Equilibria Computation in Zero-Sum Games

Intelligent Agents and Multiagent Systems | MSc Artificial Intelligence

Tatiana Boura, Andreas Sideras

NCSR Demokritos & University of Piraeus

February 2023

Outline

1. Zero Sum Games
2. Strategies & Nash equilibria
3. Repeated & Stochastic Games
4. Teaching & Learning
5. Experiments
   - Custom game 1
   - Custom game 2
   - Selling damaged goods
6. Conclusion

Table of Contents

1. Zero Sum Games
2. Strategies & Nash equilibria
3. Repeated & Stochastic Games
4. Teaching & Learning
5. Experiments
   - Custom game 1
   - Custom game 2
   - Selling damaged goods
6. Conclusion

Zero Sum Games

- In Game Theory, a Zero Sum Game is a class of games that represent situations of pure competition.
- Zero Sum Games can be formulated in **Normal form**.

Seller \ Buyer	Buy	Pass
Sell	(1,-1)	(-1,1)
Keep	(-1,1)	(-1,1)

Figure: Payoff matrix of Selling damaged goods game.

Table of Contents

1 Zero Sum Games

2 Strategies & Nash equilibria

3 Repeated & Stochastic Games

4 Teaching & Learning

5 Experiments
- Custom game 1
- Custom game 2
- Selling damaged goods

6 Conclusion

Strategies & Nash equilibria

- In order to win, a player must develop a **strategy**. There are two types of strategies, *pure* and *mixed*.
- An agent could play an utility-maximizing strategy, called **best response**.
- If everyone is best responding to each other and no one has the incentive to change his strategy, then the set of the agents' strategies is called **Nash Equilibrium**.

Table of Contents

1 Zero Sum Games

2 Strategies & Nash equilibria

3 Repeated & Stochastic Games

4 Teaching & Learning

5 Experiments
- Custom game 1
- Custom game 2
- Selling damaged goods

6 Conclusion

Repeated & Stochastic Games

- A **Repeated Game** is a game played multiple times by the same set of players.
- **Stochastic games** are a generalization of both *MDPs* and repeated games.

Table of Contents

1. Zero Sum Games
2. Strategies & Nash equilibria
3. Repeated & Stochastic Games
4. Teaching & Learning
5. Experiments
   - Custom game 1
   - Custom game 2
   - Selling damaged goods
6. Conclusion

Fictitious Play

- **Fictitious Play** (FP) is a learning rule where players form and update beliefs about the opponent's strategy (*model based learning*).
- FP can be used in computing Nash Equilibria.

Algorithm 1 FP algorithm

1: Initialize beliefs about the opponent's strategy
2: **while** played less than n times **do**
3: Play a best response to the assessed strategy of the opponent
4: Observe the opponent's actual play and update beliefs accordingly
5: **end while**

Reinforcement Learning

- **Reinforcement Learning** (RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback (reward) from its own actions and experiences.
- In a multi-agent environment the agent is not looking for an optimal strategy, but rather a strategy that performs well against their opponents.
- In two player Zero Sum Games, the strategy profile where each agent plays his **maxmin** strategy is a Nash equilibrium.

minmax Q learning

Algorithm 2 minimax Q-learning algorithm

Require: the set of states S , the set of the agent's actions A, the set of the oponent's actions O, the learning rate α, the explor probability ϵ and the discounting factor γ

for $\forall$ s $\in$ S, $\forall$ a $\in$ A and $\forall$ o $\in$ O **do** ▷ Initialize

$Q(s, a, o) \leftarrow 1$

end for

for $\forall$ s $\in$ S **do**

$V(s) \leftarrow 1$

end for

for $\forall$ s $\in$ S and a $\in$ A **do**

$\Pi(s, a) \leftarrow \frac{1}{|A|}$

end for

$\alpha \leftarrow 1.0$

When in state s, with probability ϵ choose an action uniformly at random and with probability $1 - \epsilon$ choose action a with probability $\Pi(s, a)$ ▷ Take an action

after receiving reward r for moving from state s to s' via action a and opponent's action o : ▷ Learn

$Q(s, a, o) \leftarrow (1 - \alpha) * Q(s, a, o) + \alpha * (r + \gamma * V(s'))$

