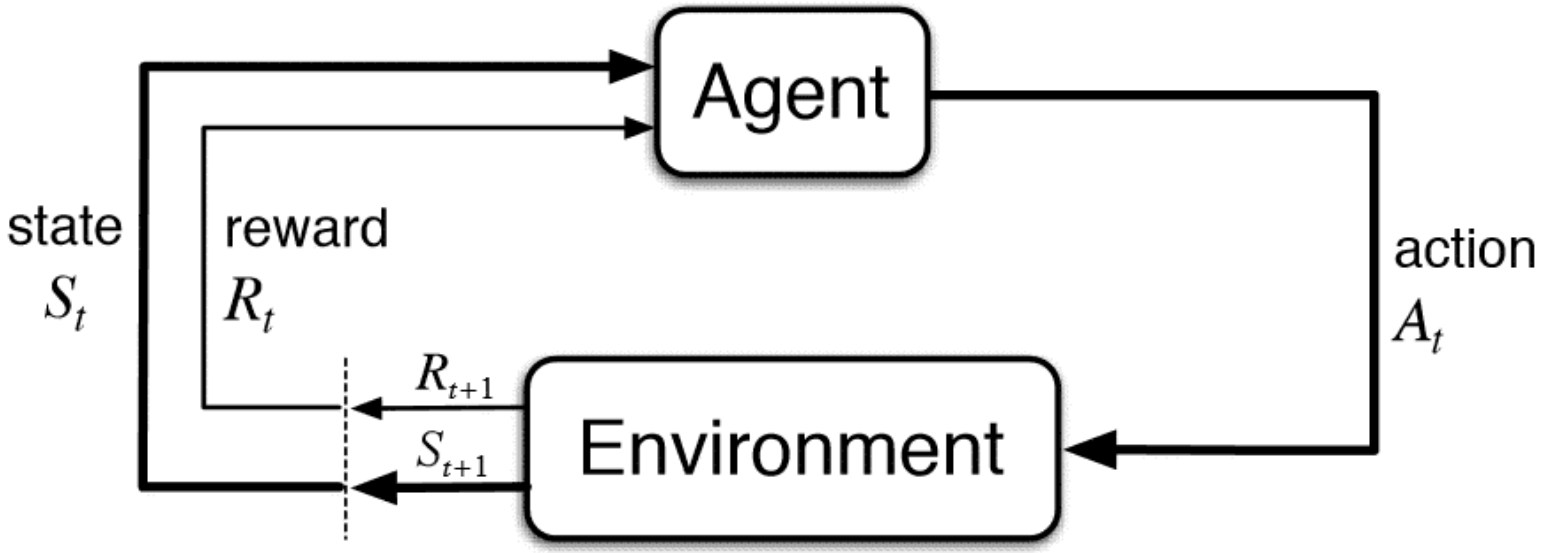
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Action Trading for Self-Interested Multi-Agent Reinforcement Learning in a Smart Factory Setting

Arnold Unterauer

Reinforcement Learning



Source: Richard S. Sutton and Andrew G. Barto **Reinforcement Learning: An Introduction**

Multi Agent Systems:

agents act selfish

no cooperation between agents

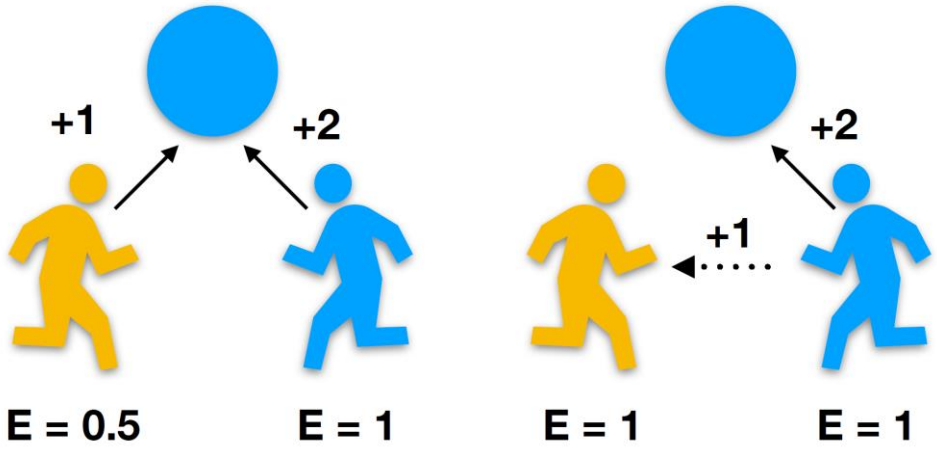
unused potential

Solution:

Cooperative Game Theory

motivate agents to cooperate with each other

Related Work – Action Trading



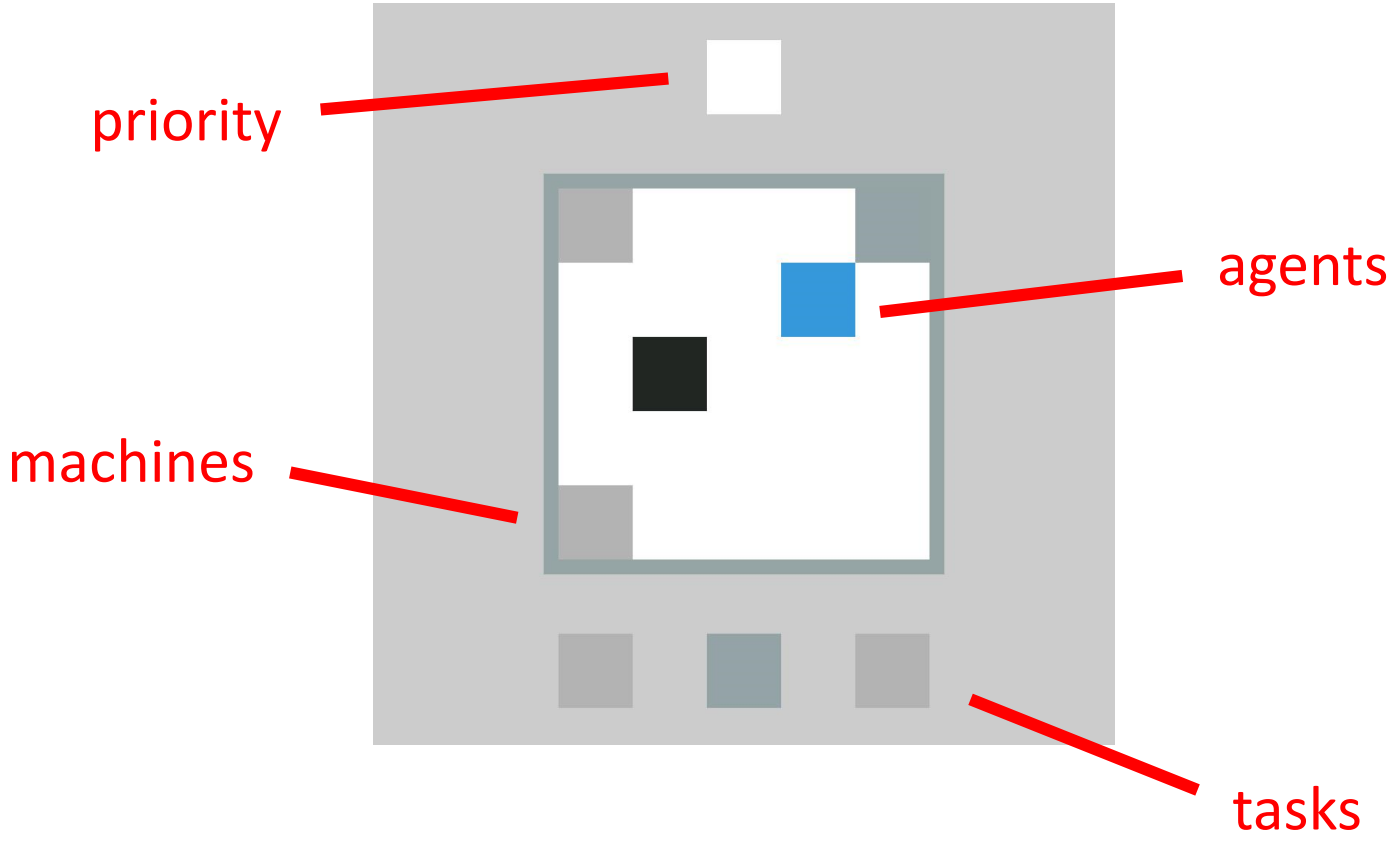
agents trade reward for the following action

agents are able to trade with extended action space

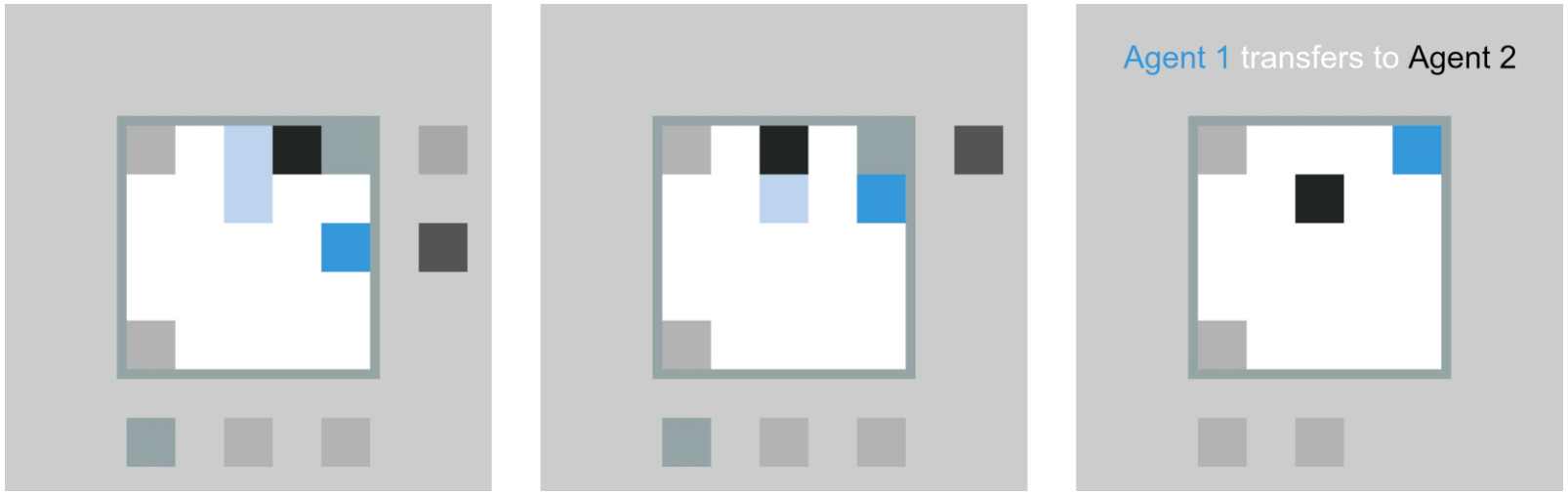
cooperating agents outperformed selfish ones

Source: Kyrill Schmid, Lenz Belzner, Thomas Gabor, and Thomy Phan **Action Markets in Deep Multi-Agent Reinforcement Learning**

