

Q-LEARNING AGENTS IN AN EXTENDED DUOPOLY GAME

Svarc Petr, Kovalcin Stanislav, Svarcova Natalie
Institute of Economic Studies, Charles University in Prague

Overview



- Motivation
 - Hamilton and Slutsky 1990
 - Huck, Muller, Normann 1999
 - Our work
- Q-learning agents in an Extended Duopoly Model
 - Note on the Q-learning
 - The model
- Simulation Results
 - No memory
 - Memory
- Conslusions

Motivation

- Starting with papers by Saloner (1987), Hamilton and Slutsky (1990), and Robson (1990), there has been a growing literature studying models of endogenous timing in oligopoly
- These papers analyze extended timing games which establish conditions under which firms are likely to play either a simultaneous-move game or a sequential-move game. The order of output or price decisions is not exogenously specified. Rather, it is derived from firms' decisions about timing
- Results from this literature may indicate whether models of simultaneous output or price decisions (Cournot, Bertrand) or sequential decisions (Stackelberg, price leadership) are preferable

Motivation cont.

- Hamilton and Slutsky 1990 (HS)
 - This game modifies the standard duopoly model by allowing for two production periods before the market clears. Firms can choose their quantities in one of the two periods, $t = 1, 2$
 - A firm can move in period 1 by committing itself to a quantity—without knowing what its competitor is doing. By waiting until period 2, a firm can observe the other firm's period-1 quantity (or its decision to wait).
 - It is assumed that the market for the homogeneous good exists only at period 2 and that production costs do not depend on the production period
 - **HS identify three (subgame-perfect) equilibria in pure strategies: the two Stackelberg equilibria in which one firm commits in period 1 to its Stackelberg leader quantity and the other firm waits and reacts with the Stackelberg follower quantity. The third equilibrium has both firms producing the simultaneous play Cournot equilibrium quantities in period 1**

Motivation cont.

- Huck, Muller, Normann 1999 (HMN)
 - ▣ An experiment designed to test the HS model with action commitment
 - ▣ In particular, they check whether there is experimental evidence for endogenous Stackelberg equilibria or whether some other (if any) equilibrium is selected by subjects
 - ▣ The data of their experimental test show, that endogenous Stackelberg leadership does not occur to the degree theory predicts. The theoretical criterion to prefer pure-strategy equilibria in undominated strategies over other equilibria turns out to be of little behavioral importance. Rather, we see the emergence of Cournot outcomes and, sometimes, collusive outcomes

Motivation cont.



- Our work
 - ▣ In our work we use HS model but we assume that agents do not deduce but rather use learning to find the optimal strategy
 - ▣ We use Q-learning (a reinforcement learning) algorithm as model of agents' learning
 - ▣ The results of the simulations of the computational model are in accordance with the experimental results of HMN –we observe simultaneous moves with Cournot and collusive equilibria more often than the theory predicts

Q-learning agents in an Extended Duopoly Model

- Following Waltman and Kaymak (2008) we apply Q-learning as a learning algorithm for the agents
- Q-learning is applied as follows
 - ▣ An agent plays a repeated game. At the beginning of the stage game in period t , the agent's memory is in some state s_t . This state may be determined by, for example, the actions played by the agent and its opponents in the stage game in period $t-1$
 - ▣ Taking into account the state of its memory, the agent chooses to play some action a_t . The choice of an action is made probabilistically based on the so-called Q-values of the agent
 - ▣ Playing action a_t results in some stage game payoff p_t that is obtained by the agent and in a transition of the state of the agent's memory from the old state s_t to some new state s_{t+1} . The agent uses the experience gained during the stage game to update its Q-values, thereby modifying the way in which it chooses actions in stage games in future periods

Q-learning agents in an Extended Duopoly Model

- Updating Qs (formaly)

$$Q_{t+1}(s, a) = \begin{cases} (1 - \alpha)Q_t(s, a) + \alpha \left(\pi_t + \gamma \max_{a' \in A} Q_t(s_{t+1}, a') \right) & \text{if } s = s_t \text{ and } a = a_t, \\ Q_t(s, a) & \text{otherwise,} \end{cases}$$

