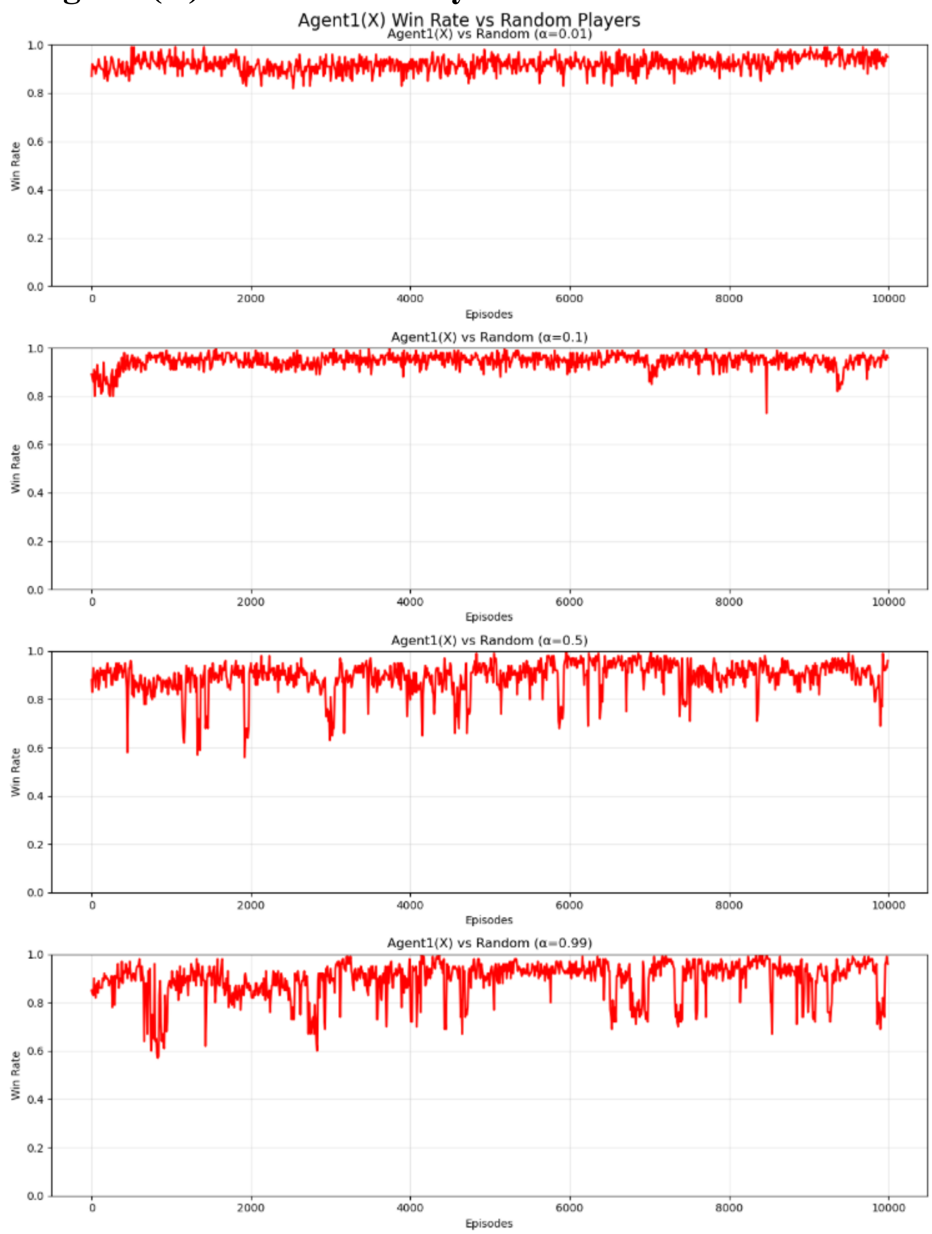
Graph 1: Agent1(X) vs Random Players

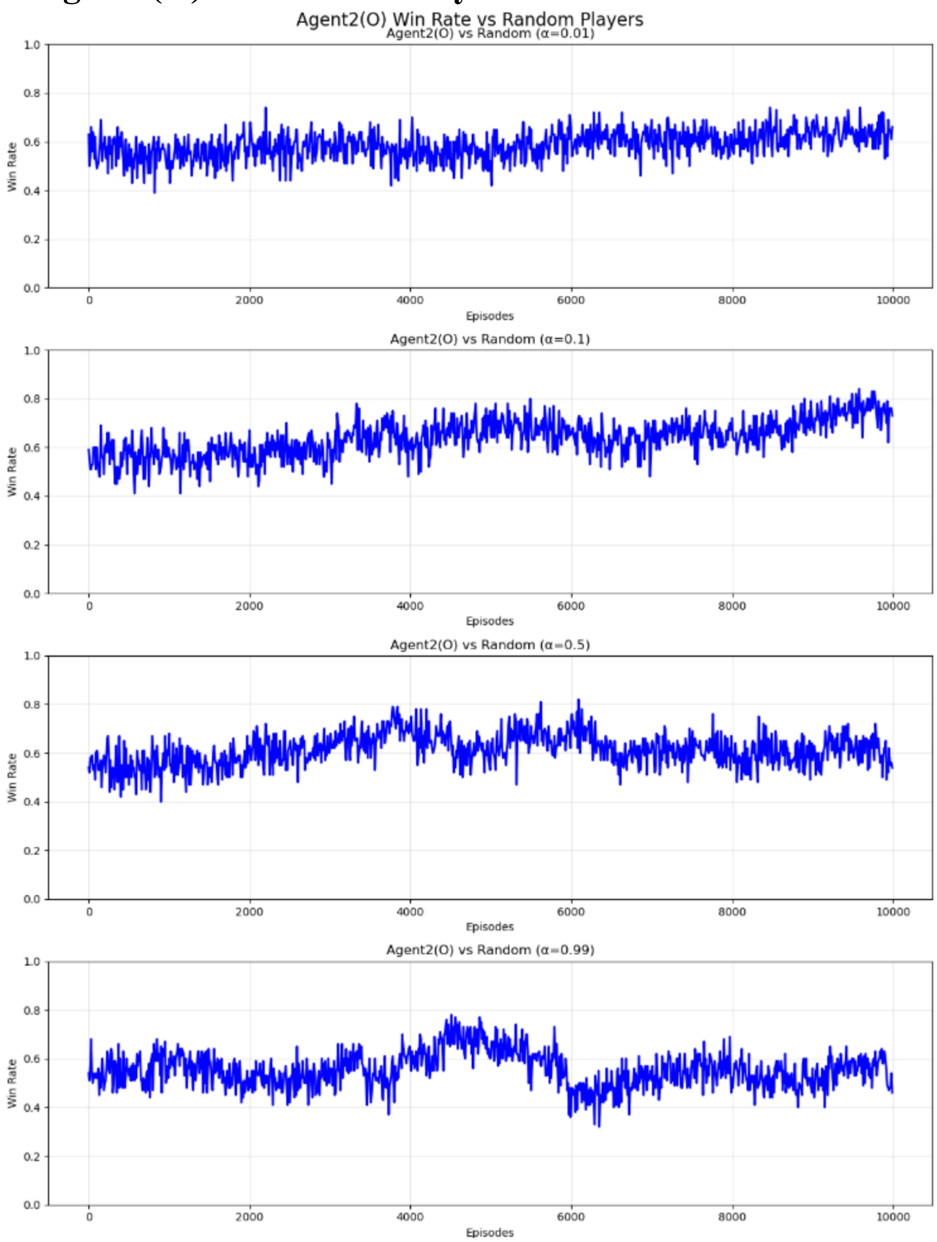


What it shows: How well X agent beats random player

1. **α = 0.01:** Smooth line at ~95% - X almost always wins, very stable
    - Why: Learns very slowly → doesn't change good strategies → stays consistent
2. **α = 0.1:** Mostly at ~95% with small wiggles - X wins consistently
    - Why: Learns slowly → keeps good moves → minor updates don't break what works
3. **α = 0.5:** Bumpy, drops to 60-80% - X wins most times but sometimes struggles
    - Why: Learns faster → sometimes overwrites good strategies → performance drop
4. **α = 0.99:** Very bumpy, big drops - X wins but very inconsistent
    - Why: Learns too fast → constantly changing strategy → forgets what worked before