Environment – Smart Factory



Action Trading

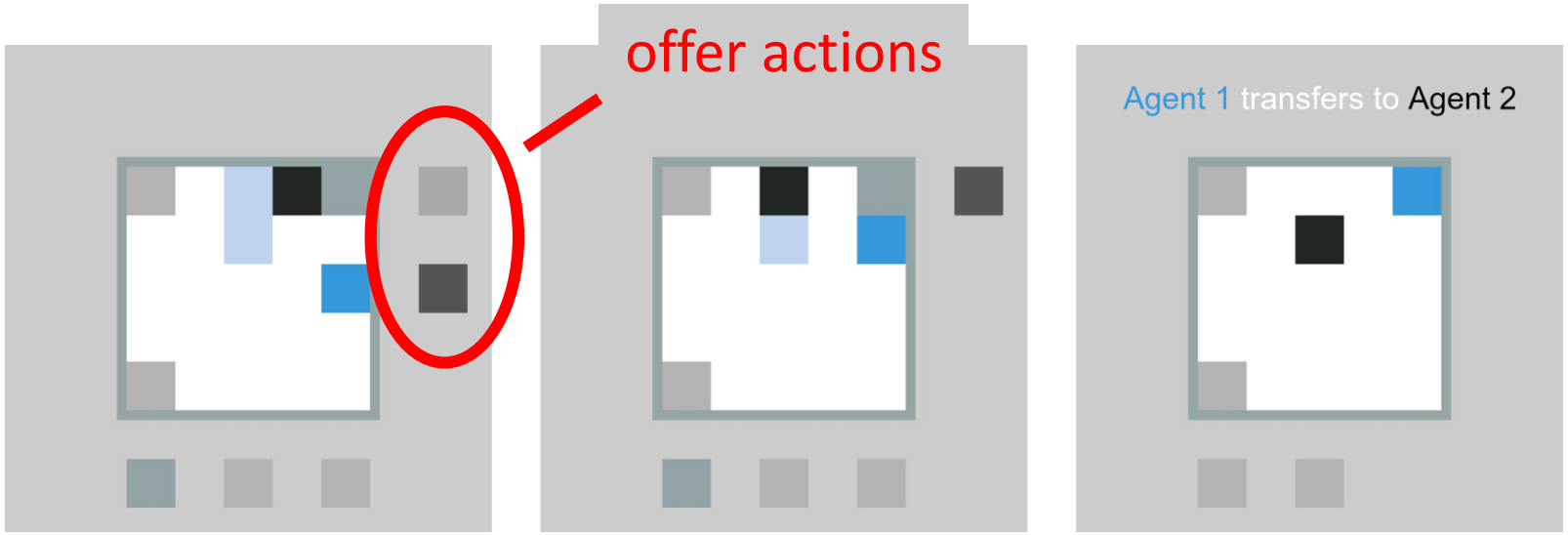


agent makes an offer consisting of n offer actions

the other agent can decide to perform the offer actions

by executing all offer actions the agent gets paid

Action Trading

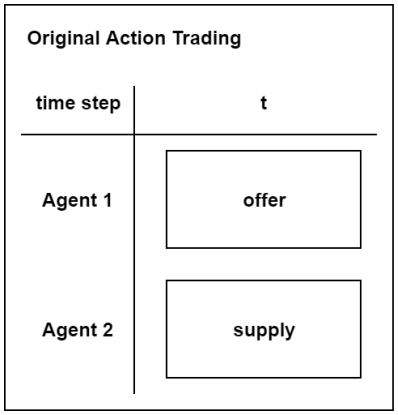


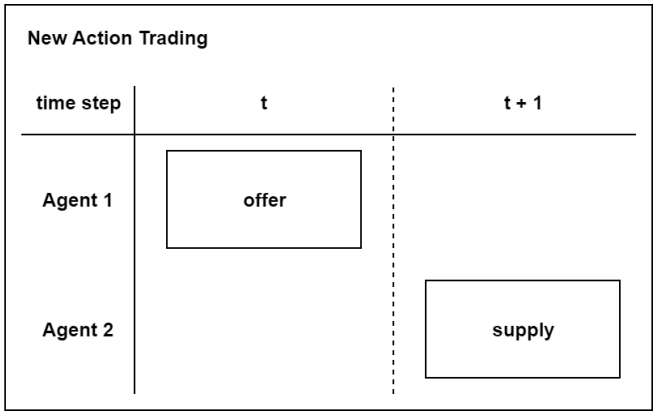
agent makes an offer consisting of n offer actions

the other agent can decide to perform the offer actions

by executing all offer actions the agent gets paid

Modification on Action Trading





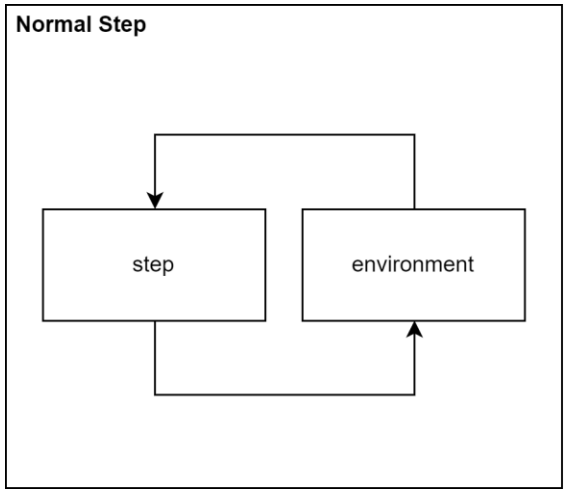
Original Action Trading:

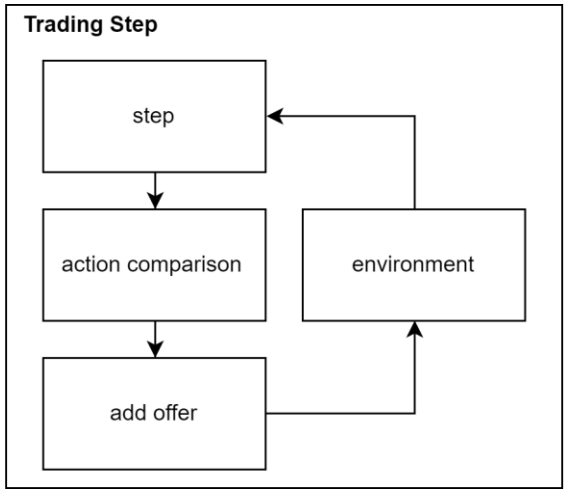
- offer and supply match at same time step
- fixed compensation
- not scalable offer actions

