Connect Four Experimentation

1 Goal and experimental protocol

We compare three training strategies for a *Predictive MinMax* agent on Connect Four (6×7):

- **Strategy A (baseline)**: constant parameters ($L = 2, K = 4$).
- **Strategy D (aggressive curriculum)**: staged increase of search depth and width.
- **Strategy E (constrained curriculum)**: rule $K = 2L + 1$.

The analysis combines: (i) training dynamics (loss), (ii) in-training performance (win-rate vs random), (iii) offline evaluation (MSE/correlation vs deeper MinMax), and (iv) online evaluation (100 games vs random).

2 Training logs and monitored variables

For each strategy, 100 training iterations (fixed seed) are available, tracking

`iter`, L, K, `batch_size`, `loss`,

and (every 10 iterations) an evaluation against a random opponent:

`wins_vs_random`, `draws_vs_random`, `losses_vs_random`.

Win-rate is defined as

$$\text{winrate} = \frac{\text{wins}}{\text{wins} + \text{draws} + \text{losses}}.$$

3 Training dynamics

3.1 Training loss

Figure 1 shows the loss (MSE) over training iterations for A, D, and E.

Figure 1: Training loss (MSE) vs iteration for the three strategies.

Key observations. All strategies exhibit an overall decreasing trend; oscillations and occasional spikes are expected in self-play with non-stationary targets. Strategy E tends to reduce loss faster in the mid-training range (approximately iterations 30–60), while D appears more regular in the late phase.

3.2 Win-rate vs random during training

Table 1 reports win-rate vs random measured every 10 iterations.

Table 1: Win-rate vs random during training (evaluation points every 10 iterations).

Strategy	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
A	0.20	0.50	0.45	0.20	0.35	0.20	0.30	0.20	0.00	0.10
D	0.70	0.15	0.15	0.70	0.45	0.50	0.75	0.80	0.55	0.60
E	0.70	0.15	0.40	0.75	0.35	0.60	0.50	0.45	0.35	0.70

Figure 2 visualizes the same points.

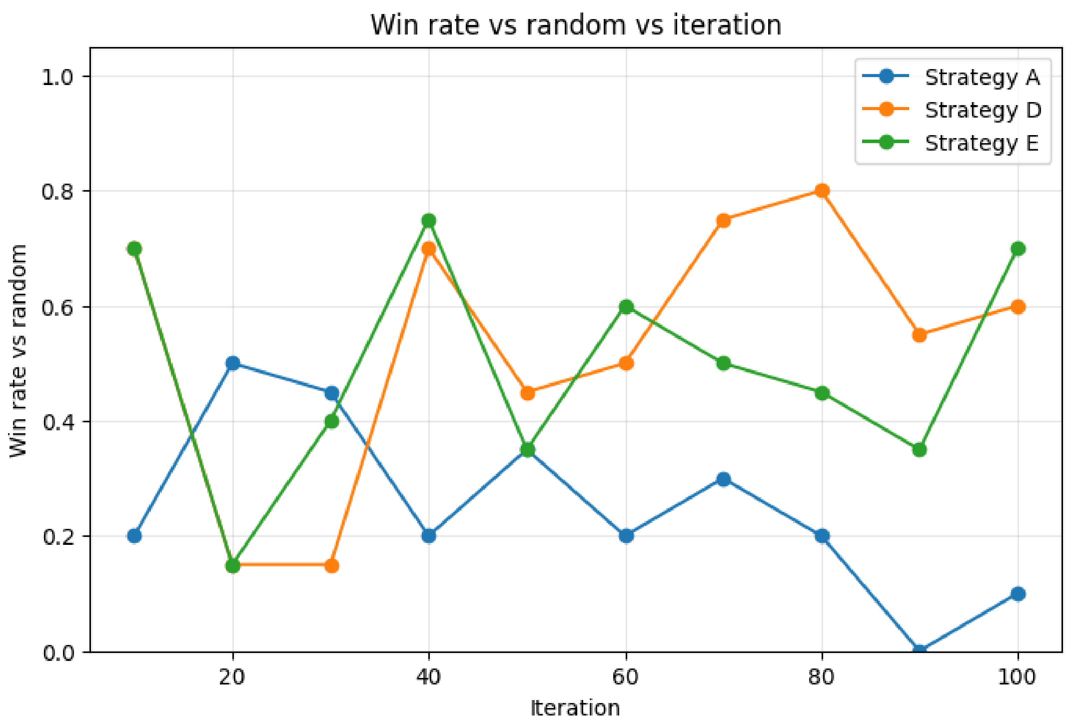

Figure 2: Win-rate vs random (evaluations every 10 iterations) for A, D, E.

Summary. Curriculum-based strategies (D, E) outperform the baseline A at most evaluation points. D shows a more consistent profile in late training (70–100), whereas E achieves high peaks but with higher variability.

4 Offline evaluation (vs deeper MinMax)

We compare network predictions $H_{\text{true}}(s)$ to "ground-truth" values computed via a deeper MinMax search (without heuristic) on 200 test states. We report MSE (lower is better) and Pearson correlation (higher is better).

Table 2: Offline evaluation (200 states): MSE and correlation vs deeper MinMax.

Strategy	Name	MSE	Corr.
A	Baseline (constant L, K)	0.273999	0.099216
D	Aggressive curriculum (Strategy D)	0.174261	0.174923
E	Curriculum v3 ($K = 2L + 1$)	0.181715	0.138203

Interpretation. D achieves the closest approximation of deeper MinMax (lowest MSE and highest correlation); E follows; A is clearly weaker on both metrics.

5 Online evaluation vs random (100 games)

We run 100 games against a random opponent using (L, K) pairs consistent with each strategy:

$$\text{A: } (2, 4), \quad \text{D: } (3, 7), \quad \text{E: } (3, 7).$$

Table 3: Results vs random (100 games).

Strategy	(L,K)	Wins	Draws	Losses	Win-rate
A	(2,4)	94	0	6	0.94
D	(3,7)	97	0	3	0.97
E	(3,7)	94	0	6	0.94

6 Overall summary

Table 4: Summary: heuristic quality (offline) and playing strength (vs random).

Strategy	Offline MSE	Offline Corr.	Win-rate vs random
A	0.273999	0.099216	0.94
D	0.174261	0.174923	0.97
E	0.181715	0.138203	0.94

Conclusion. **Strategy D** is best overall: strongest offline agreement with deeper MinMax and highest win-rate vs random. **Strategy E** ranks second in offline quality but does not yield a stable advantage over A in the online test considered. **Strategy A** is consistently weaker offline and does not match D online.

Supplementary material

The complete Jupyter notebook (full code and outputs) is available at:
`https://github.com/marco-calabrese93/MinMaxExp/blob/4e2f814bc54a03a3cffbd749f4dc935966a85437/experiments/ANALYSIS_FINAL.ipynb`