Q-learning agents in an Extended Duopoly Model

- Artificial market setting

$$price = \max(u - v(q_1 + q_2), 0)$$

$$profit_i = \max(q_i \cdot (price - cost_i), -q_i \cdot cost_i)$$

$u = 40$ – denotes maximum possible price

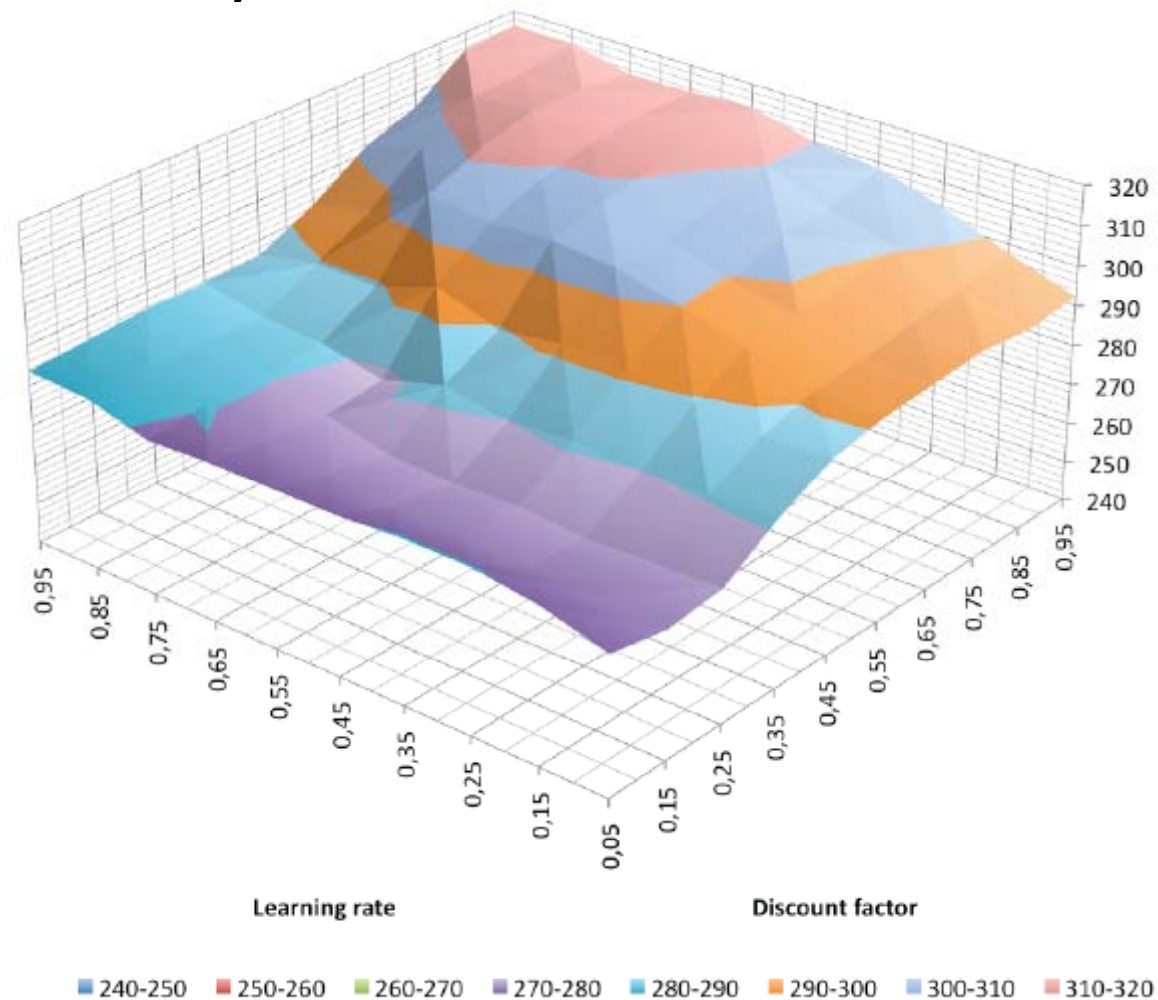
$v = 1$ – denotes slope of inverse demand function

$w = 4$ – denotes firm's marginal cost

- Agents can choose discrete quantities only from the range $[0,40]$
- Profit in Cournot is 288, profit in Stackelberg is 243

