Effective Short-Term Opponent Exploitation in Simplified Poker

Supplemental Online Appendix

Finnegan Southey, Bret Hoehn, Robert C. Holte
University of Alberta, Dept. of Computing Science

October 6, 2008

This appendix contains plots for a wider range of opponents than was practical to include in the main article. It also includes some sample results for P2 modelling P1.

1 Player 1 vs. Player 2 Results

1.1 Convergence Rate Study

Figures 1 through 6. The long-term behaviour of the different methods is similar for all opponents but note the variation in the early performance of parameter learning due to effectiveness of the prior against each opponent. There are also several examples of the early non-monotonicity discussed at length in the main paper.

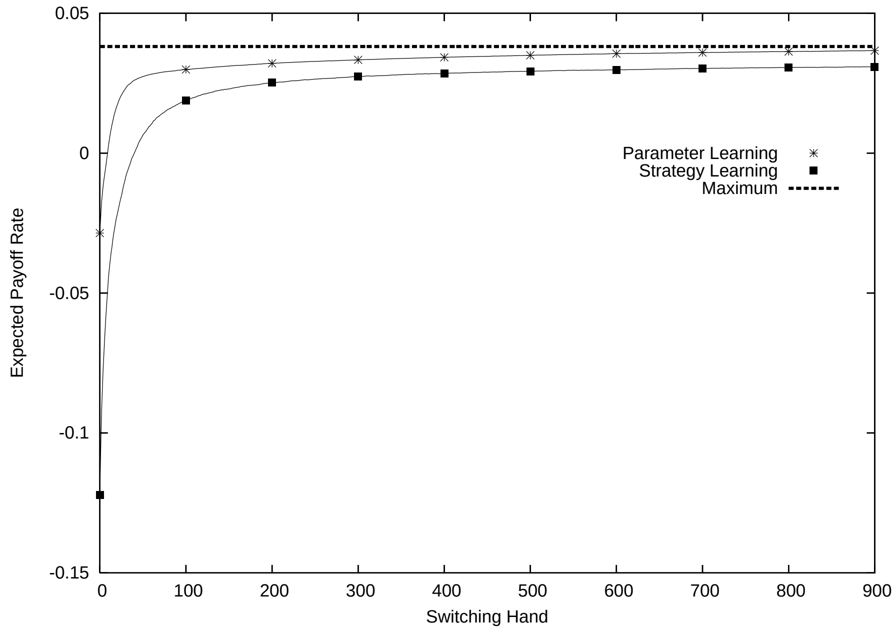

Figure 1: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against O_1.

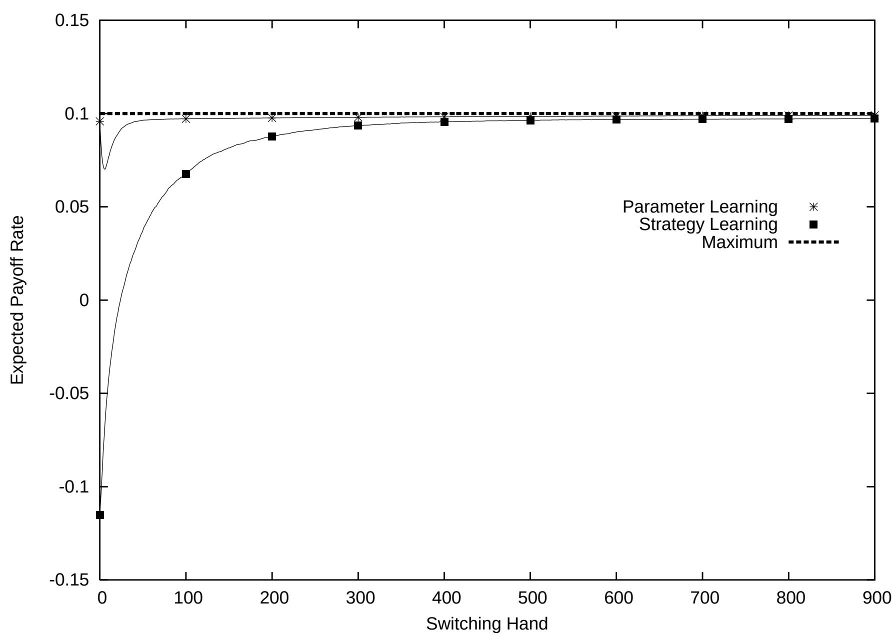

Figure 2: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against O_2.

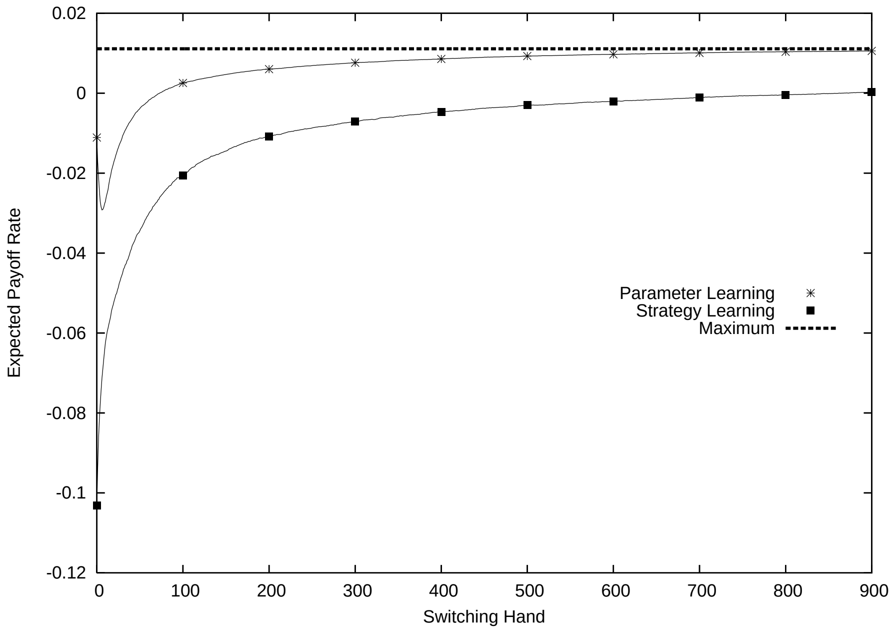

Figure 3: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against O_3.

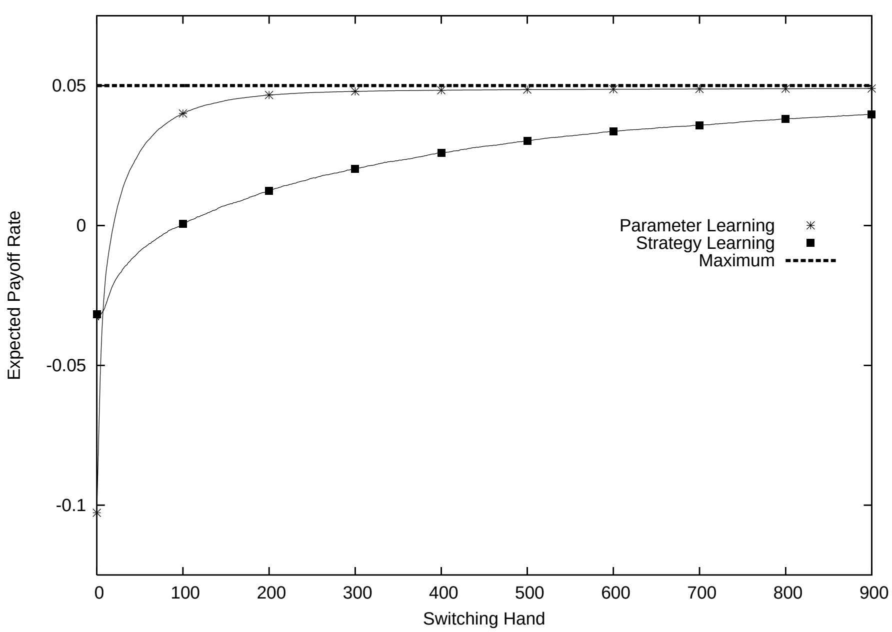

Figure 4: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against O_4.

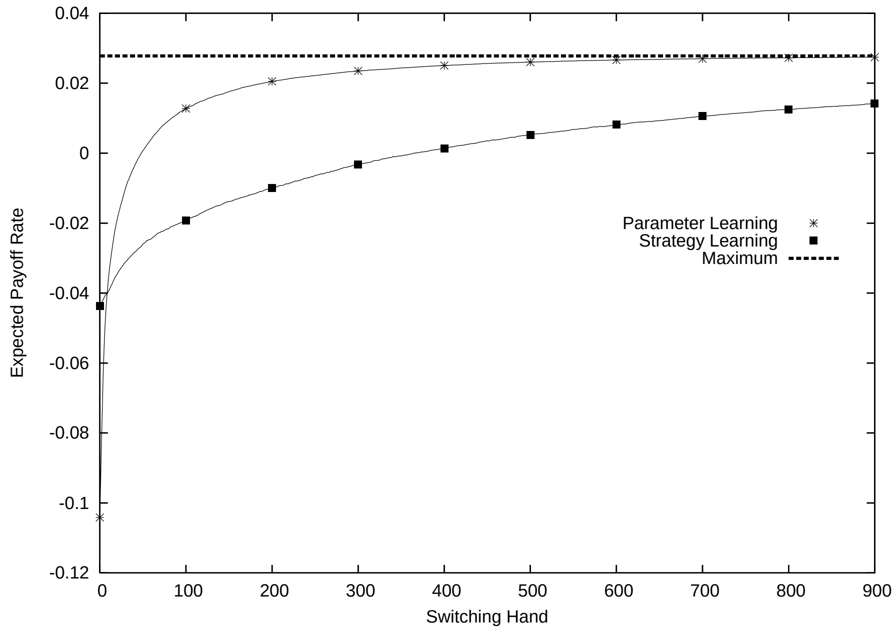

Figure 5: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against O_5.