$\Pi(s, \cdot) \leftarrow argmax_{\Pi'(s,\cdot)}(min_{o'} \sum_{a'}(\Pi(s, a') * Q(s, a', o')))$

$V(s) \leftarrow min_{o'} \sum_{a'}(\Pi(s, a') * Q(s, a', o'))$

Update α

Table of Contents

1 Zero Sum Games

2 Strategies & Nash equilibria

3 Repeated & Stochastic Games

4 Teaching & Learning

5 Experiments
- Custom game 1
- Custom game 2
- Selling damaged goods

6 Conclusion

Custom game 1

	C	D
A	(3,-3)	(1,-1)
B	(2,-2)	(4,-4)

There is a Nash equilibrium in *mixed strategies*, but not in *pure*. Let the column player choose action C with probability p and action D with probability $1 - p$. Since the row player must be indifferent between his actions (A, B):

$$u_1(A) = u_1(B)$$
$$3 \cdot p + 1 \cdot (1 - p) = 2 \cdot p + 4 \cdot (1 - p)$$
$$p = \frac{3}{4}$$

and so, the mixed strategy of the column player is $(0.75, 0.25)$. Following the same procedure, the mixed strategy of row player is $(0.5, 0.5)$. The expected payoff of row player is equal to 2.5, whereas the expected payoff of column player is equal to -2.5 (equal to the *Value* of the game in the RL case).