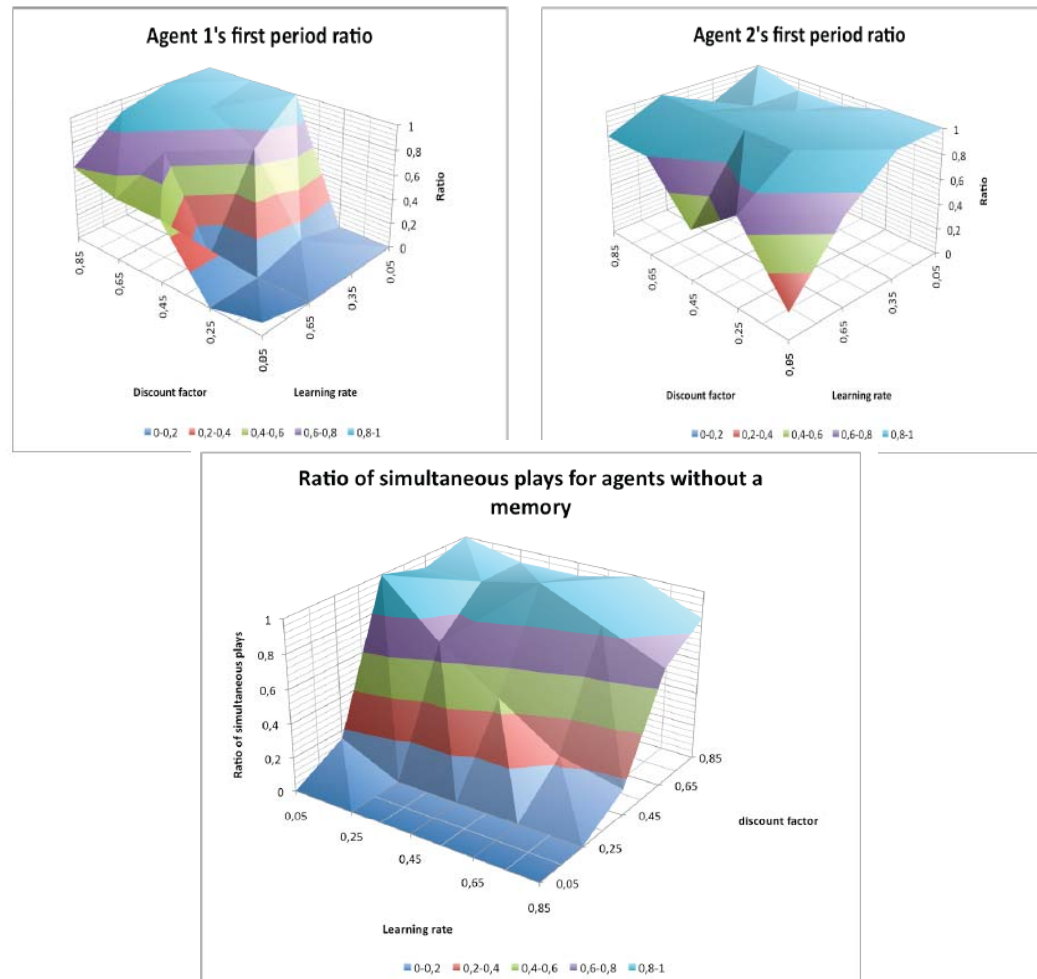
Simulation Results

□ No memory



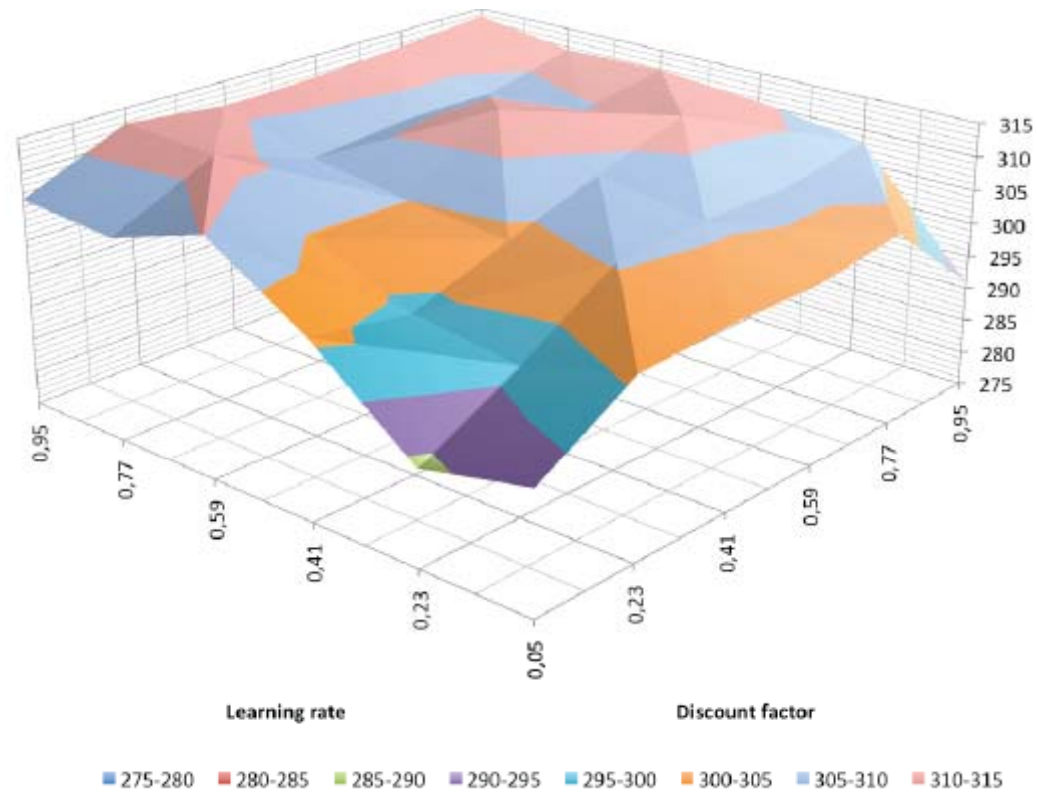
Simulation results cont.