New Action Trading:

- offer and supply in separate time steps
- compensation based on state
- scalable offer actions

Environment with Action Trading



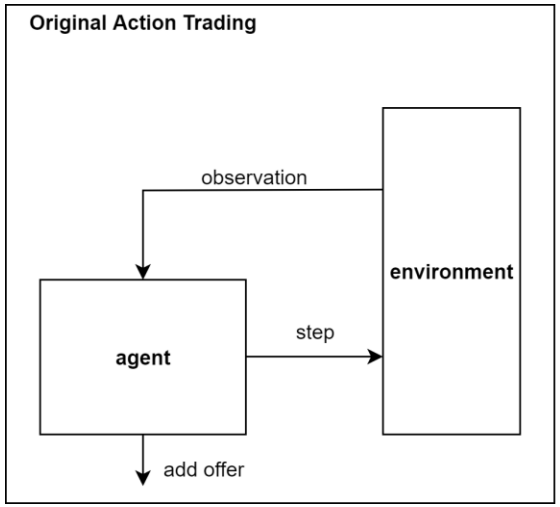


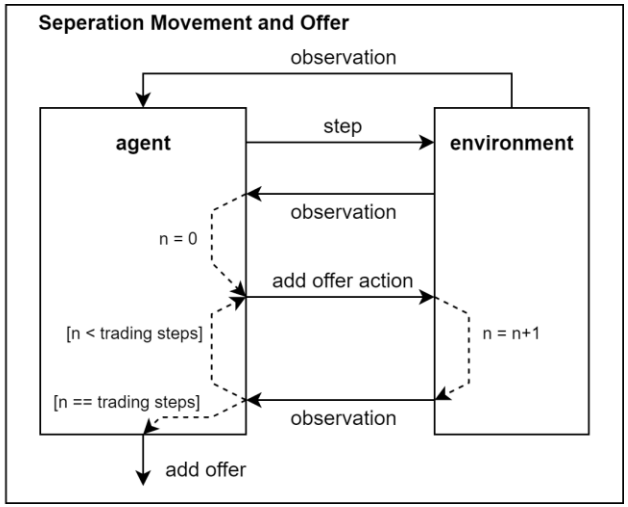
Extending normal environment step:

action comparison: checks if last action matched offer action

add offer: agent makes a new offer

Trading Modes





Original Action Trading:

one single action determines movement and offer

→ exponentially growing action space

Separation Movement and Offer:

seperates the selection of movement and offer actions

offer actions are added one at a time

→ constant action space

Offer:

OFFER ACTION 0
LEFT

Expected Rewards:

Q0 UP	Q1 DOWN	Q2 LEFT	Q3 RIGHT
0.05	0.01	- 0.07	- 0.04

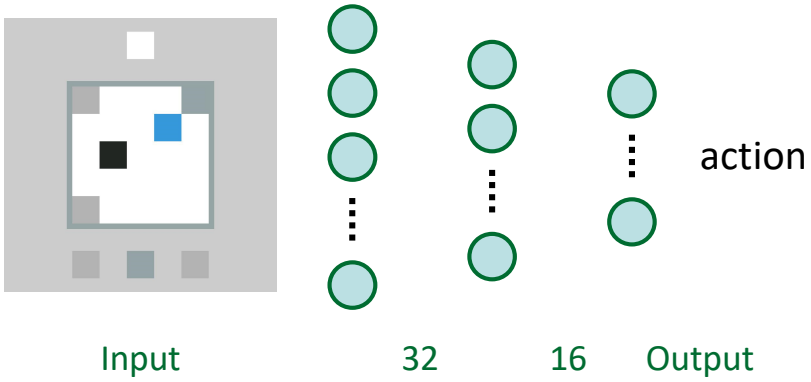
Difference D: $Q_{max} - Q_{offer}$ $0.05 - (-0.07) = 0.12$

Compensation in Future: D/γ $0.12/0.95 = 0.126$

$$\rightarrow \quad \boldsymbol{Compensation = Mc * \sum_{l=0}^{n-1} \frac{Q_{max}(t+l) - Q_{offer}(t+l)}{\gamma}}$$

Experiments

DQN



Decaying ε-greedy: $\varepsilon = 1 \rightarrow \varepsilon_{decay} = 8e^{-6} \rightarrow \varepsilon_{min} = 0.01$