Custom game 1 - Convergence

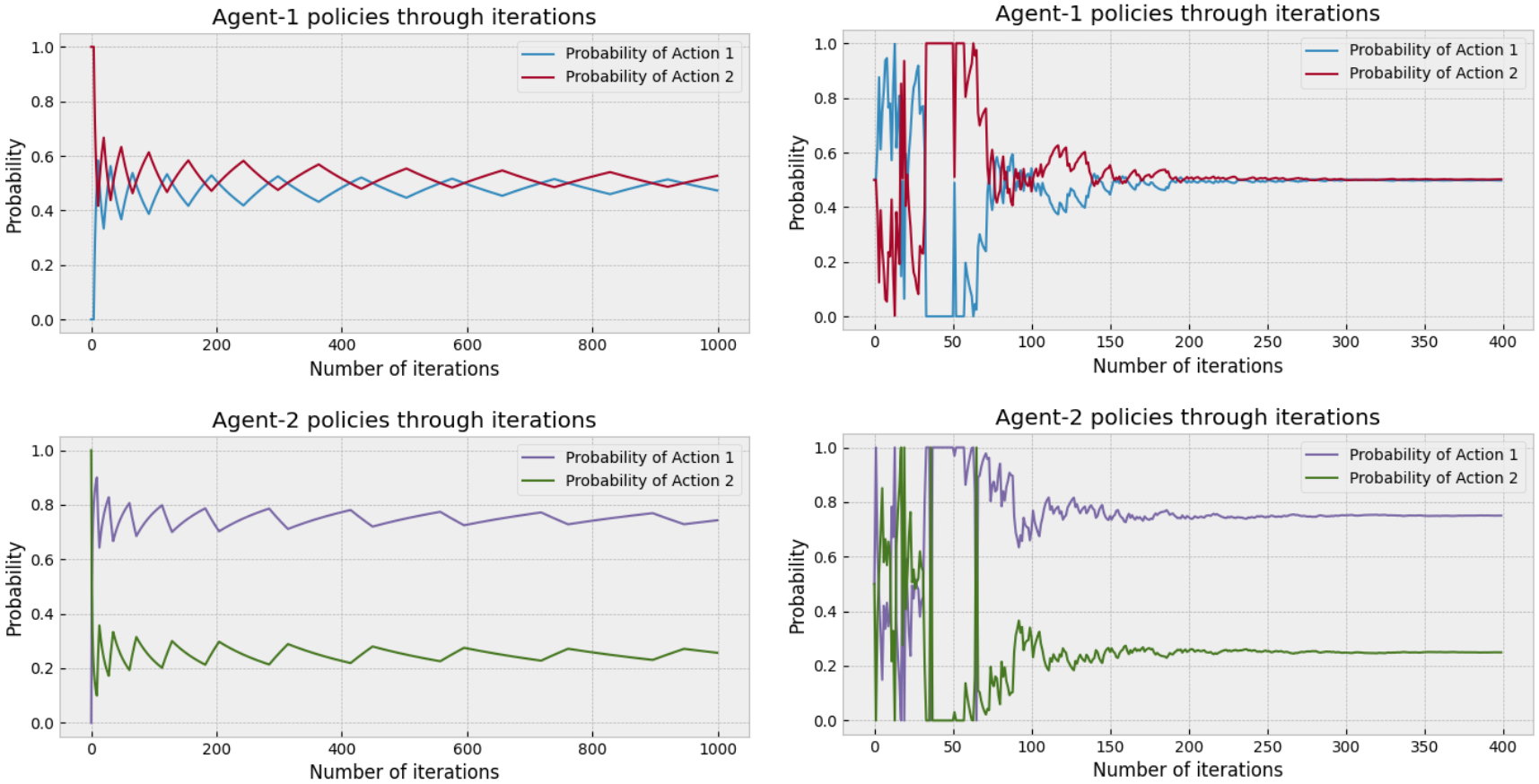

Figure: FP (Left) and RL (Right)

Custom game 2

	C	D
A	(2,-2)	(2,-2)
B	(1,-1)	(3,-3)

There are both pure Nash and mixed strategy equilibria. The pure strategy equilibrium is the action profile (A, C). Row player plays the pure strategy of action A, and the column player plays the mixed strategy of (0.5, 0.5). The expected payoff of row player is equal to 2, whereas the expected payoff of column player is equal to -2.

Custom game 2 - Convergence

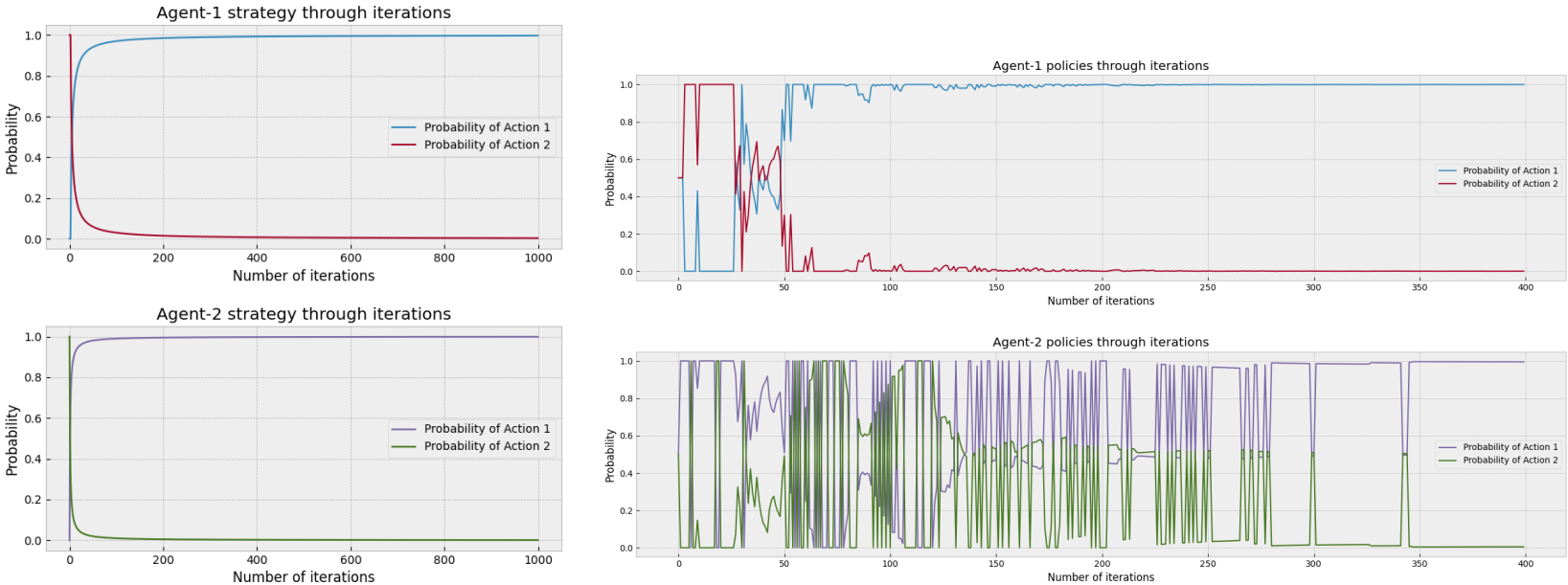

Figure: FP (Left) and RL (Right)

