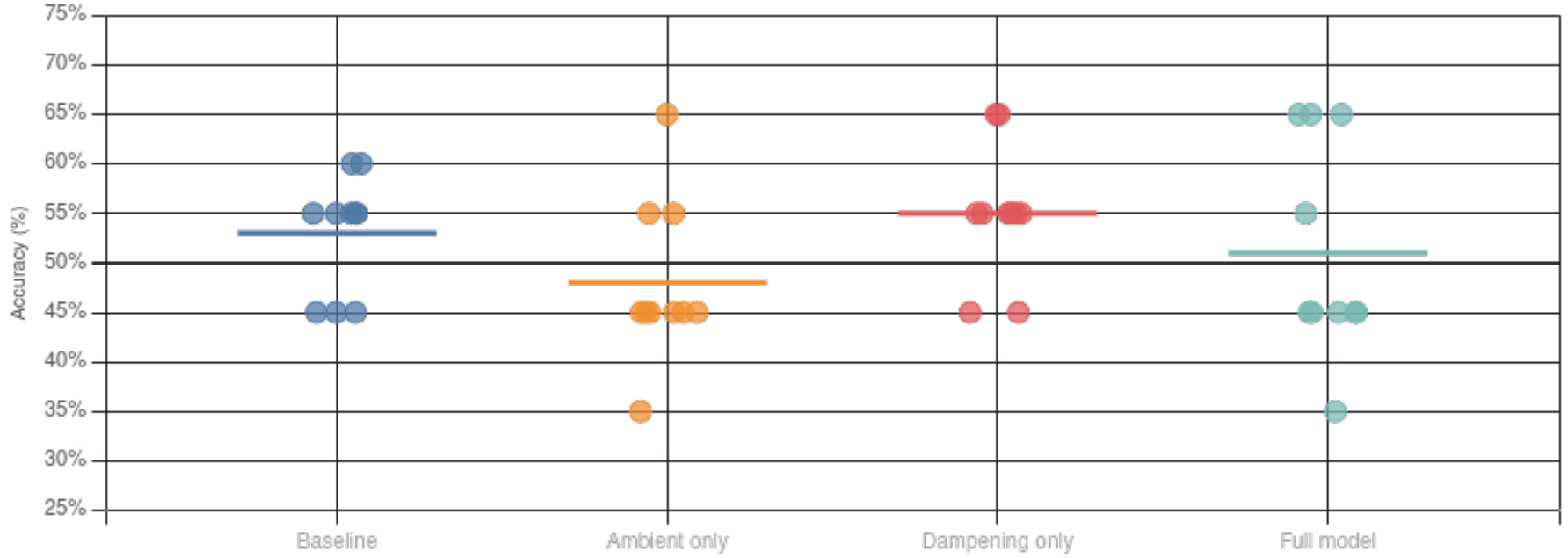
blend = clamp((ep − 350) / 350, 0, 1) · reward = (1−blend)×linear + blend×squared

Verdict: incremental — annealing partially recovers exp-005 regression but doesn't beat exp-004.

Full model reaches 51.0% mean (vs 53.0% in exp-004, 45.0% in exp-005). The 3 late seeds (888, 999, 1234) still hit 65%, but variance is now the worst yet at ±10.7% — one seed dropped to 35%, worse than any previous run. Annealing adds noise to the transition without delivering consistent benefit. **Dampening only is now the best condition at 55%**, suggesting the flag gate is interfering with what dampening alone could achieve. The reward shape is not the bottleneck — the 7-seed failure mode is upstream.

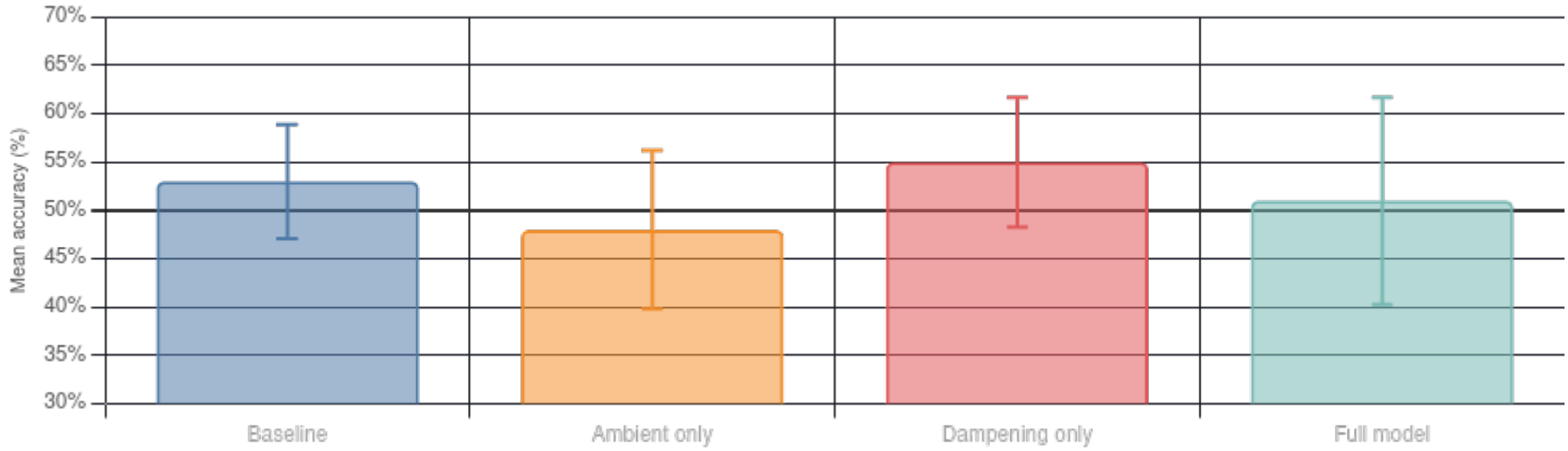
Baseline Ambient only Dampening only Full model

1 — ACCURACY DISTRIBUTION ACROSS SEEDS



Each dot = one seed. Horizontal line = mean. Dashed line = 50% random chance.