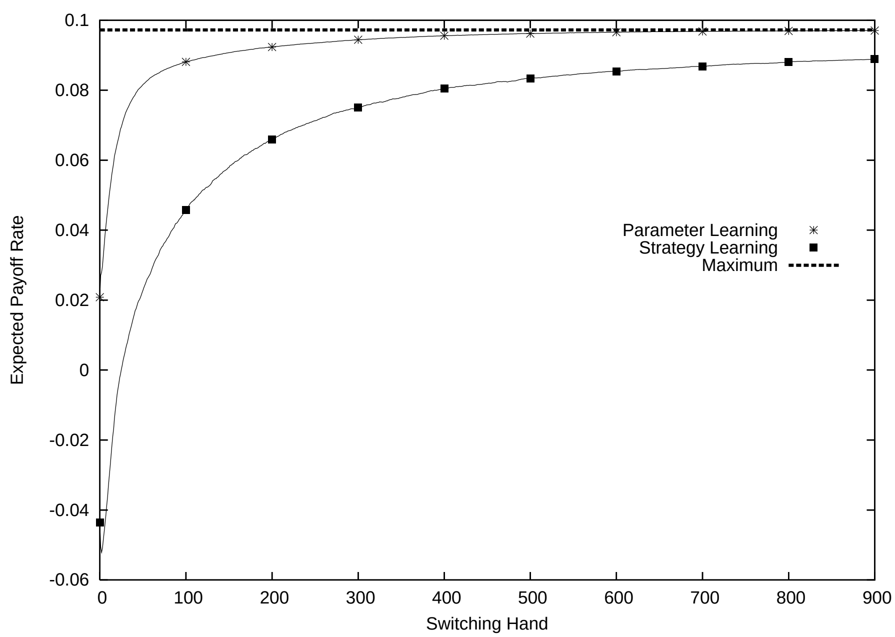

Figure 6: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against O_6.

1.2 Prior Study

Figures 7 through 12. Opponent 3 is particularly interesting in that the stronger default priors perform better than the weaker, unlike the other opponents. Also note how the weak bad prior often outperforms the strong default prior, reinforcing the notion that quick adaptation is paramount and that little or no prior is generally better than any strong guess.

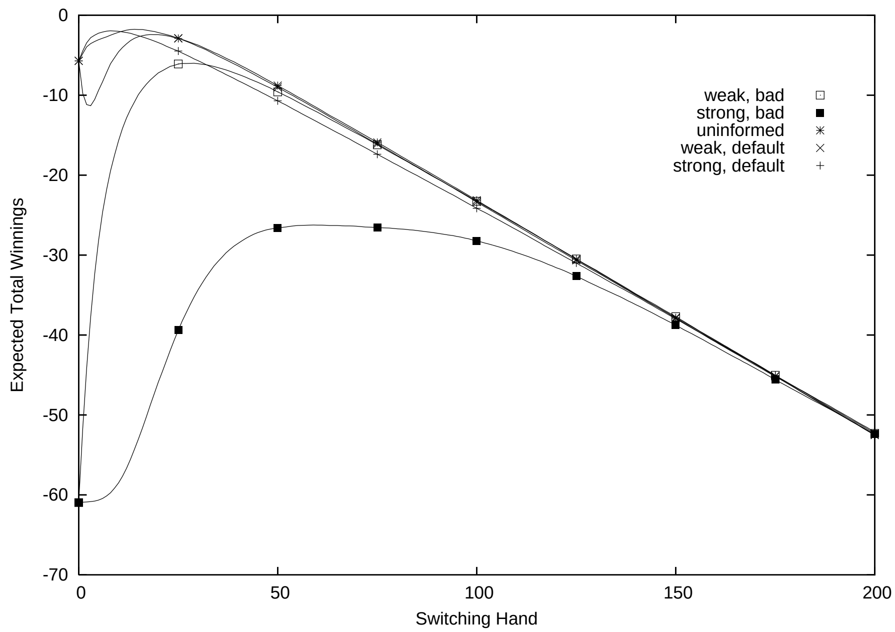

Figure 7: Prior Study: Four different priors for parameter learning against O_1.

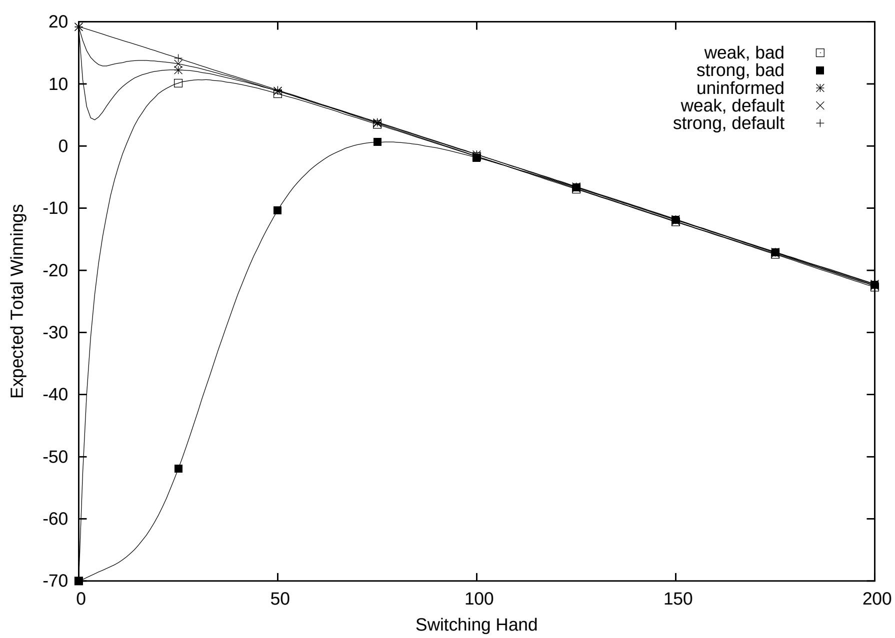

Figure 8: Prior Study: Four different priors for parameter learning against O_2.

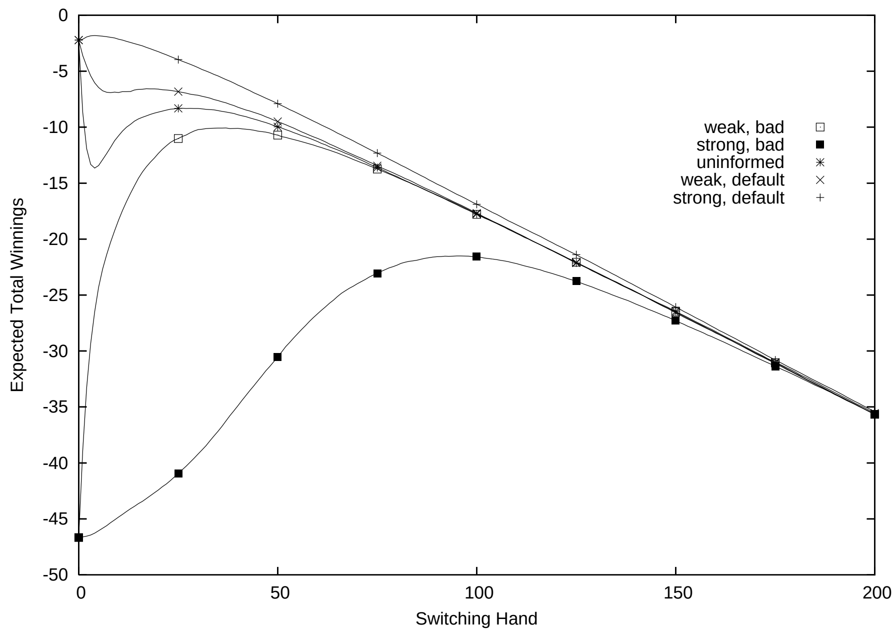

Figure 9: Prior Study: Four different priors for parameter learning against O_3.

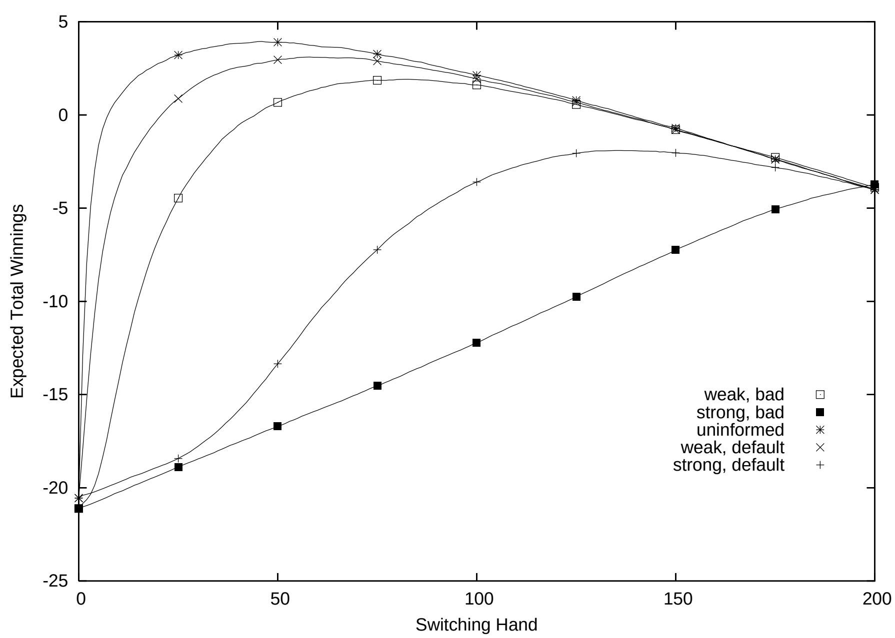

Figure 10: Prior Study: Four different priors for parameter learning against O_4.