□ No memory



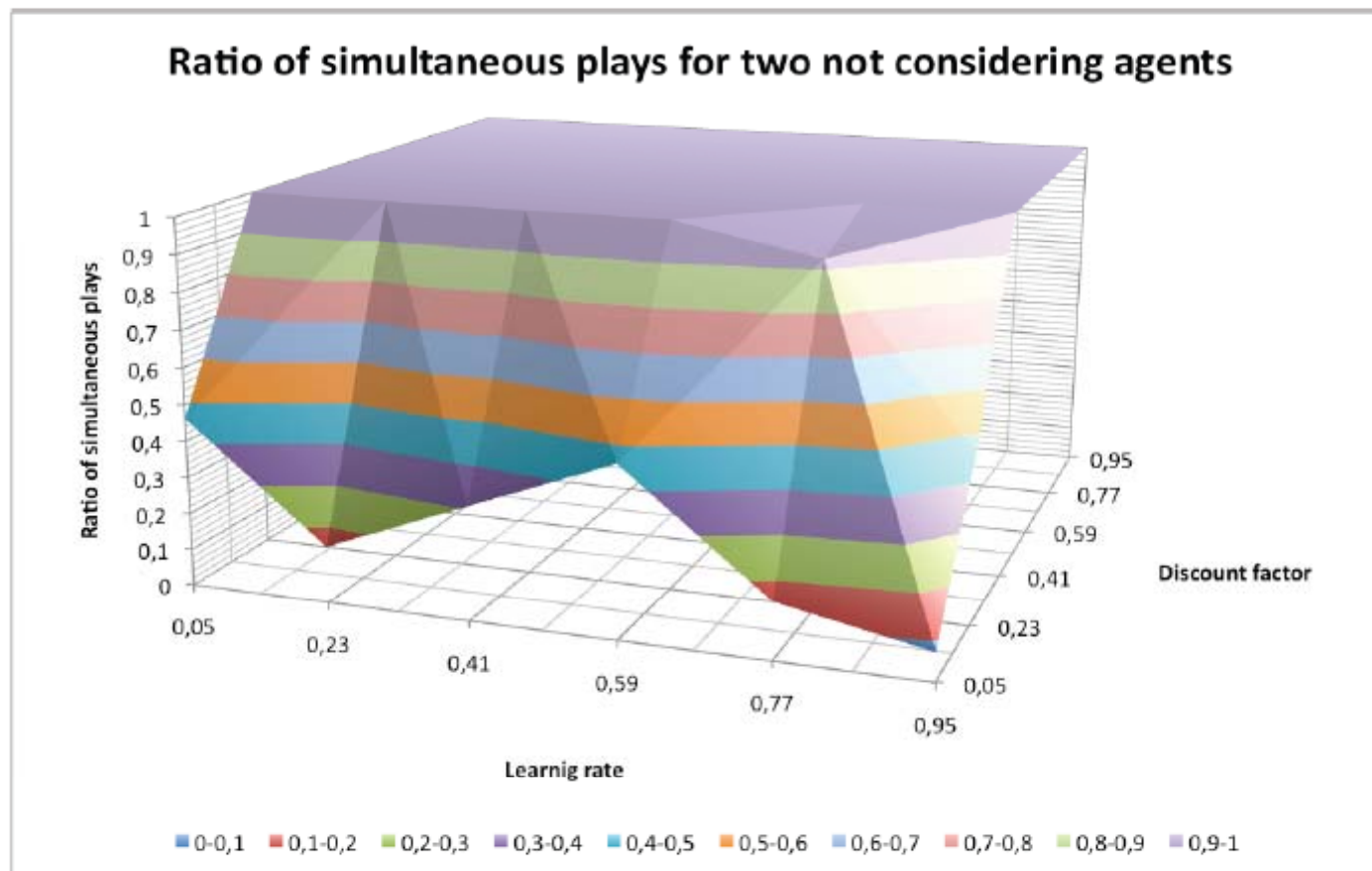
Simulation Results

- Memory (only considering current game)