Meaning: Lower alpha = X beats random players more reliably

📊 Graph 2: Agent2(O) vs Random Players

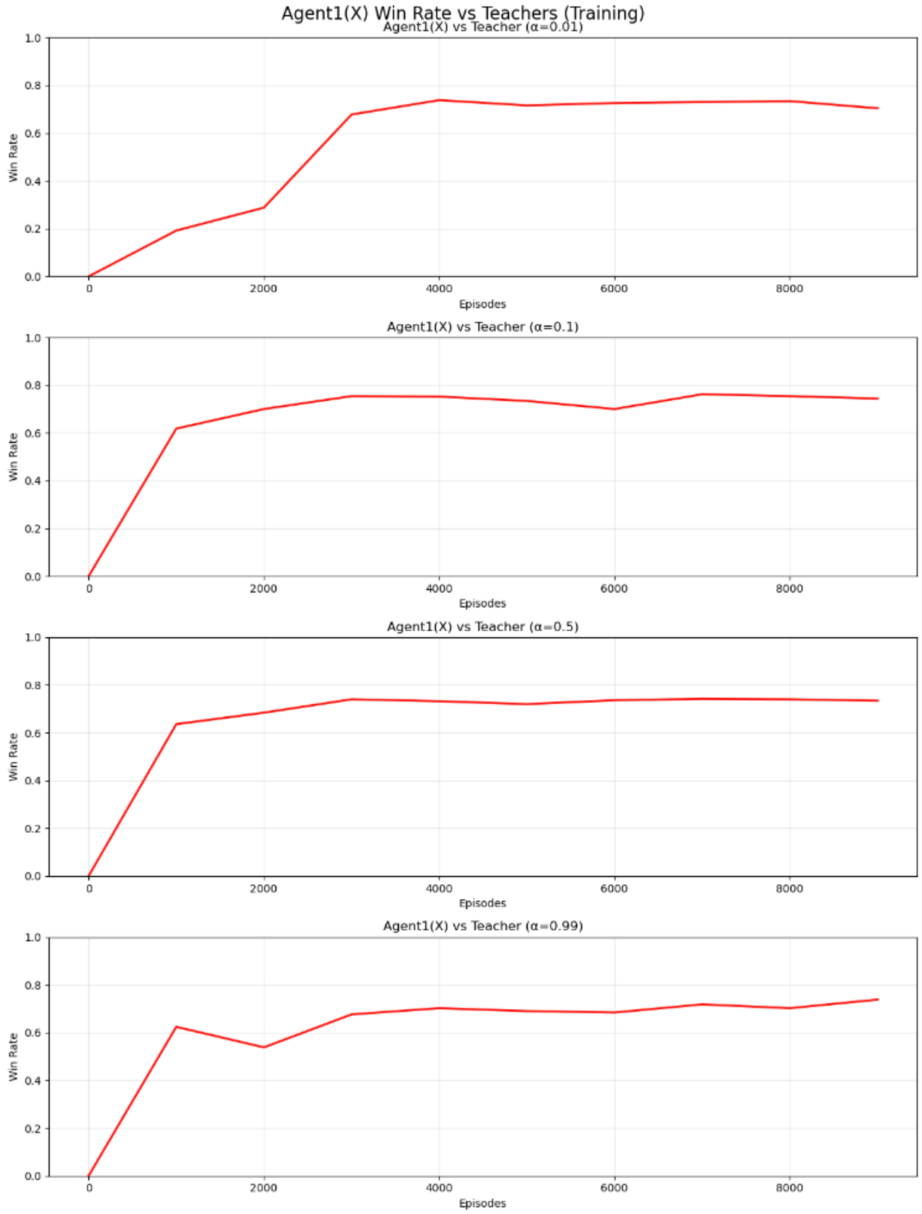


What it shows: How well O agent beats random players

1. **α = 0.01:** Steady line at ~60% - O wins about half the time, stable
   a. Why: Slow learning keeps O consistent, but O naturally weaker (goes second)
2. **α = 0.1:** Slowly improves 50% → 75% - O gets better over time
   a. Why: Good learning pace → gradually finds better strategies → steady improvement
3. **α = 0.5:** Wiggly around 60% - O wins sometimes, inconsistent
   a. Why: Medium learning speed → sometimes finds good moves, sometimes loses them
4. **α = 0.99:** Very bumpy 40-70% - O performance all over the place
   a. Why: Changes strategy too often → can't stick to what works → unstable results