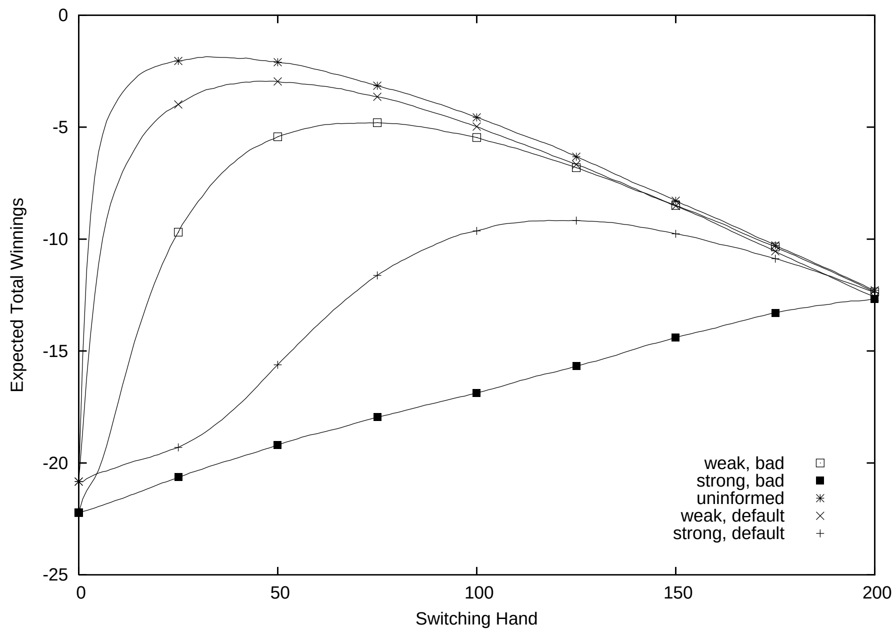

Figure 11: Prior Study: Four different priors for parameter learning against O_5.

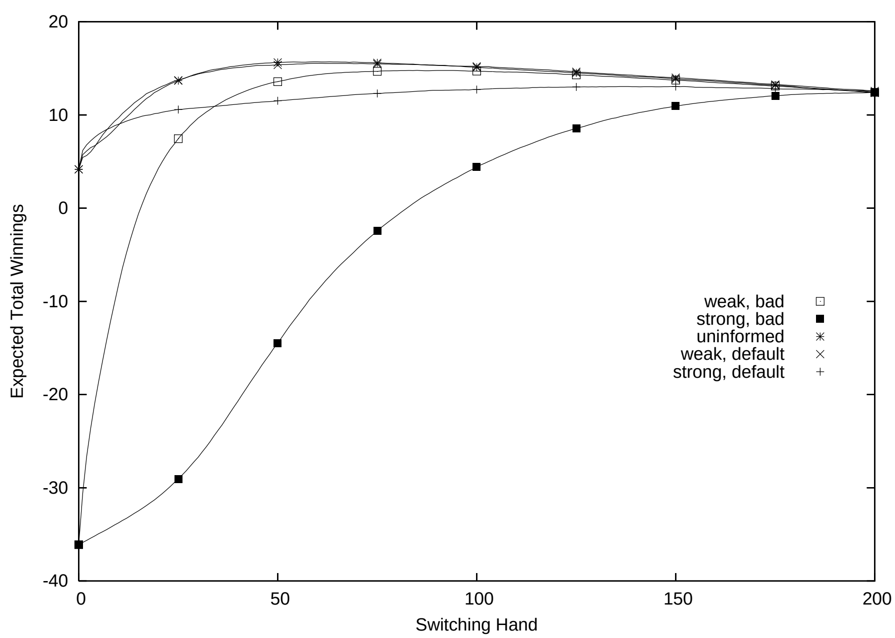

Figure 12: Prior Study: Four different priors for parameter learning against O_6.

1.3 Nash Exploration Study

Figures 13 through 18. This study shows that $\gamma = 1$ is the most reliably choice and that $\gamma = 0$ varies dramatically depending on opponent and without apparent reference to its nearest neighbour, $\gamma = 0.25$. We expect that in more complex pokers, it will be unlikely that any single Nash strategy dominates the rest in terms of exploration.

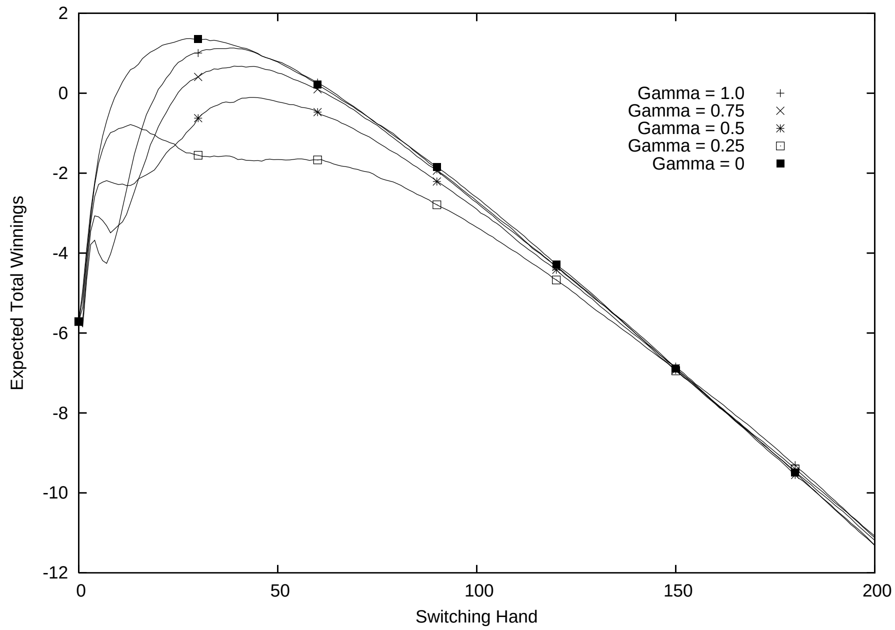

Figure 13: Nash Study: Expected total winnings vs. switching hand for parameter learning using various Nash strategies for exploration against O_1.

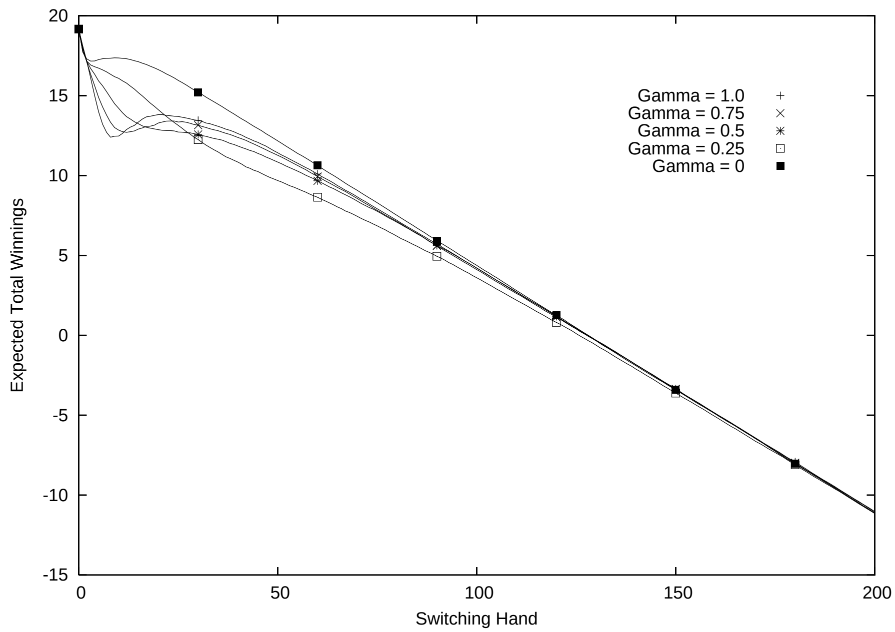

Figure 14: Nash Study: Expected total winnings vs. switching hand for parameter learning using various Nash strategies for exploration against O_2.