Selling damaged goods

	Buy	Pass
Sell	(1,-1)	(-1,1)
Keep	(-1,1)	(-1,1)

In this game exist two Nash equilibria in pure strategies. These equilibria are (*Sell*, *Pass*) and (*Keep*, *Pass*). There, also, exists an equilibrium in the pure strategy of *Pass* action of the column player and the mixed strategy (0.5, 0.5) of the row player. In each equilibrium, the row player's expected payoff is -1 and column player's expected payoff is 1.

Selling damaged goods - FP Convergence

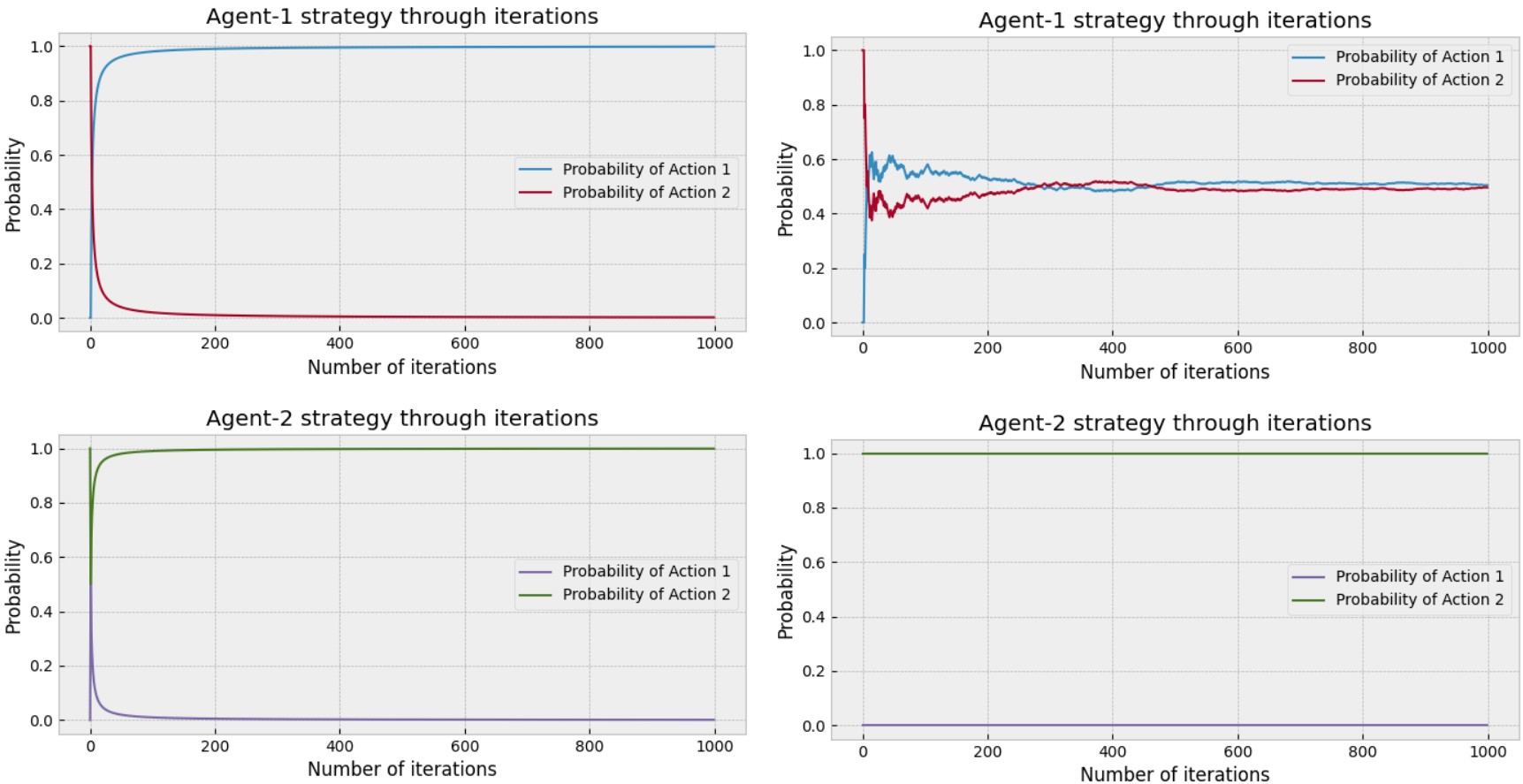

Figure: FP Convergence with same beliefs

Selling damaged goods - RL Convergence

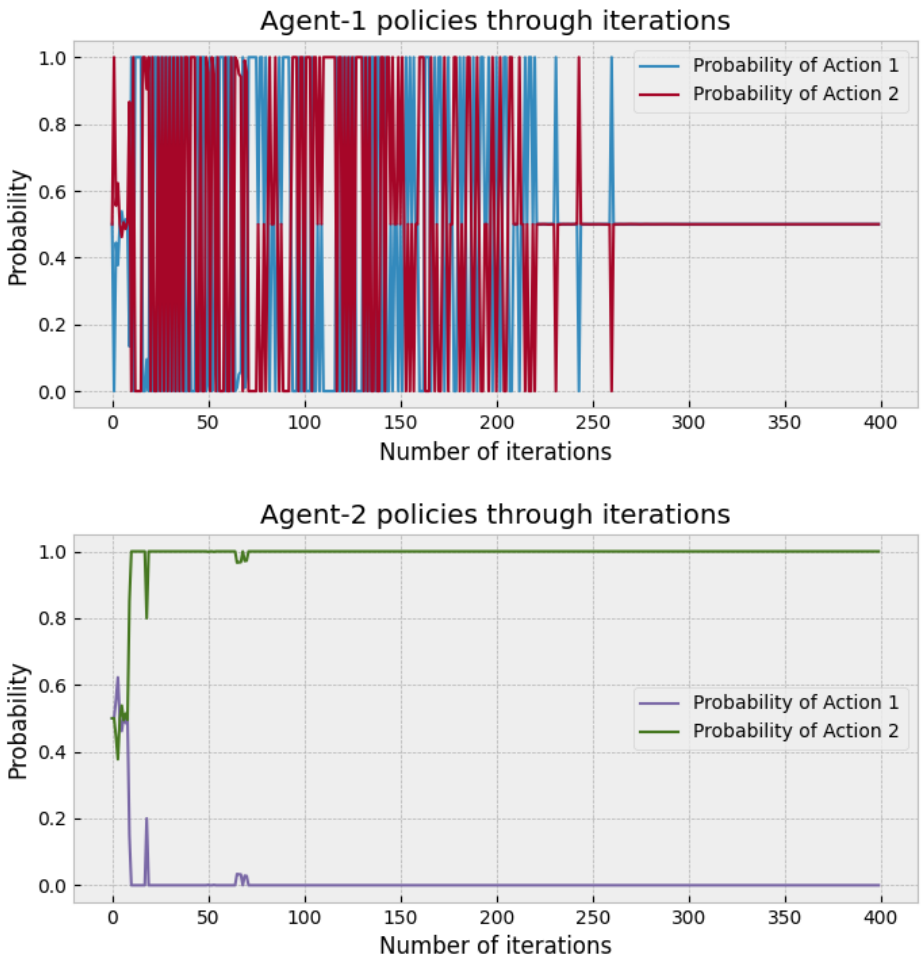

Figure: RL Convergence

Table of Contents

1. Zero Sum Games
2. Strategies & Nash equilibria
3. Repeated & Stochastic Games
4. Teaching & Learning
5. Experiments
   - Custom game 1
   - Custom game 2
   - Selling damaged goods
6. Conclusion

Convergence

- In the case of multiple equilibria, FP converged to any of them.
  - e.g. in the *selling damaged goods* game, after running the FP 100.000 times, 33% of them converged to the mixed strategy equilibrium, 66% to one of the pure equilibria and 1% to the other.
- RL in our experiments converged always to the same equilibria.
- RL converges much faster than FP to the equilibria.
- FP converges more smoothly than RL, which oscillates much more.
- FP is highly dependent on the initialization of its prior beliefs.
- RL is much more robust to its initial parameters.