Meaning: O is naturally weaker than X, lower alpha helps it be more consistent

📊 Graph 3: Agent1(X) vs Teachers

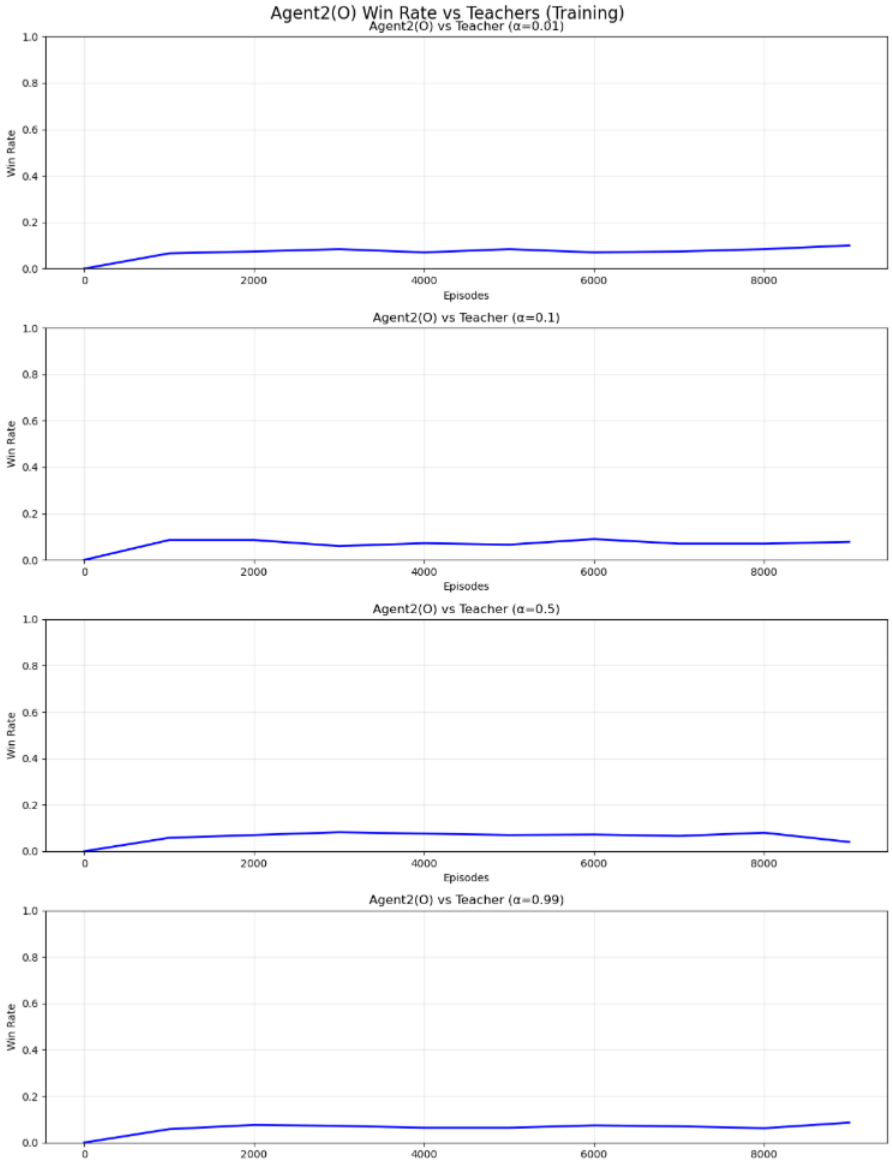


What it shows: How well X learns to beat smart teachers

1. α = 0.01: Smooth climb 0% → 75% - X learns steadily, reaches 75% wins
   a. Why: Slow, careful learning → builds solid strategy against teachers → no forgetting
2. α = 0.1: Good climb to 75% then flat - X learns well, stays at 75%
   a. Why: Balanced learning → finds good counterstrategies → maintains performance
3. α = 0.5: Quick rise to 70% then flat - X learns fast but stops at 70%
   a. Why: Fast learning → quickly finds some strategies but overwrites before perfecting
4. α = 0.99: Bumpy rise then dip - X learns fast but forgets, ends lower
   a. Why: Too fast learning → finds strategies quickly but immediately forgets them → chaos

Meaning: Lower alpha = X learns better against smart opponents

📊 Graph 4: Agent2(O) vs Teachers



What it shows: How well O learns to beat smart teachers

1. **α = 0.01:** Flat line at ~10% - O barely wins, but steady
    a. Why: Even slow learning can't help much → O's disadvantage too big → at least stable
2. **α = 0.1:** Flat line at ~8% - O almost never wins
    a. Why: Teachers too good + O goes second → very hard to find winning strategies
3. **α = 0.5:** Flat line at ~5% - O rarely wins
    a. Why: Fast learning makes it worse → keeps changing losing strategies → no improvement
4. **α = 0.99:** Flat line at ~5% - O almost never wins
    a. Why: Chaotic learning → can't develop any consistent strategy → stays bad

Meaning: O really struggles against smart teachers no matter what alpha

Learning Rate Effects

1. **Lower α (0.01-0.1): Conservative Learning**

    a. Stable updates: Small changes preserve good knowledge
    b. Consistent performance: Less volatility in win rates
    c. Better retention: Doesn't overwrite successful strategies

2. **Higher α (0.5-0.99): Aggressive Learning**

    a. Rapid adaptation: Quick learning but unstable
    b. Volatility: Overwrites previous knowledge too quickly
    c. Inconsistent: Performance swings wildly