Discount rate: $\gamma = 0.95$

20 runs: 1500 training episodes 2000 evaluation episodes

Experiments

Action Trading vs No Action Trading

Trading Mode: Original Action Trading

Trading Mode: Separation of Movement and Offer

Comparison between Trading Modes

Mark up

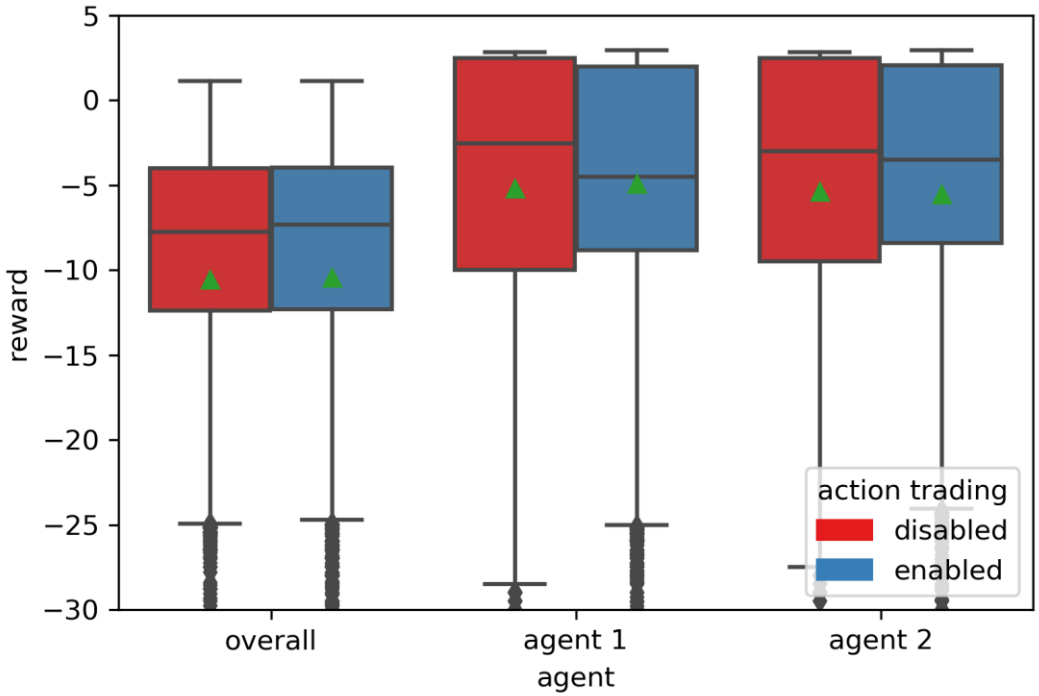
Budget

Payment Timing

Partial Payment

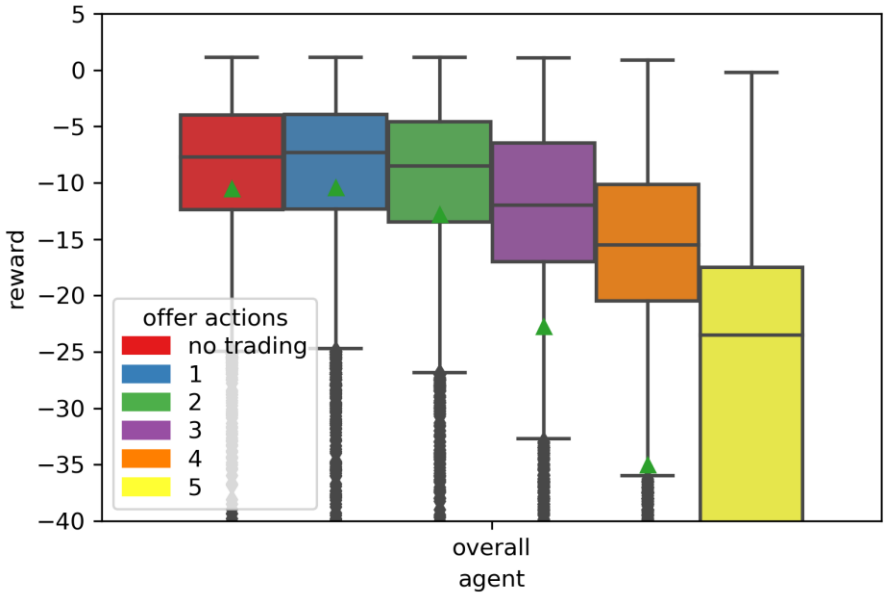
Experiments – Trading vs No Trading

Trading Mode: Original Action Trading

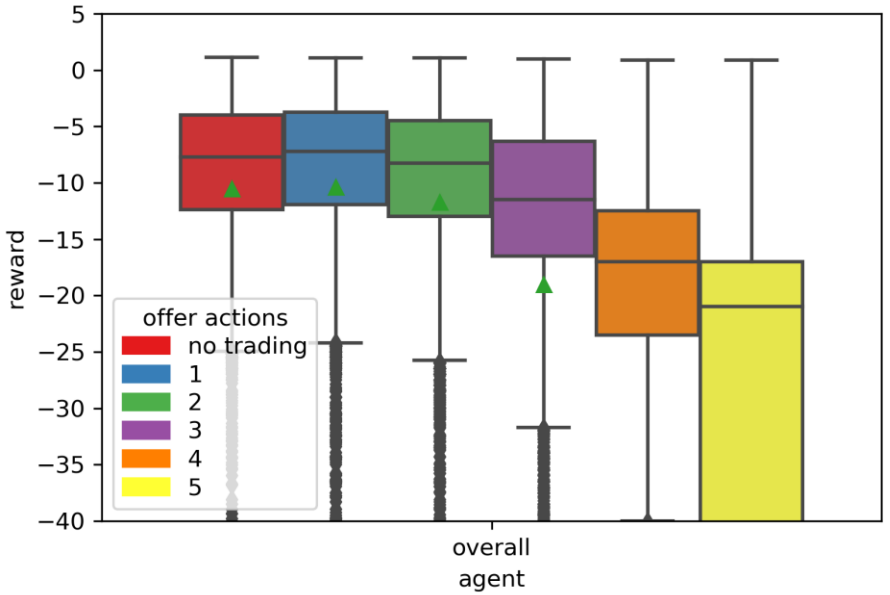


Experiments – Original Action Trading

Trading Mode: Original Action Trading



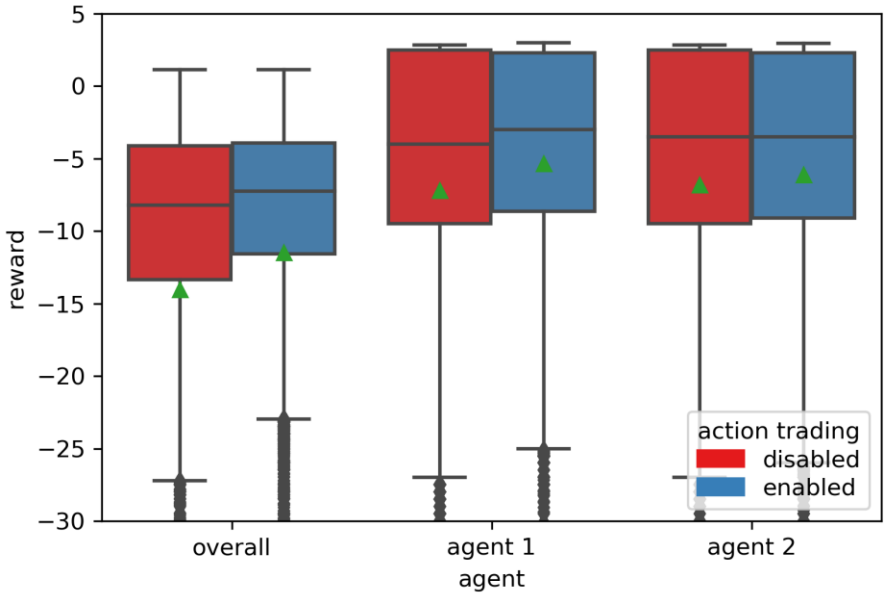
mark up: 1.00