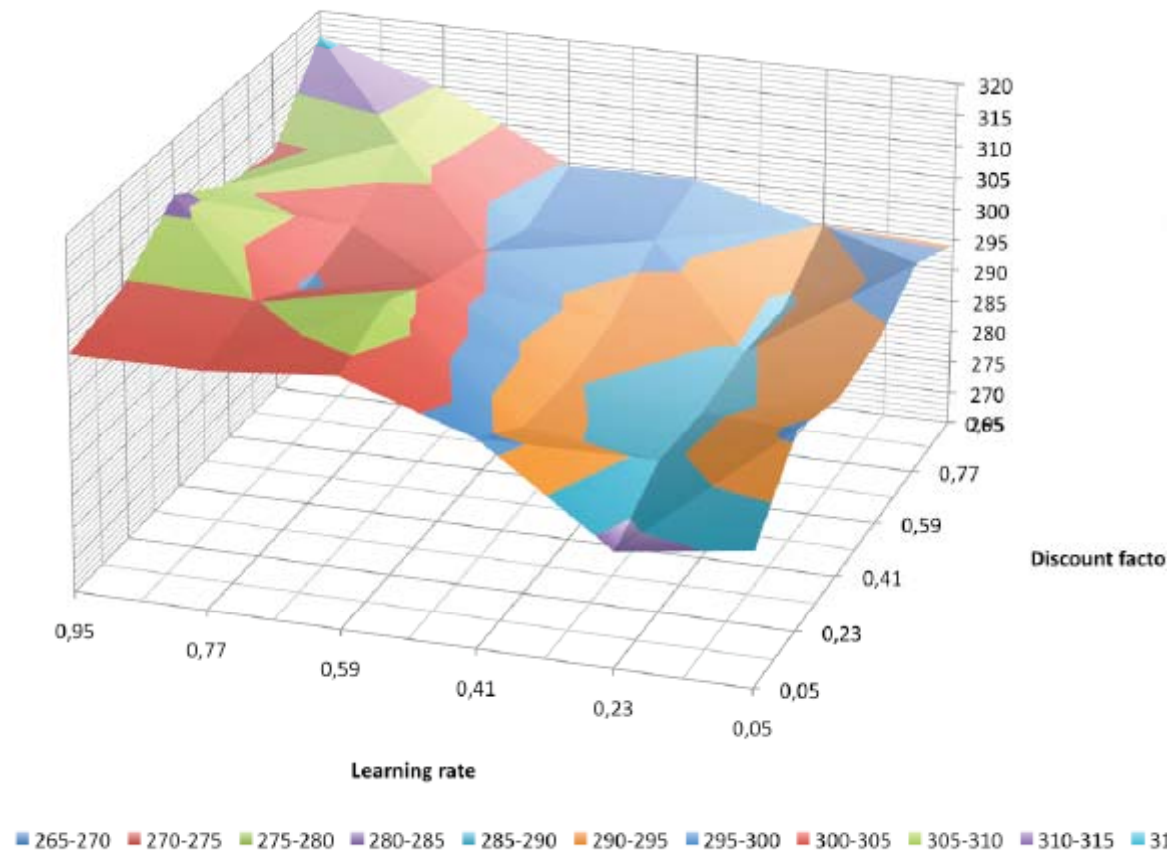
Simulation Results

- Memory (only considering current game)