Agent Asymmetry

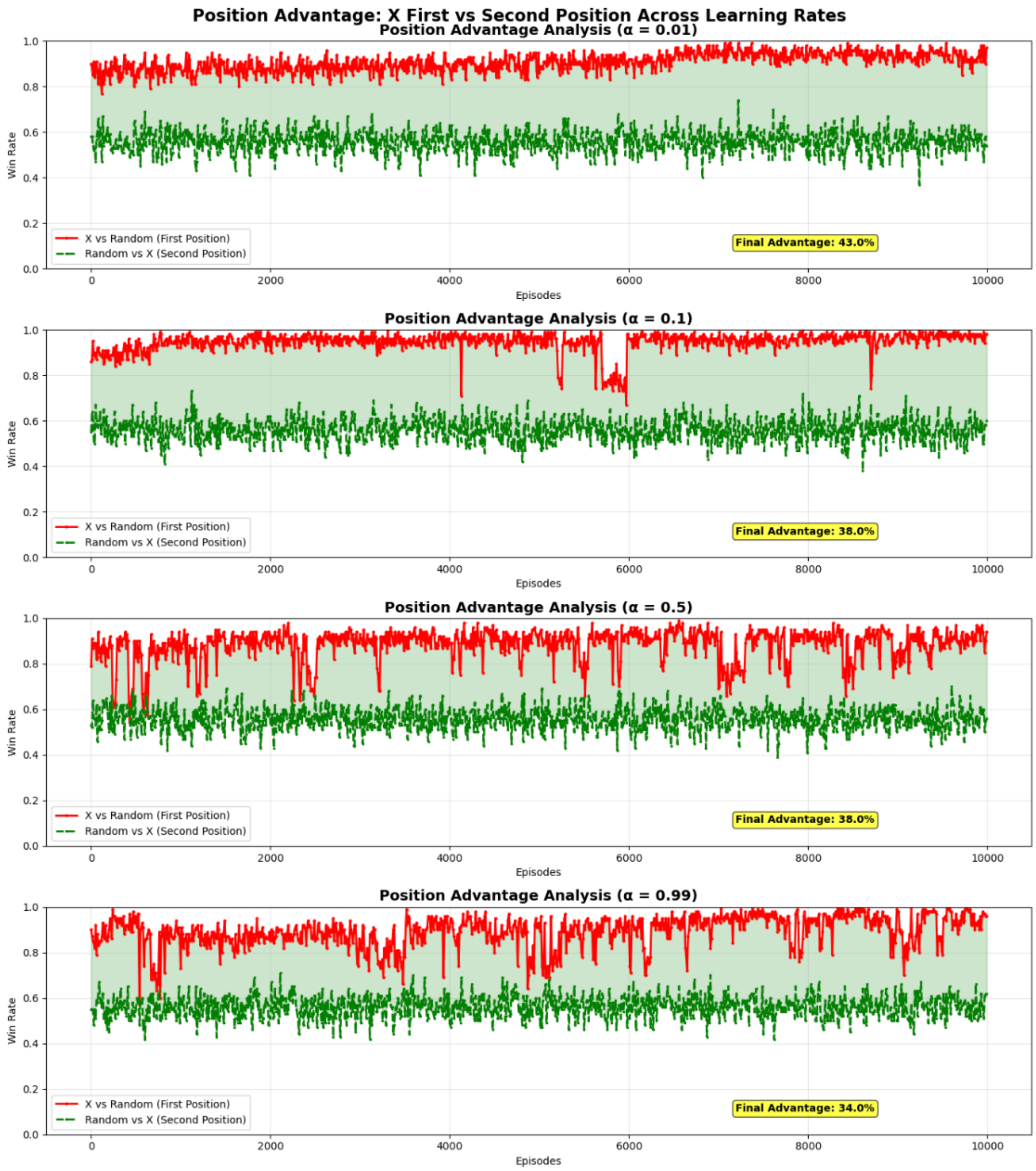
1. **Agent1 (X) Superior:**

    a. First-move advantage: X plays first in Tic-Tac-Toe
    b. Strategic position: Easier to control center/corners
    c. Training benefit: Better learning from teacher interactions

2. **Agent2 (O) Struggles:**

    a. Reactive role: Always responding to X's moves
    b. Harder learning: Must counter optimal X play
    c. Teacher mismatch: May not learn effective counter strategies

X Agent (Trained to Play First)