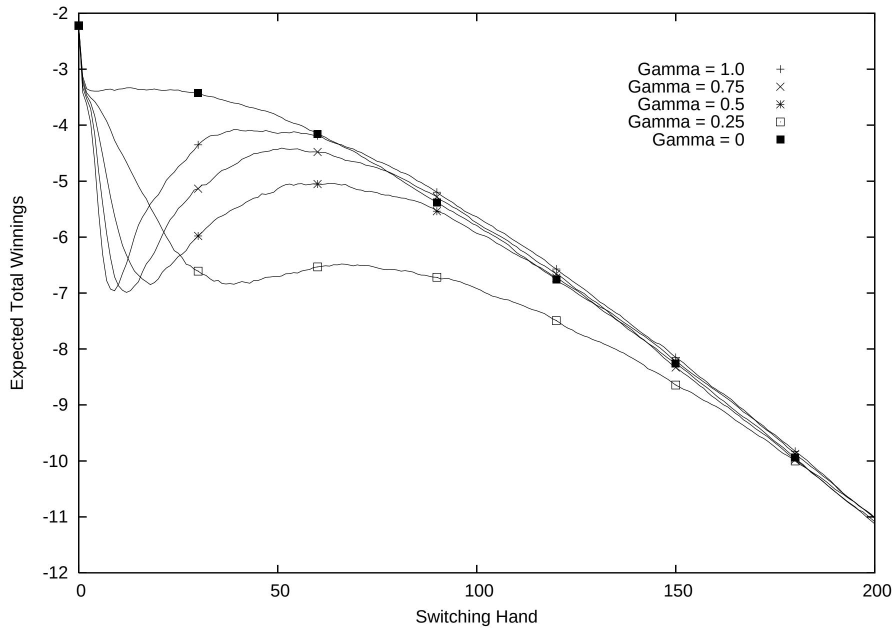

Figure 15: Nash Study: Expected total winnings vs. switching hand for parameter learning using various Nash strategies for exploration against O_3.

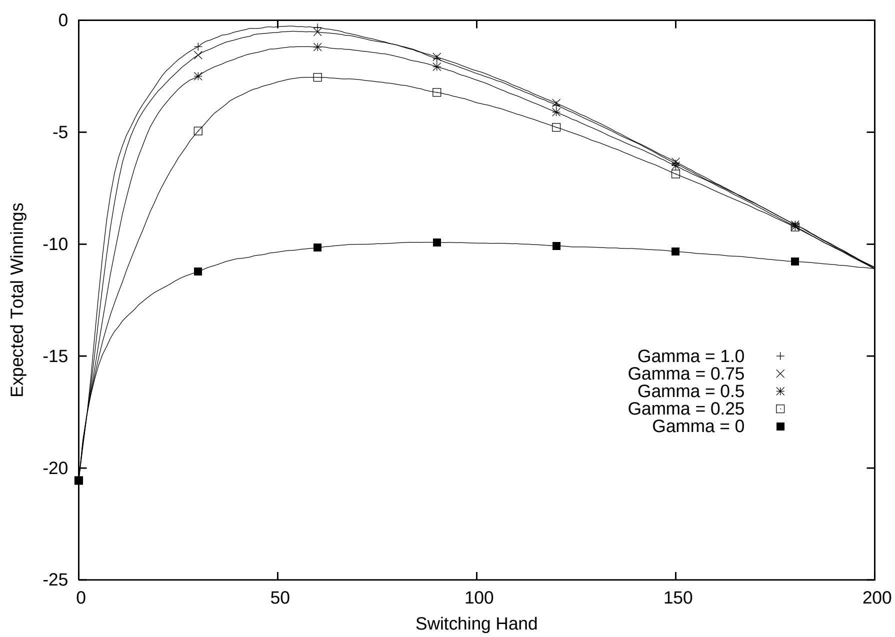

Figure 16: Nash Study: Expected total winnings vs. switching hand for parameter learning using various Nash strategies for exploration against O_4.

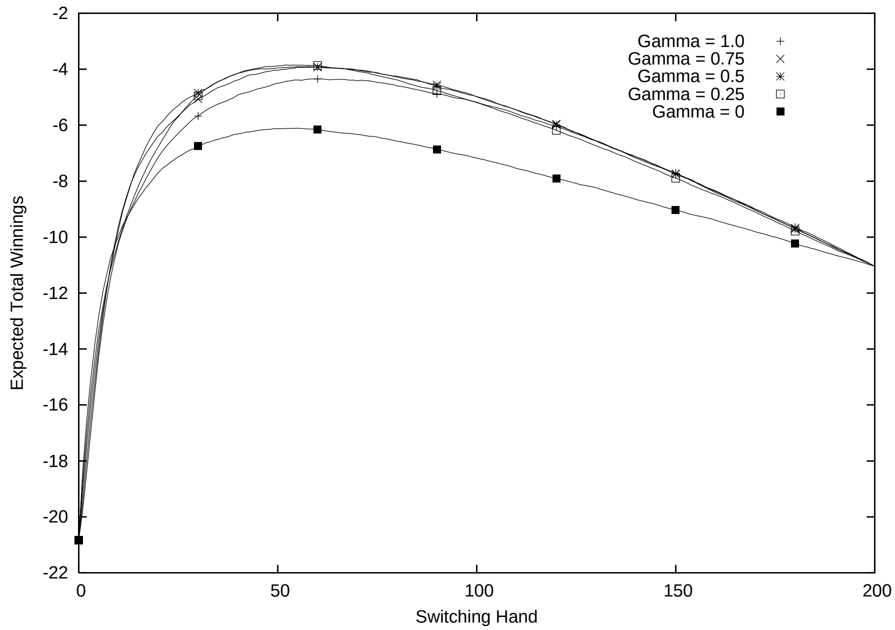

Figure 17: Nash Study: Expected total winnings vs. switching hand for parameter learning using various Nash strategies for exploration against O_5.

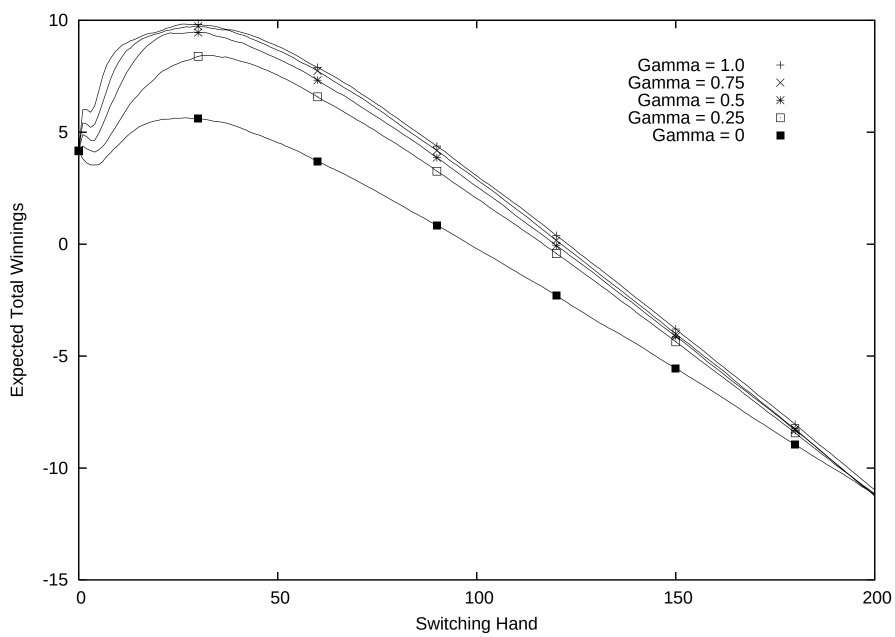

Figure 18: Nash Study: Expected total winnings vs. switching hand for parameter learning using various Nash strategies for exploration against O_6.

1.4 Learning Method Comparison

Figures 19 through 24. It is important when viewing these plots to remember that the peak of each curve is the point to be compared. Overall, strategy learning fare poorly and the two different exploration choices each work best for half of the opponents. Note that for O_2 and O_3, it is best to skip learning altogether as the game is not long enough to allow any learner to recover from initial missteps.

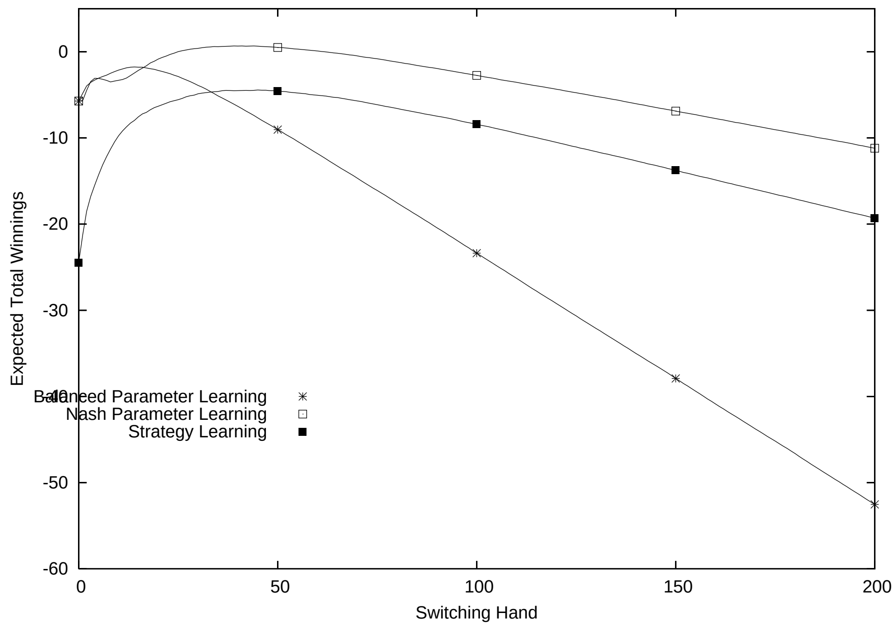

Figure 19: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against O_1.