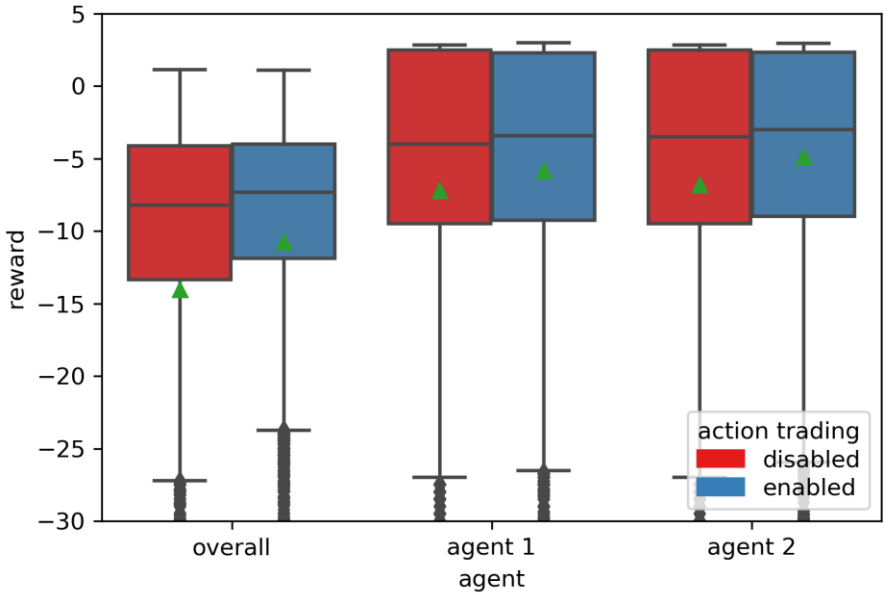
mark up: 1.05

Experiments – Separation Movement and Offer

Trading Mode: Separation Movement and Offer



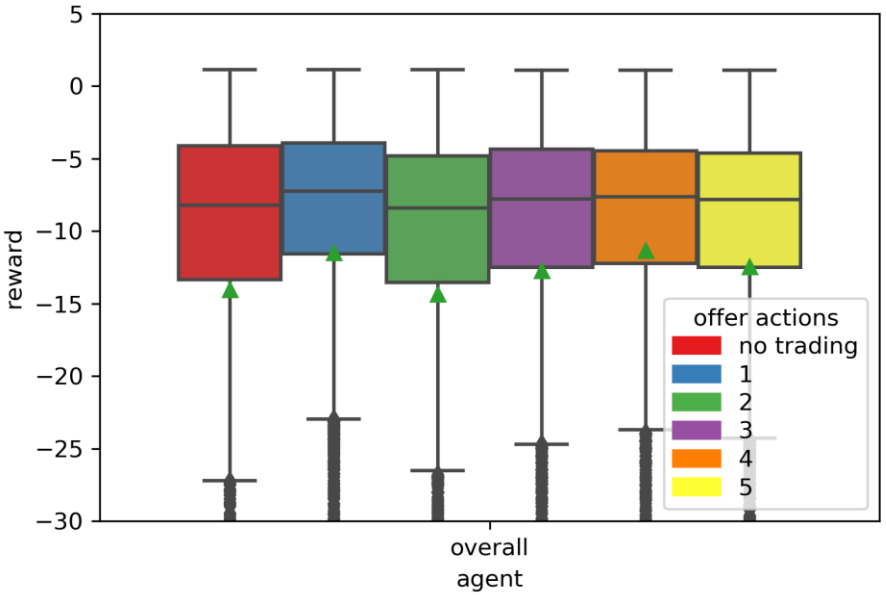
mark up: 1.00



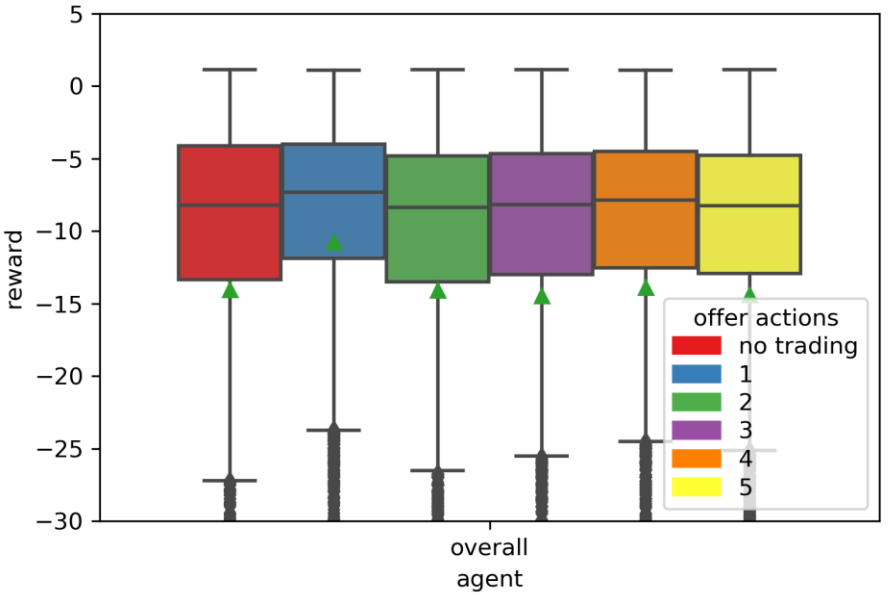
mark up: 1.05

Experiments – Separation Movement and Offer

Trading Mode: Separation Movement and Offer



mark up: 1.00



mark up: 1.05

Experiments – Comparison Trading Modes

Original Action Trading

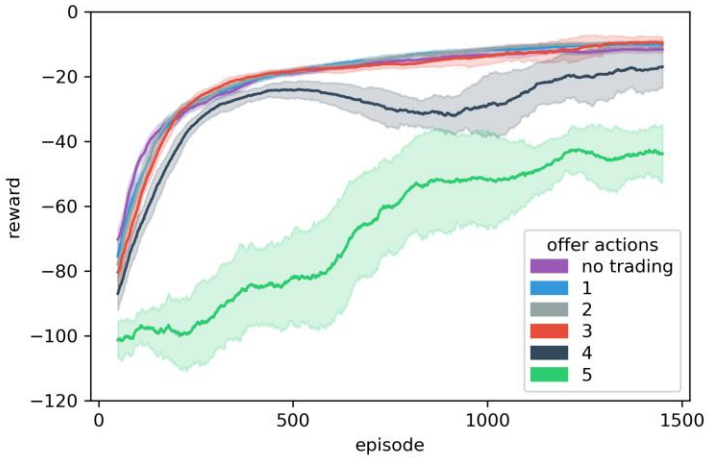
n	mark up	action space	trades/episode	mean reward	median reward
-	-	4	-	-10.57	-7.74
1	1.0	20	2.80	-10.45	-7.32
2	1.0	68	0.60	-12.82	-8.5