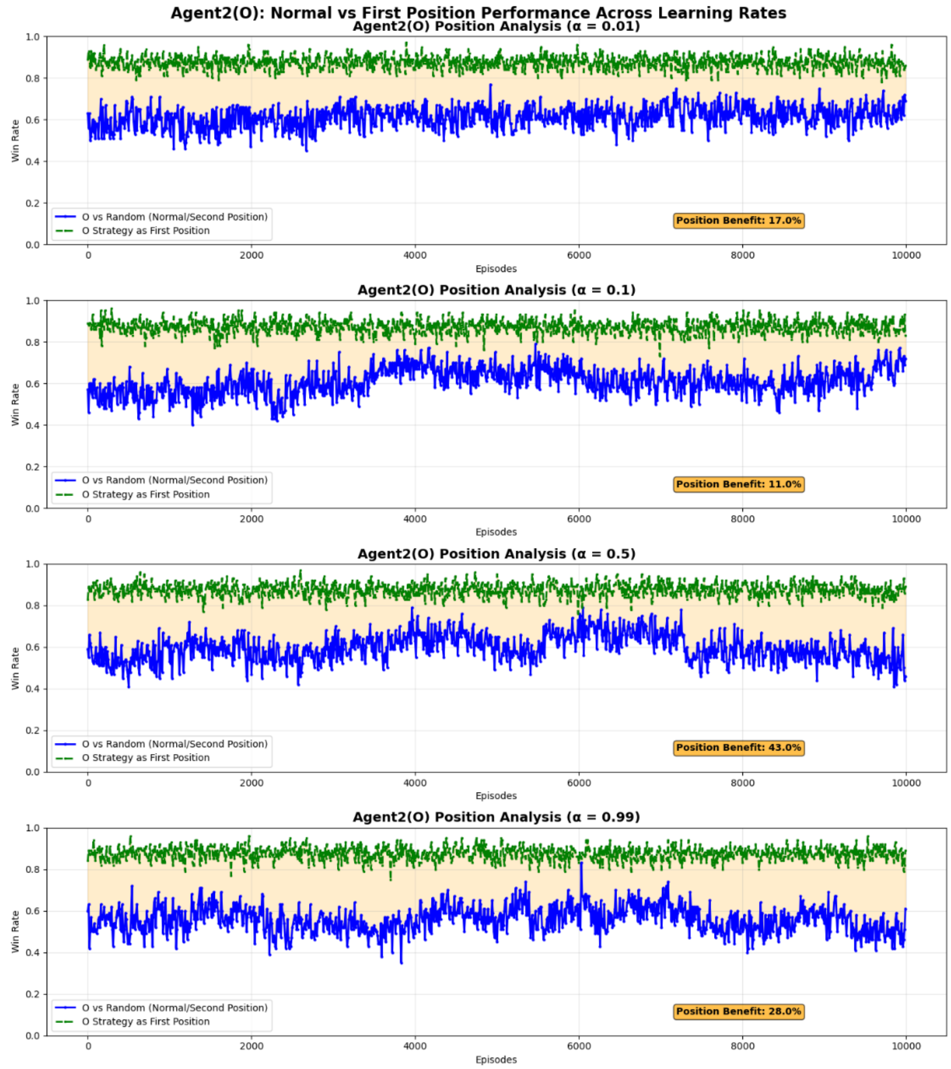
What the graphs show:

- **Red line (Role Trained)**: X playing normally (going first) - does great!
- **Green line (Role Switched)**: X trying to play second - does okay but much worse

Main findings:

- **All learning speeds work well** - X always wins about 95% of games when playing first
- **But X is terrible at switching roles** - drops to only 55% wins when playing second
- **Slower learning (α=0.01)** = more steady, less jumpy lines
- **Faster learning (α=0.99)** = very jumpy, unstable performance

O Agent (Trained to Play Second)



What the graphs show:

- **Blue line (Role Trained):** O playing normally (going second) - struggles more
- **Green line (Role Switched)**: O trying to play first - does better!

Main findings:

- **Learning speed matters A LOT** for O agent:
  - α=0.1 (medium speed): Best performance - wins 73% of games
  - α=0.99 (very fast): Worst performance - only wins 46% of games
- **O improves when switching to go first -** gets easier games
- **High learning speeds make O very unstable -** performance jumps around wildly

O's training created a well-rounded understanding of the game, and when combined with the natural first-player advantage, this knowledge becomes more effective than when constrained to the disadvantaged second position.