2 — MEAN ± 1 STD



Error bars show ±1 standard deviation across 10 seeds. Full model has the highest variance of any condition across all experiments.

3 — PAIRED T-TESTS VS BASELINE

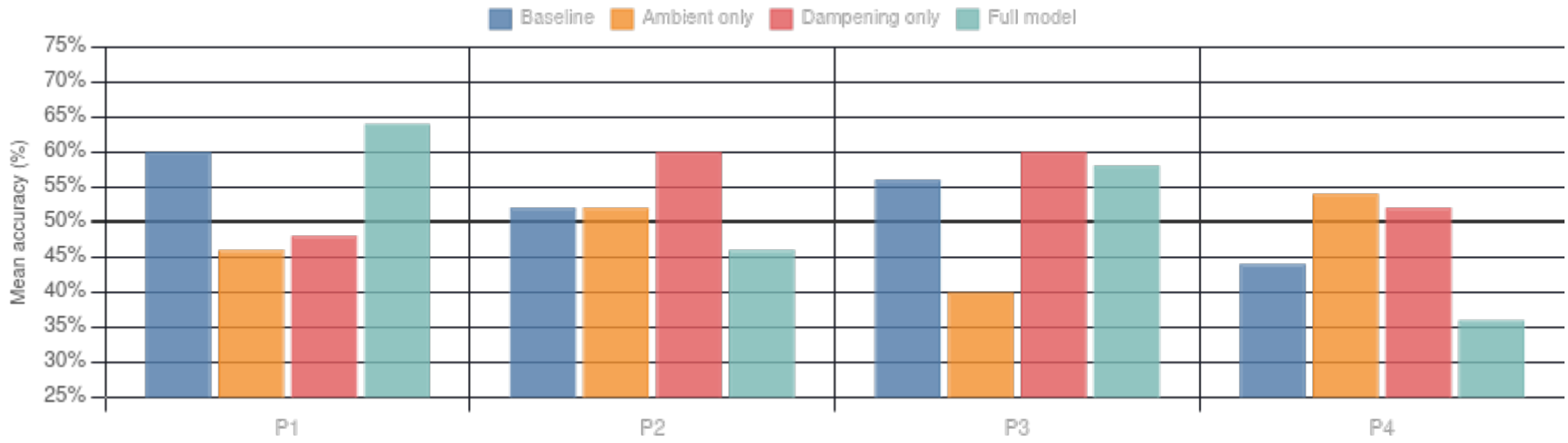
Comparison	Mean diff	t	p	Result
Ambient only vs Baseline	-5.0%	-1.3156	0.1630	**ns**
Dampening only vs Baseline	+2.0%	+0.8402	0.3118	**ns**
Full model vs Baseline	-2.0%	-0.5695	0.4302	**ns**

Two-tailed paired t-test, df=9. First experiment where Dampening only trends positive vs Baseline.

4 — RAW DATA (ALL SEEDS)

Seed	Baseline	Ambient	Dampening	Full
42	55%	45%	55%	35%
137	55%	45%	55%	45%
271	45%	45%	55%	45%
314	55%	45%	55%	45%
500	55%	45%	65%	45%
618	45%	55%	55%	45%
777	55%	65%	55%	55%
888	60%	35%	65%	65%
999	45%	55%	45%	65%
1234	60%	45%	45%	65%
Mean	**53.0%**	**48.0%**	**55.0%**	**51.0%**
Std	±5.9%	±8.2%	±6.7%	±10.7%

5 — PER-PATTERN ACCURACY (MEAN ACROSS SEEDS)



Full model again shows P1 specialisation (64%) at the cost of P4 (36%) — same asymmetry as experiment-004. The pattern is consistent across reward shapes, pointing to a structural wiring bias rather than a reward artefact.

The 7-seed failure is structural, not reward-shaped:

Seeds 42–618 fail in every experiment that uses the flag gate (004, 005, 006). Seeds 888–1234 succeed in every one of those experiments. The reward schedule (linear, squared, annealed) does not change which seeds succeed. This points to a wiring initialisation effect: some random graphs produce enough consistent co-activation signal for the flag gate to latch; others don't. Tuning reward won't fix this — the flag gate threshold or gain needs to be relaxed, or more training episodes are needed to let weak signals accumulate.

Progress across iterations (Full model mean):

Iter 1 (original): 45.5% max 55%

Iter 3 (flipped signs): 46.0% max 55%

Iter 4 (flag gate): 53.0% max 65% ← best mean

Iter 5 (flag + squared): 45.0% max 65% regression

Iter 6 (flag + anneal): 51.0% max 65% partial recovery

Next: return to exp-004 base, investigate the 7-seed failure — relax flag threshold or extend training.