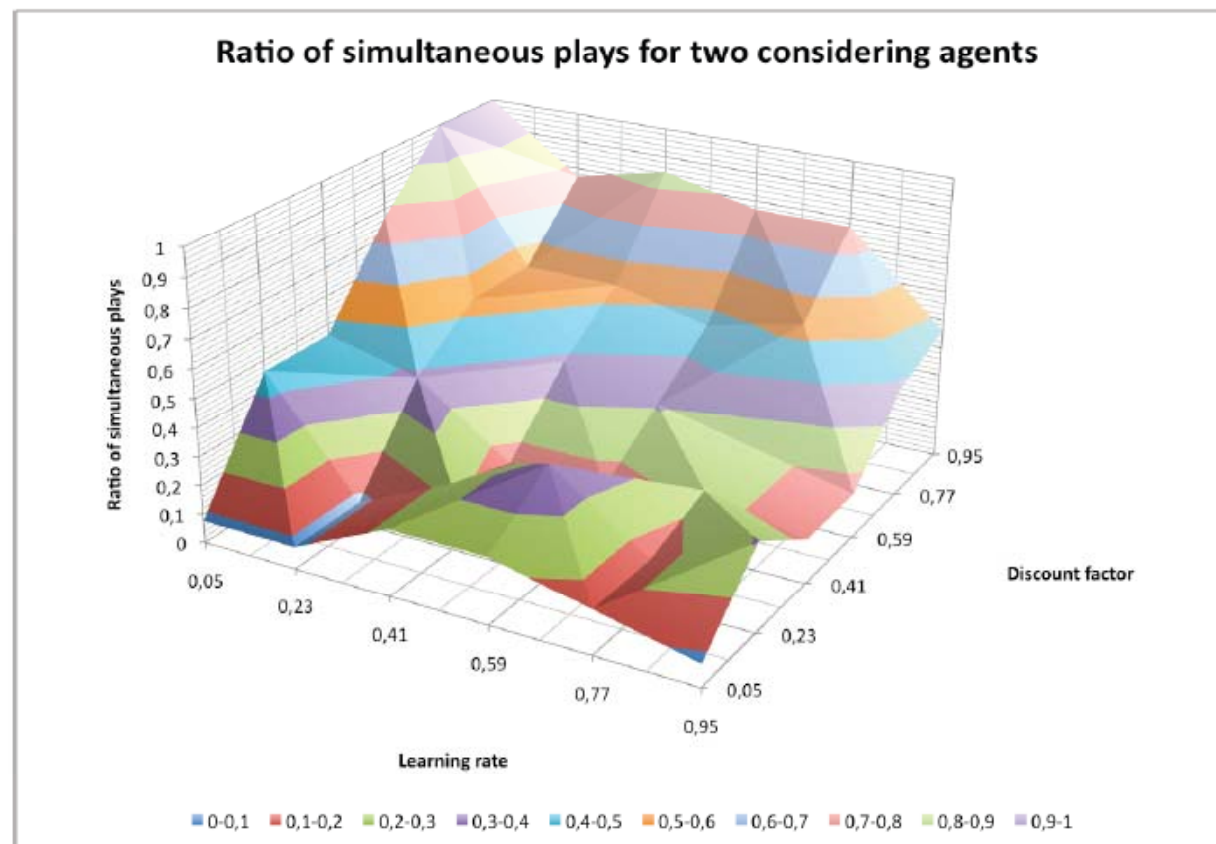
Simulation Results

- Memory (considering next game)