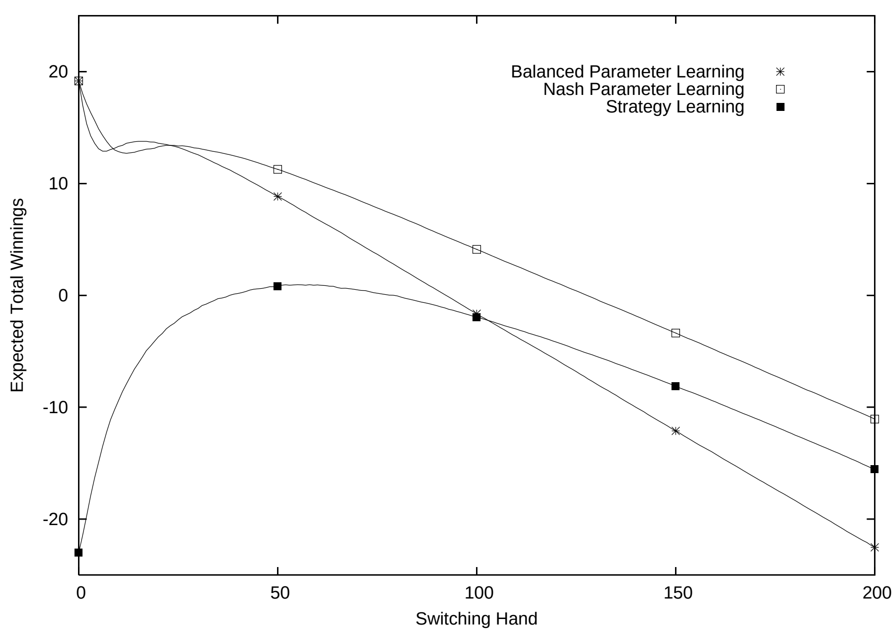

Figure 20: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against O_2.

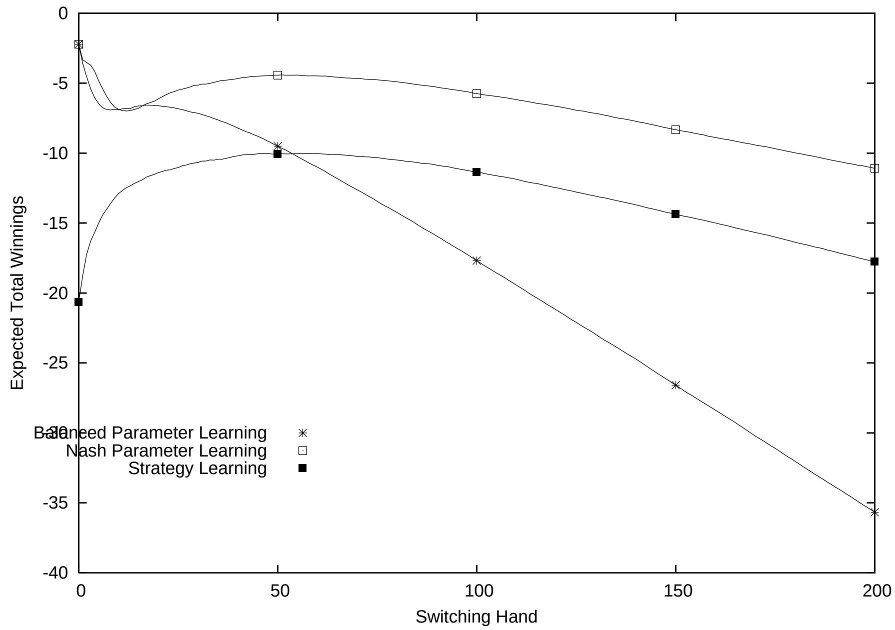

Figure 21: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against O_3.

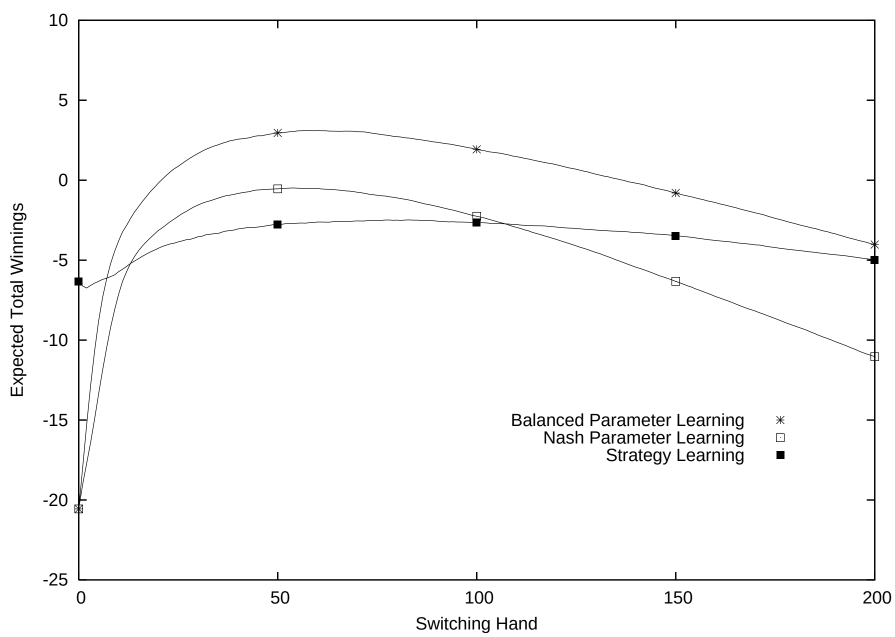

Figure 22: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against O_4.

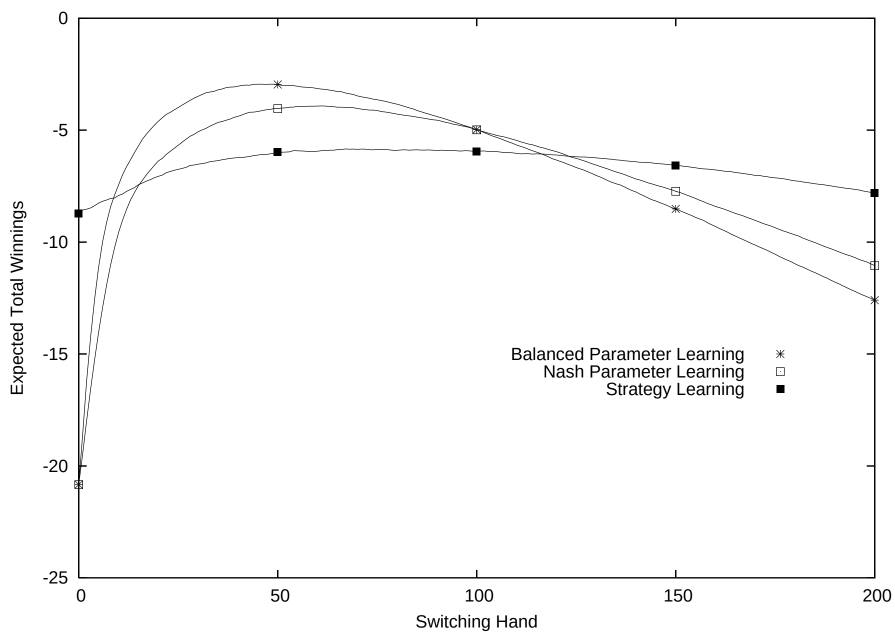

Figure 23: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against O_5.

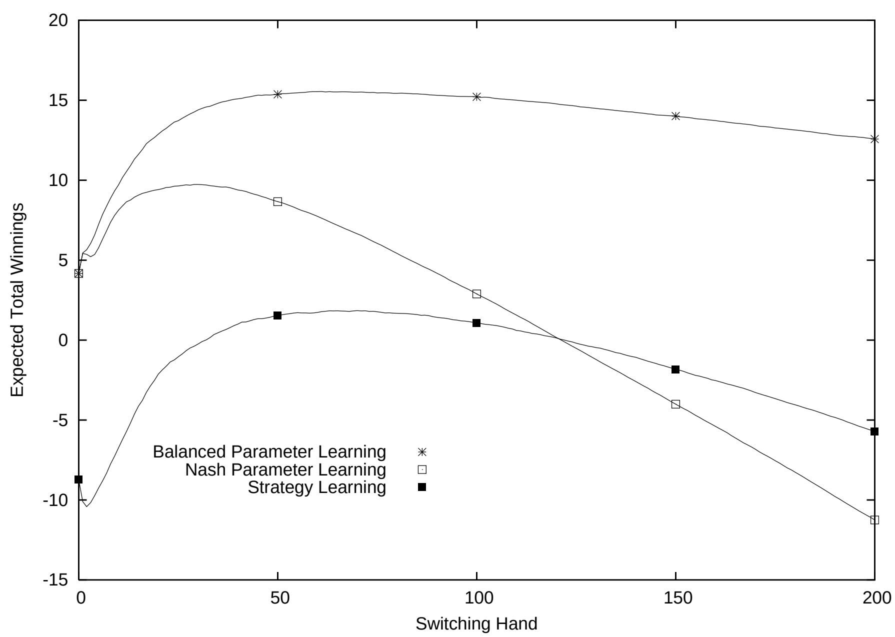

Figure 24: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against O_6.