3	1.0	260	0.13	-22.77	-12.00
4	1.0	1028	0.03	-35.02	-15.50
5	1.0	4100	0.00	-86.71	-23.50
1	1.05	20	2.78	-10.40	-7.20
2	1.05	68	0.62	-11.74	-8.24
3	1.05	260	0.13	-19.03	-11.50
4	1.05	1028	0.03	-49.30	-17.00
5	1.05	4100	0.00	-70.66	-21.00

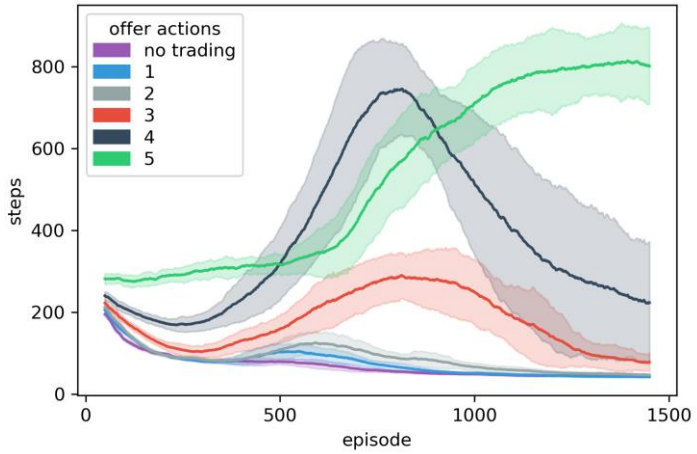
Separation Movement and Offer

n	mark up	action space	trades/episode	mean reward	median reward
-	-	8	-	-14.04	-8.22
1	1.0	8	1.88	-11.49	-7.24
2	1.0	8	0.40	-14.34	-8.4
3	1.0	8	0.16	-12.73	-7.8
4	1.0	8	0.11	-11.33	-7.64
5	1.0	8	0.04	-12.45	-7.84
1	1.05	8	1.70	-10.75	-7.34
2	1.05	8	0.51	-14.07	-8.36
3	1.05	8	0.17	-14.48	-8.16
4	1.05	8	0.09	-13.91	-7.88
5	1.05	8	0.05	-14.41	-8.26