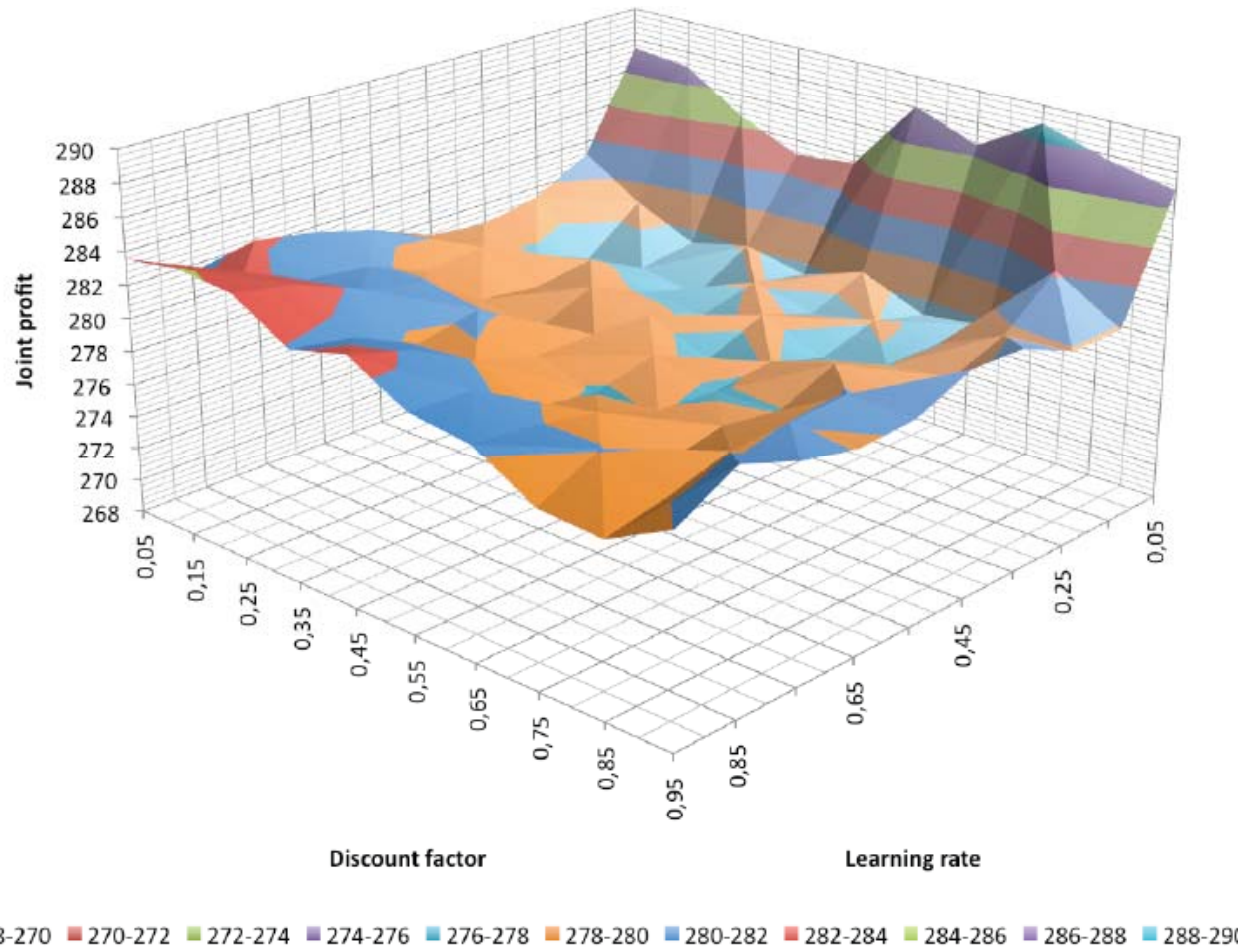
Simulation Results

- Memory (considering next game)