1.5 Game Length Study

Figures 25 through 30. This collection of plots shows a good deal of variation but in general supports the notion that switching at around 50 hands is a reasonable strategy over the set of game lengths considered.

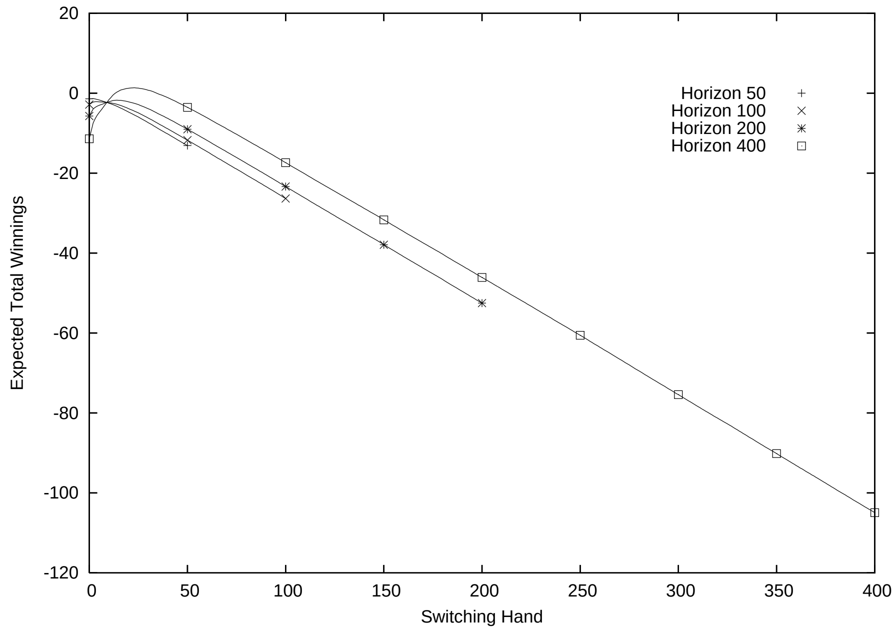

Figure 25: Game Length Study: Expected total winnings vs. switching hand for game lengths of 50, 100, 200, and 400 hands of parameter learning against O_1.

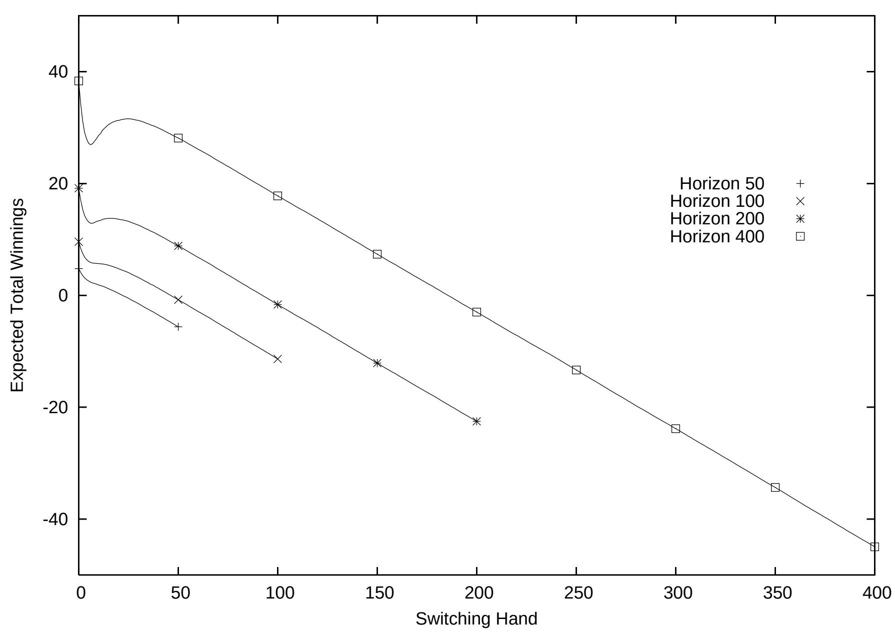

Figure 26: Game Length Study: Expected total winnings vs. switching hand for game lengths of 50, 100, 200, and 400 hands of parameter learning against O_2.

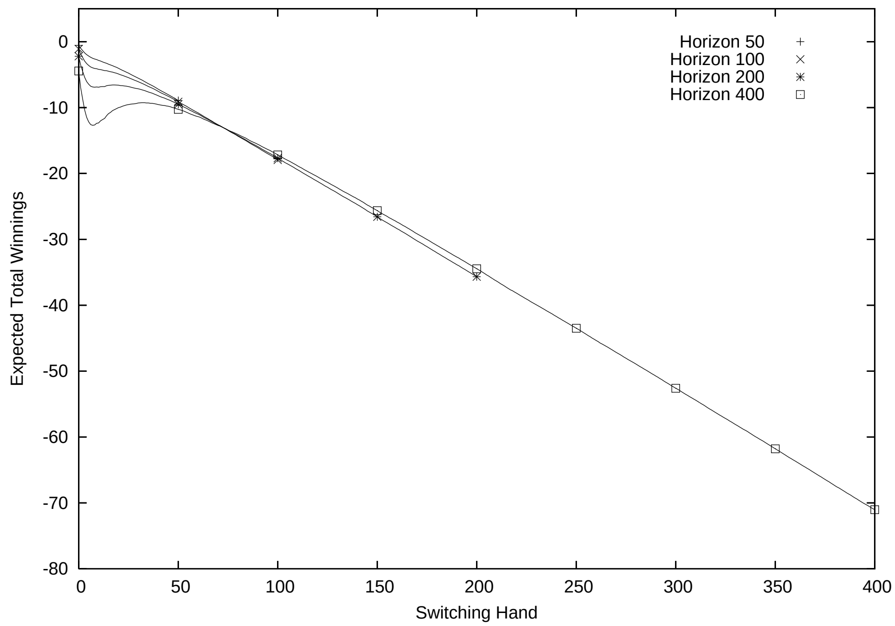

Figure 27: Game Length Study: Expected total winnings vs. switching hand for game lengths of 50, 100, 200, and 400 hands of parameter learning against O_3.

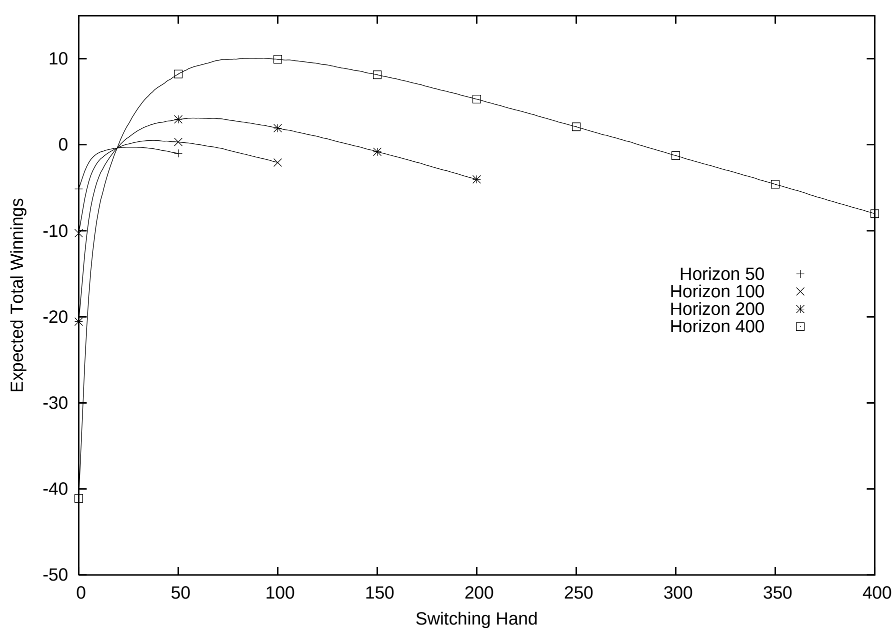

Figure 28: Game Length Study: Expected total winnings vs. switching hand for game lengths of 50, 100, 200, and 400 hands of parameter learning against O_4.

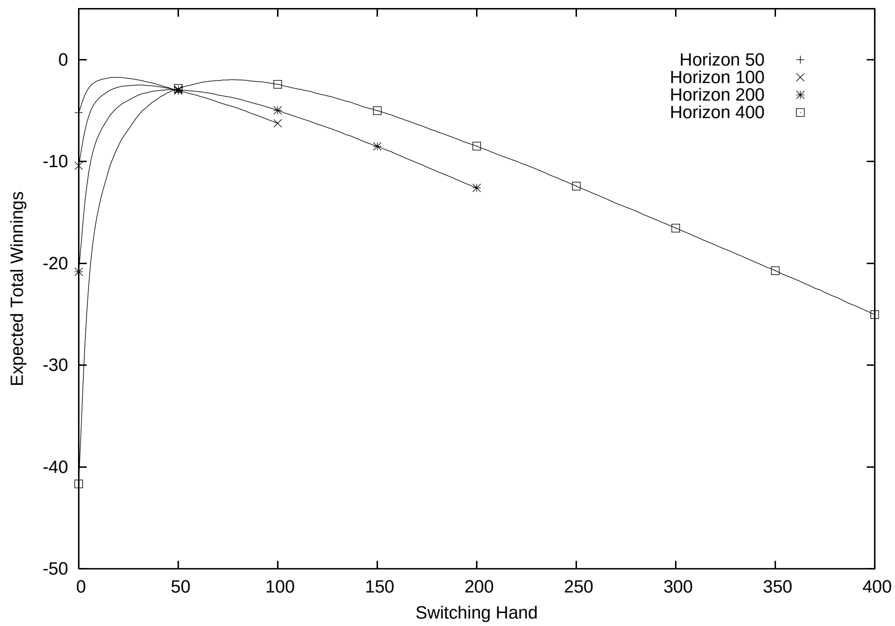

Figure 29: Game Length Study: Expected total winnings vs. switching hand for game lengths of 50, 100, 200, and 400 hands of parameter learning against O_5.