FP Parameters

- Initial beliefs affect the convergence rate.

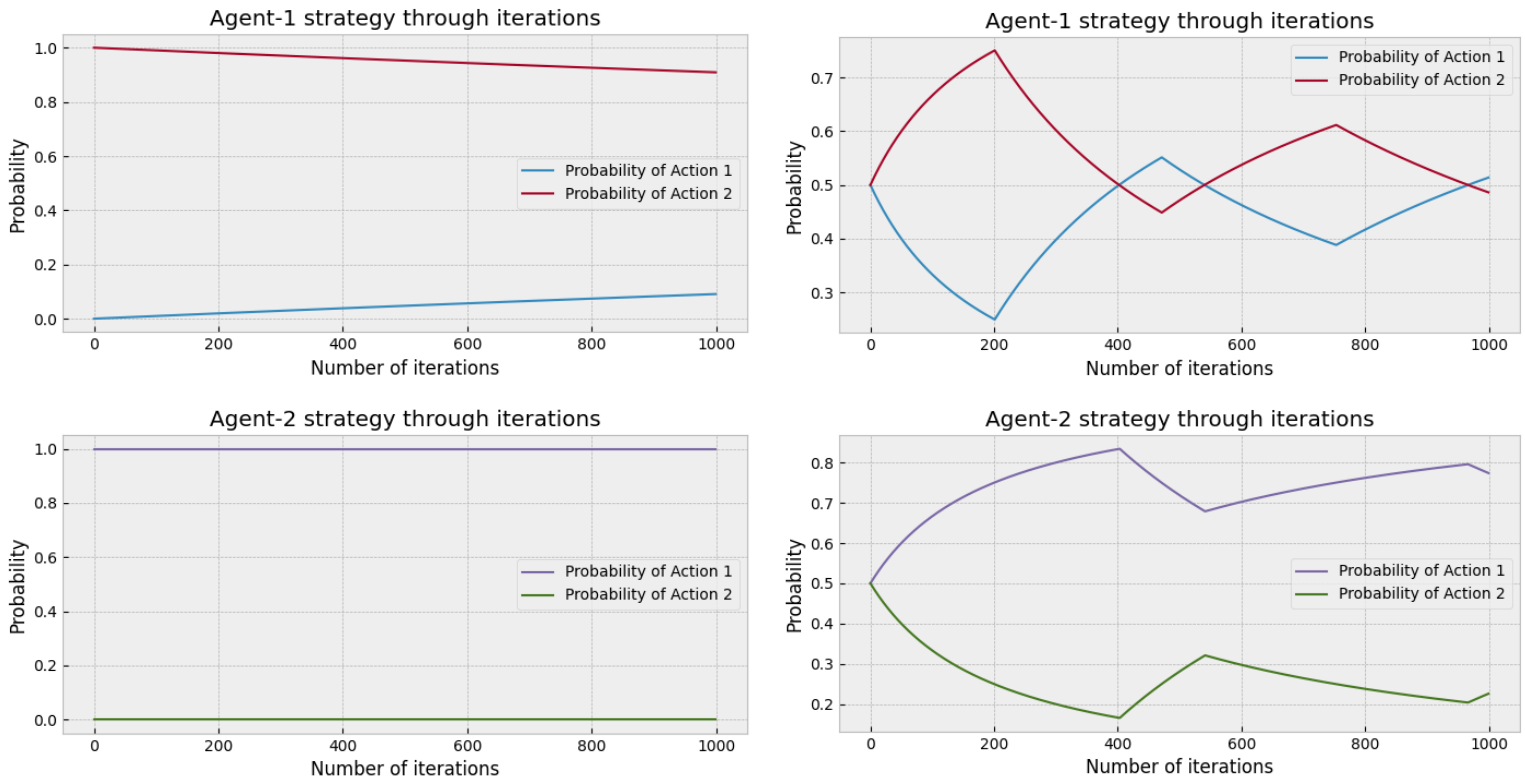

Figure: Custom Game 1 : Initial beliefs (10000,0), (0,10000) [Left] and (100,100), (100,100) [Right]

FP Parameters

- Initial beliefs affect the equilibria found.