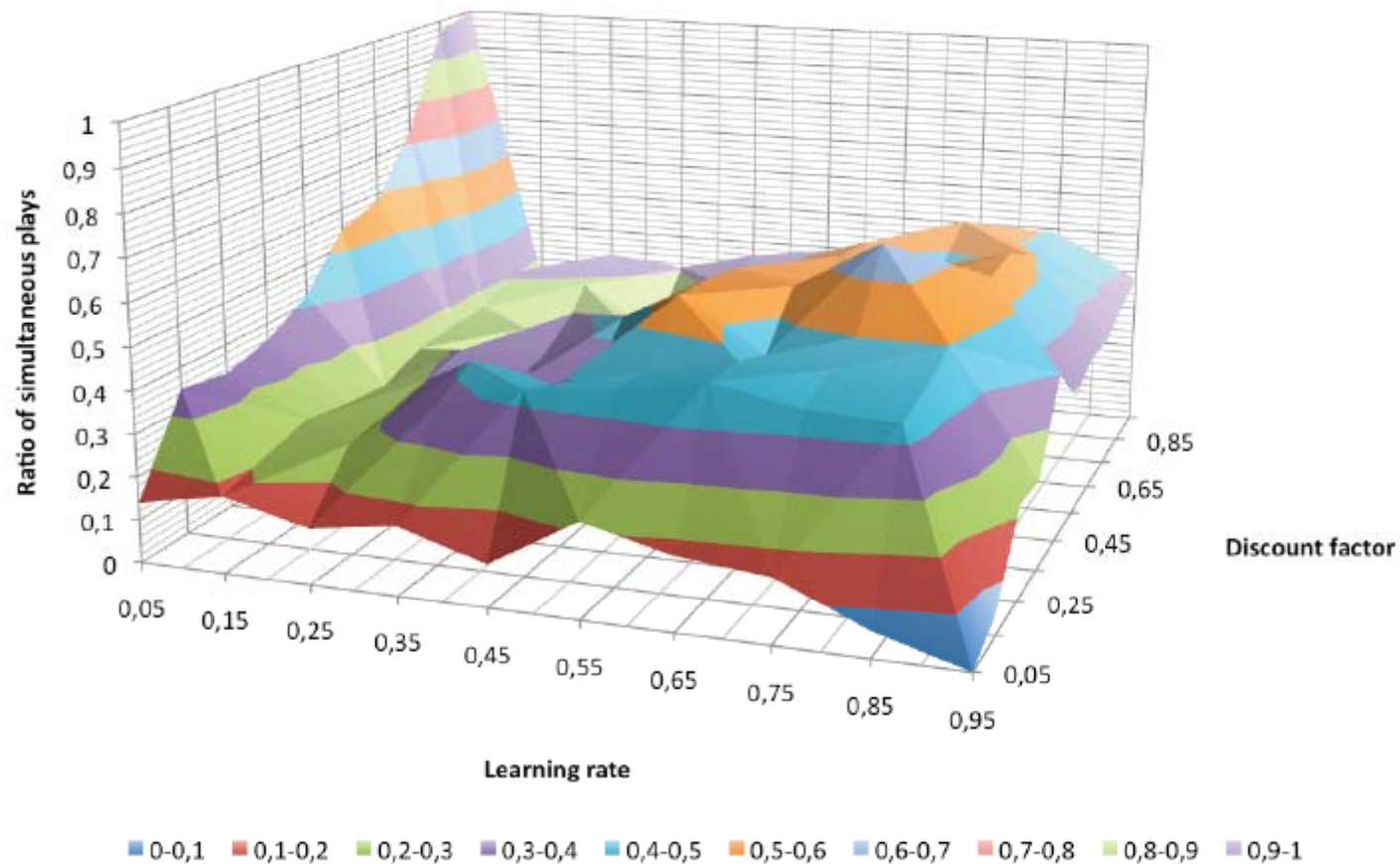
Simulation Results

- Memory (one considering the other not)



Simulation Results

- Memory (one considering the other not)



Conclusions

- Results of the simulations are in accordance with the experimental findings – we observe simultaneous moves and collusive behavior in many cases
- The Q-learning seems to be a good approximation of learning of real agents
- We believe that the difference between the theory and experiments can be explained by the way how the optimal moves are chosen – game theorists vs. normal people 😊