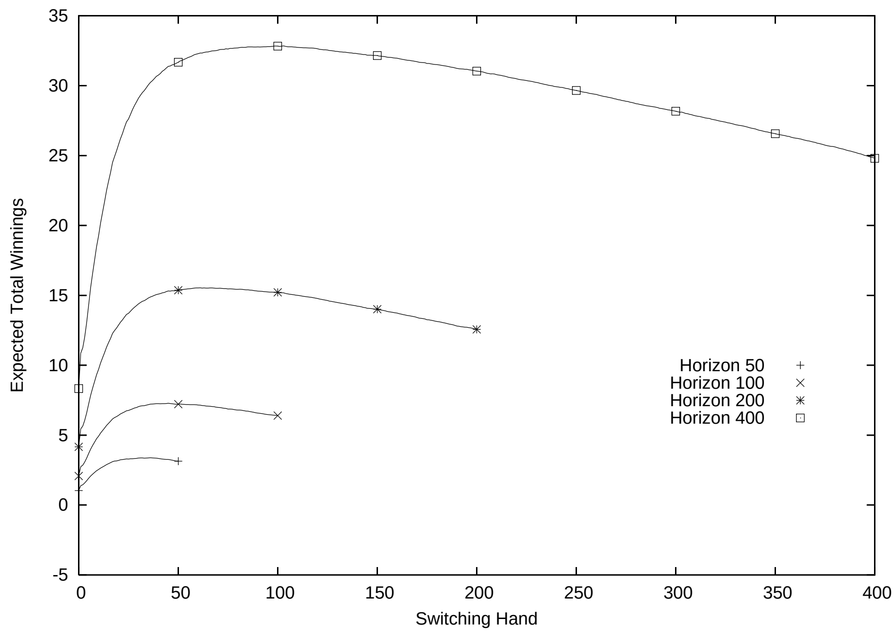

Figure 30: Game Length Study: Expected total winnings vs. switching hand for game lengths of 50, 100, 200, and 400 hands of parameter learning against O_6.

2 Player 2 vs. Player 1 Results

We present additional results for parameter learning against four different P1 opponents ($P_1 = (.2, .5, .9), P_2 = (.12, .65, .6), P_3 = (.25, .35, .3), P_4 = (.35, .65, .7)$).

2.1 Convergence Study

Figures 31 through 34. Here our experiments are limited to parameter learning with two different exploration techniques. Note that the balanced exploration strategy outperforms Nash exploration in all cases.

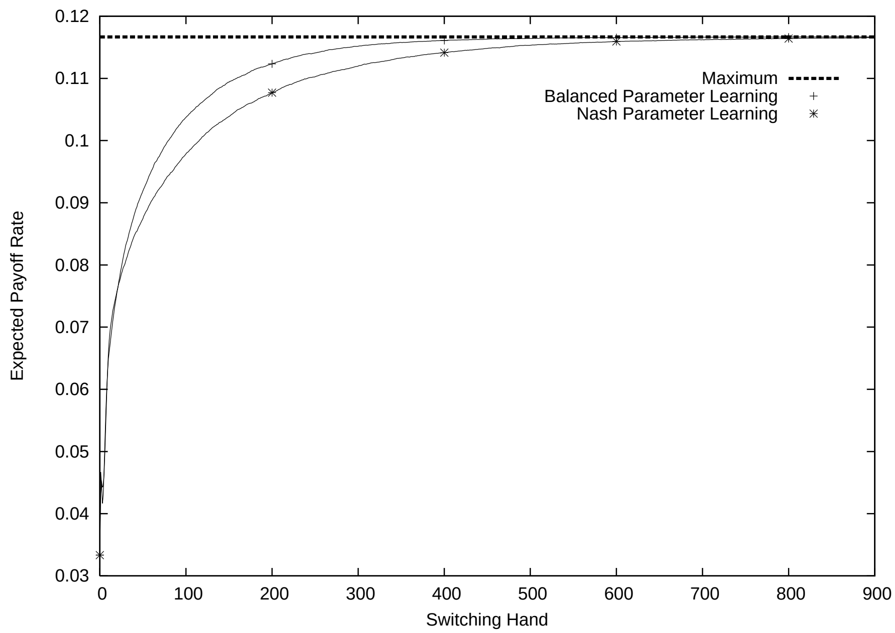

Figure 31: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against P1 Opponent P_1.

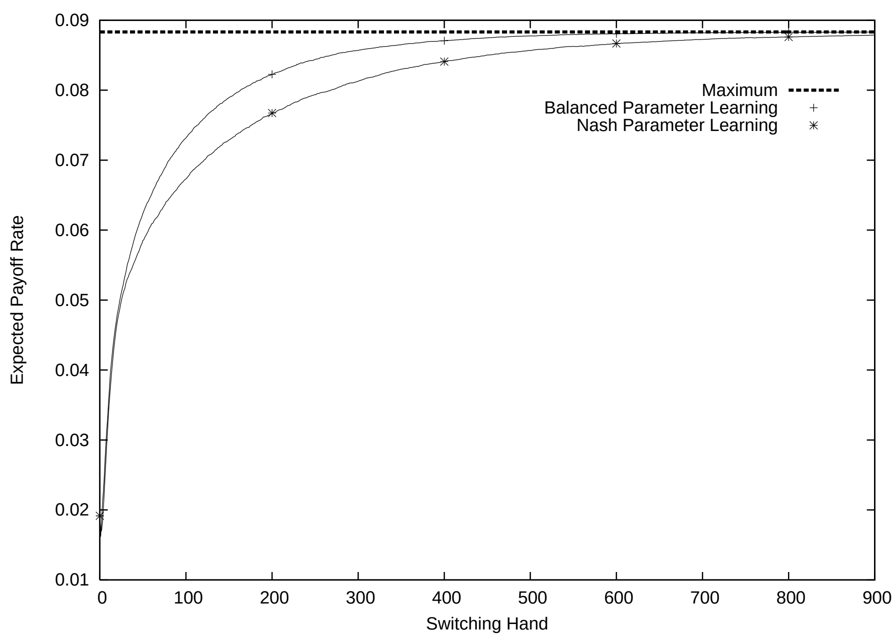

Figure 32: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against P1 Opponent P_2.

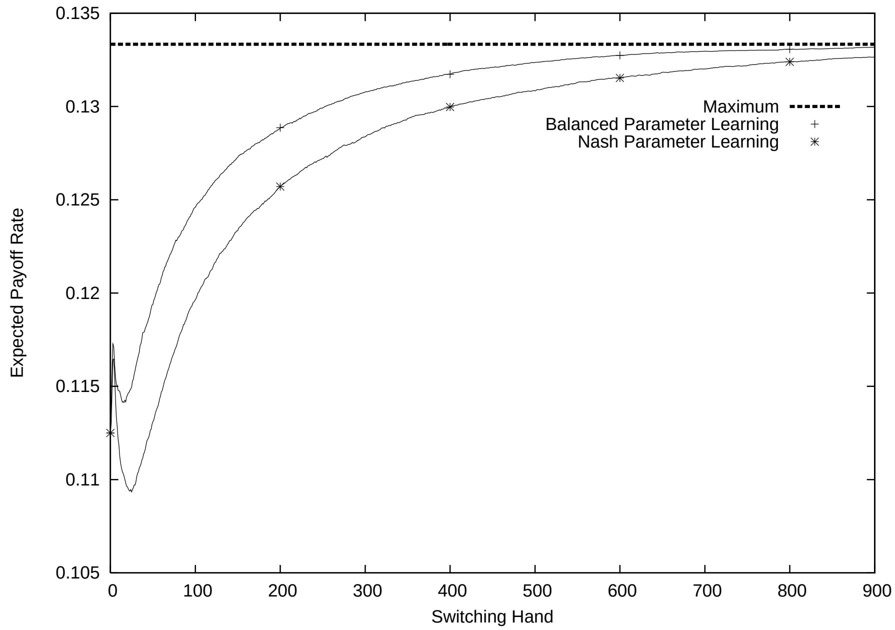

Figure 33: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against P1 Opponent P_3.