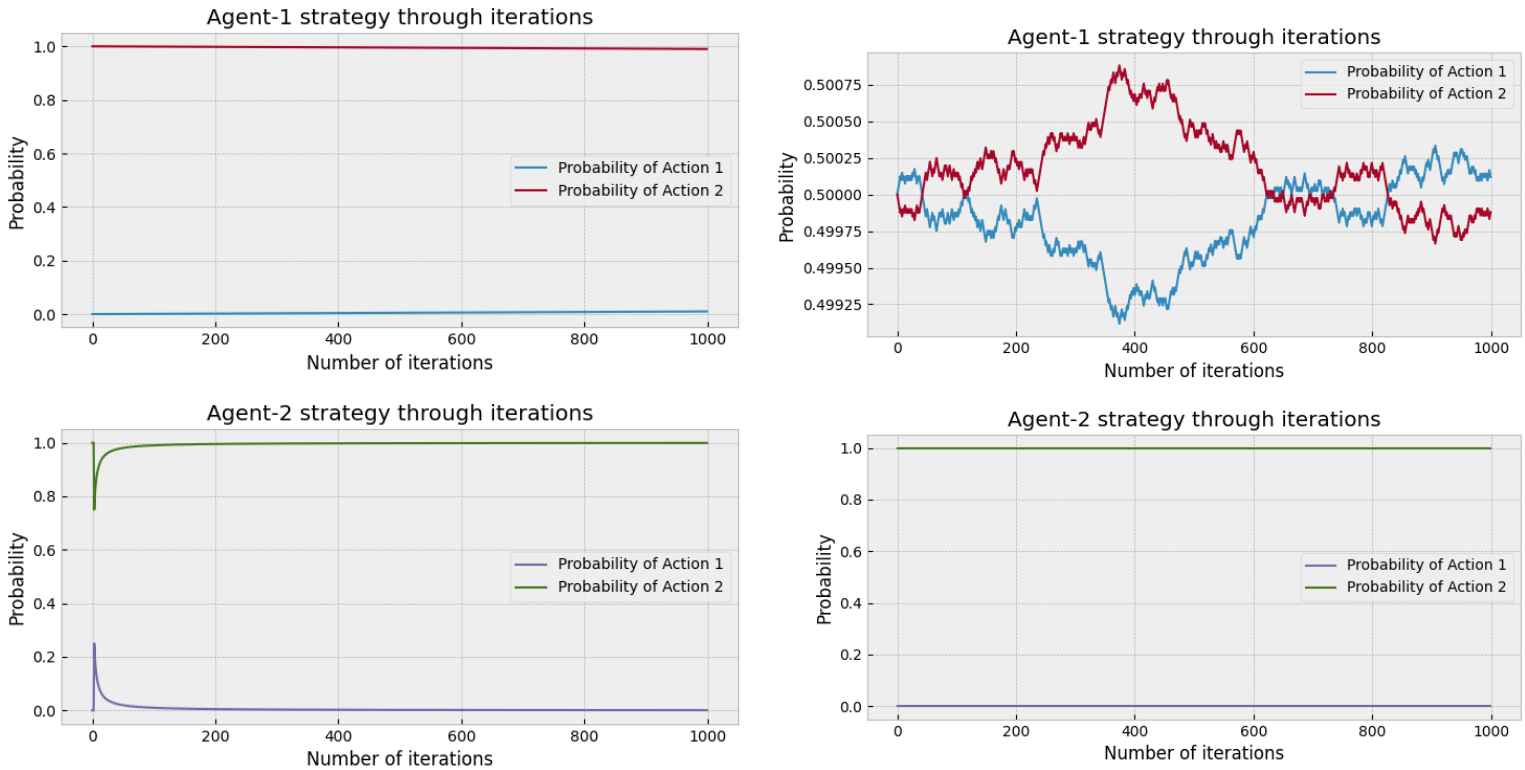

Figure: Selling damaged goods : Initial beliefs (0,1), (0,100000) [Left] and (0,1), (1000,1000) [Right]

Thank you for your attention and interest.