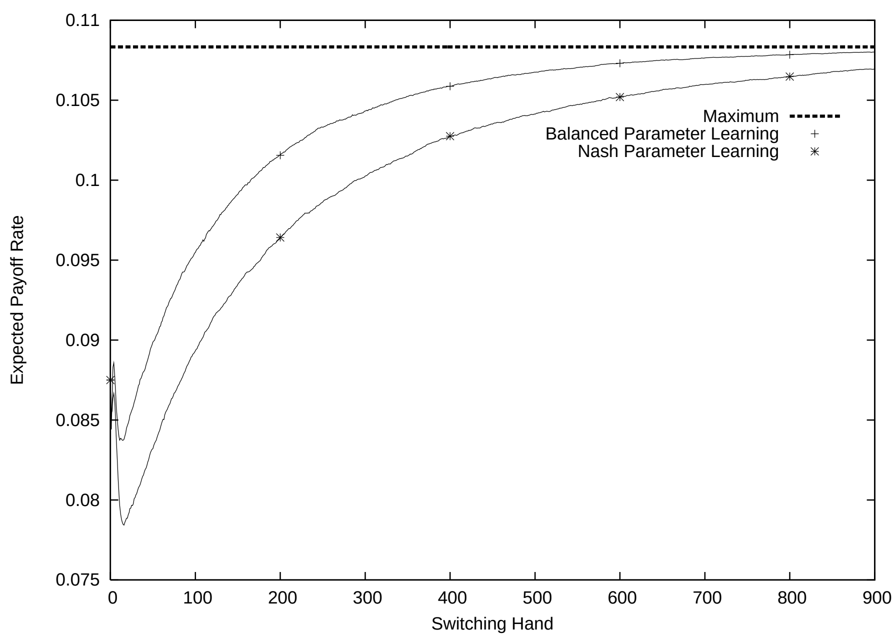

Figure 34: Convergence Rate Study: Expected payoff rate vs. switching hand for parameter and strategy learning against P1 Opponent P_4.

2.2 Learning Method Comparison

Figures 35 through 38. Here again, balanced exploration outperforms Nash exploration. Nash exploration has a particularly egregious loss against opponent P_2 where it performs worse than playing the Nash strategy throughout. In this case, even balanced exploration offers only a small edge over playing Nash, and is still well below the best response value.

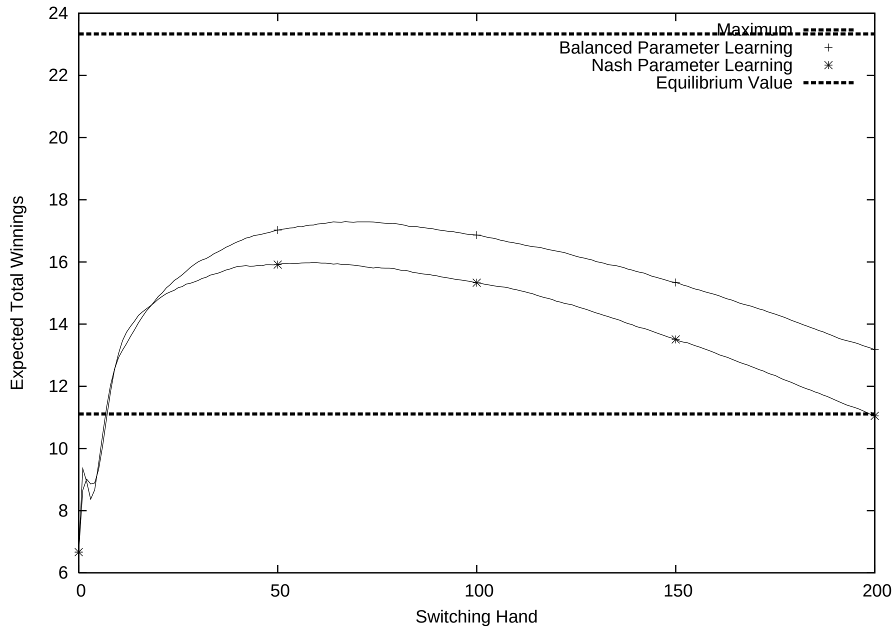

Figure 35: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against P1 Opponent P_1.

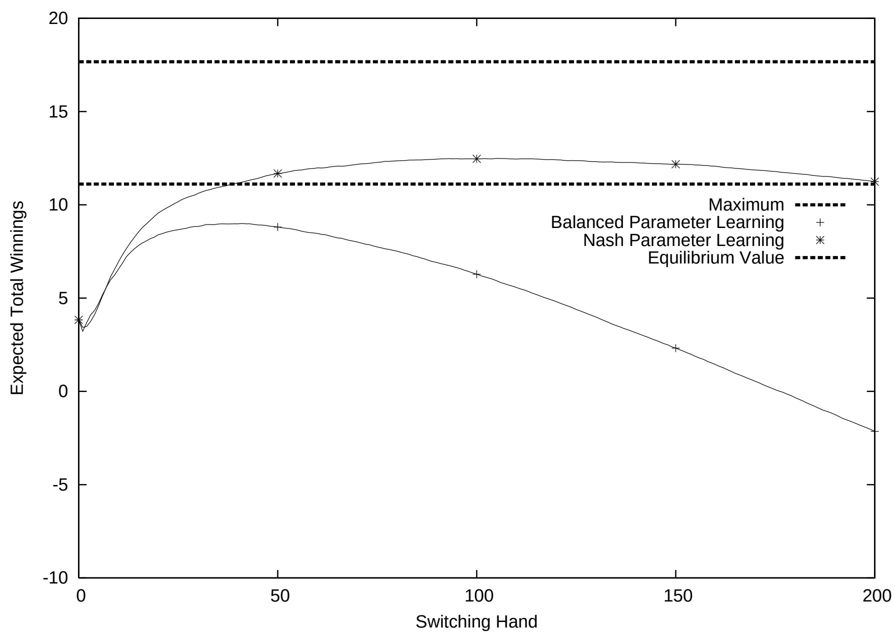

Figure 36: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against P1 Opponent P_2.

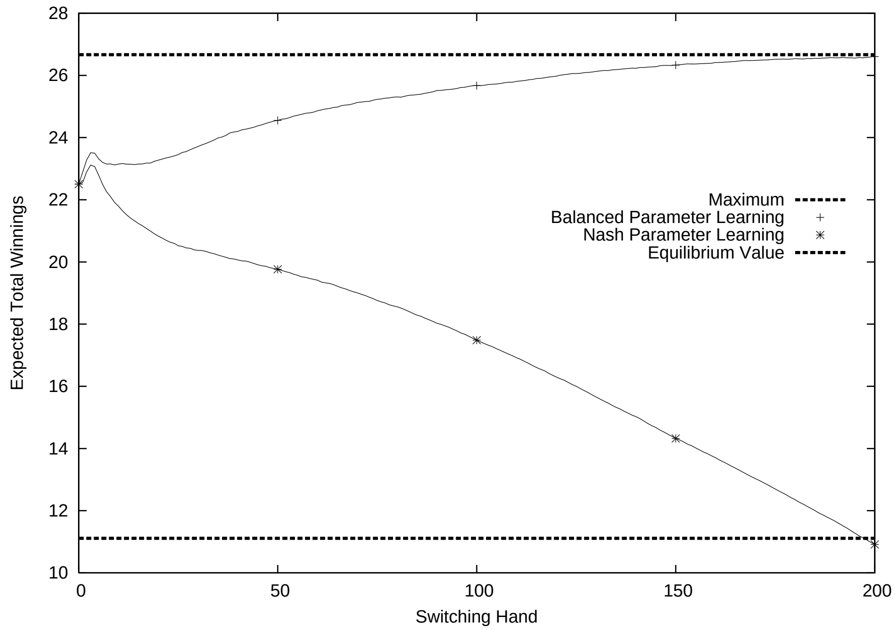

Figure 37: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against P1 Opponent P_3.

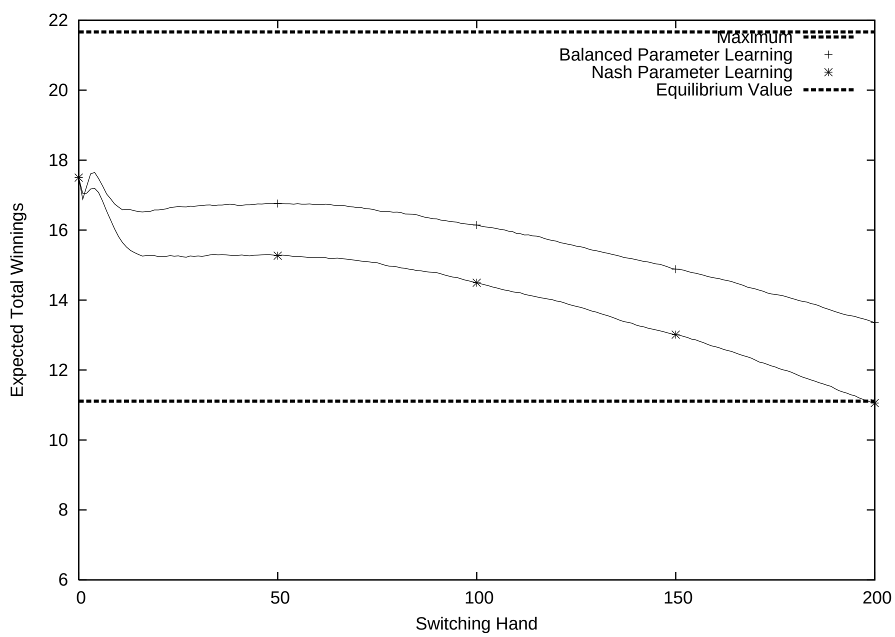

Figure 38: Learning Method Comparison: Expected total winnings vs. switching hand for both parameter learning and strategy learning against P1 Opponent P_4.