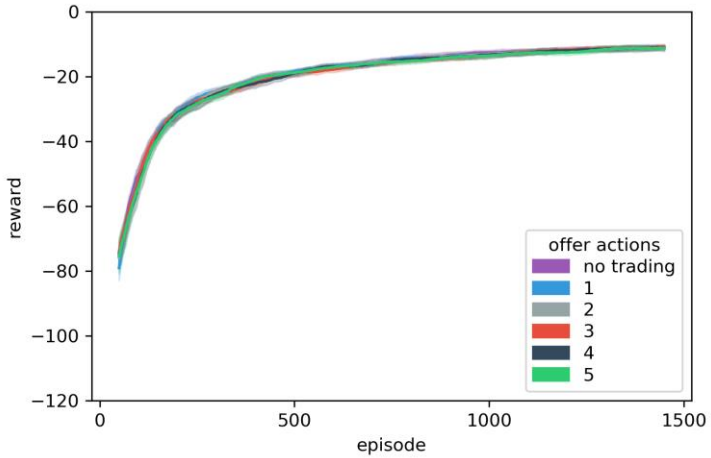
Experiments – Comparison Trading Modes

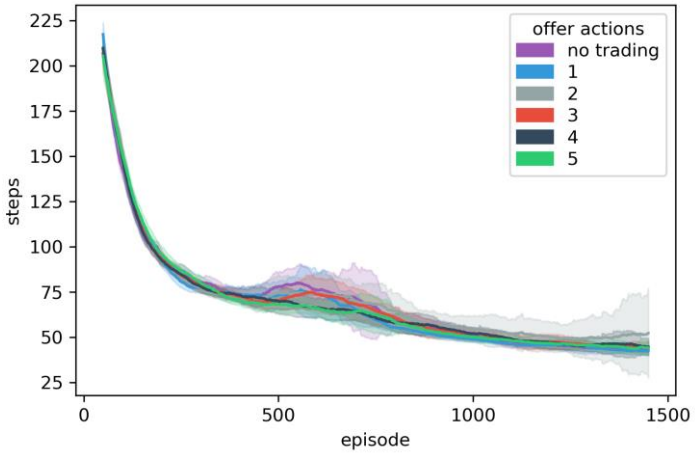
Original Action Trading





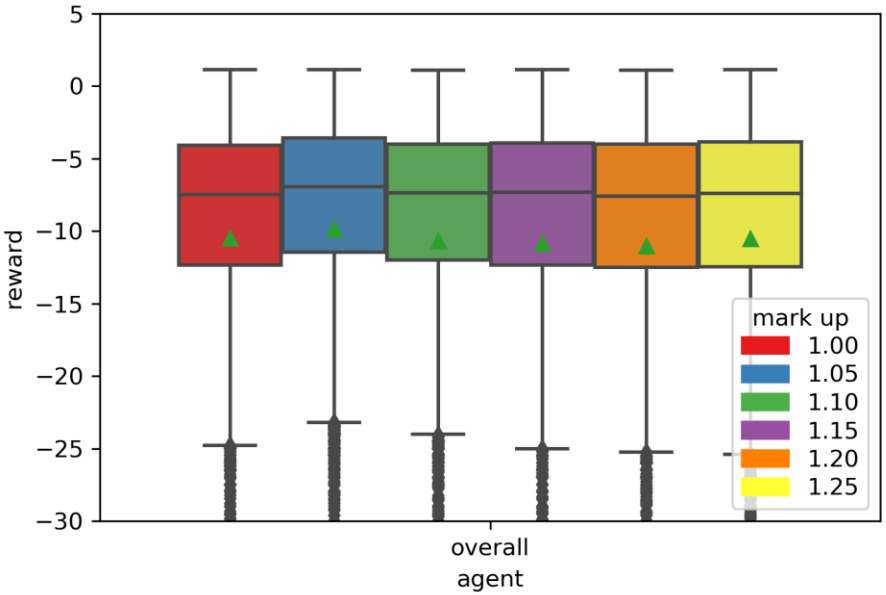
Separation Movement and Offer



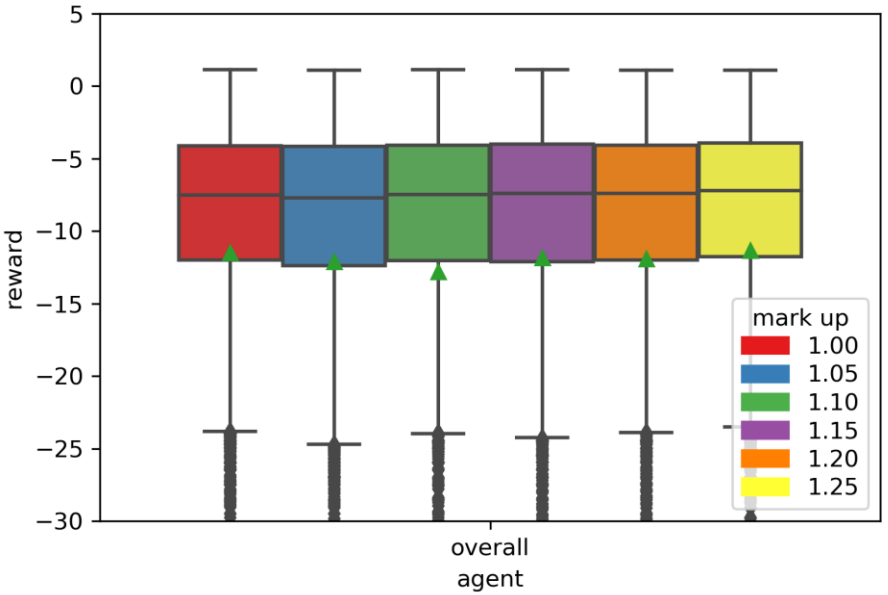


Experiments – Mark up

Original Action Trading

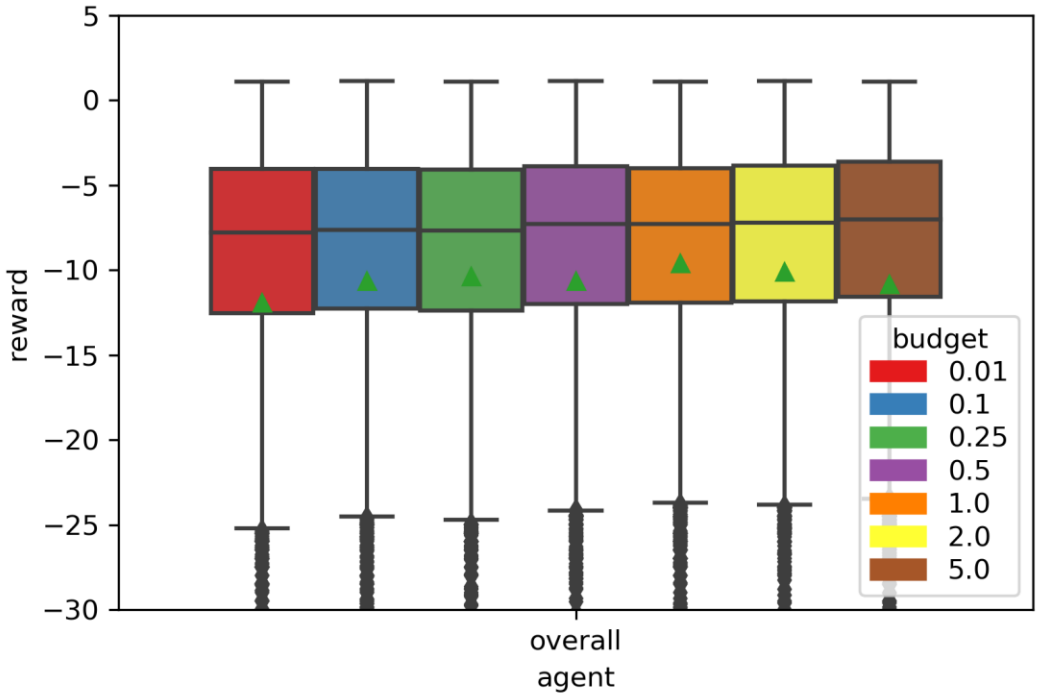


Separation Movement and Offer



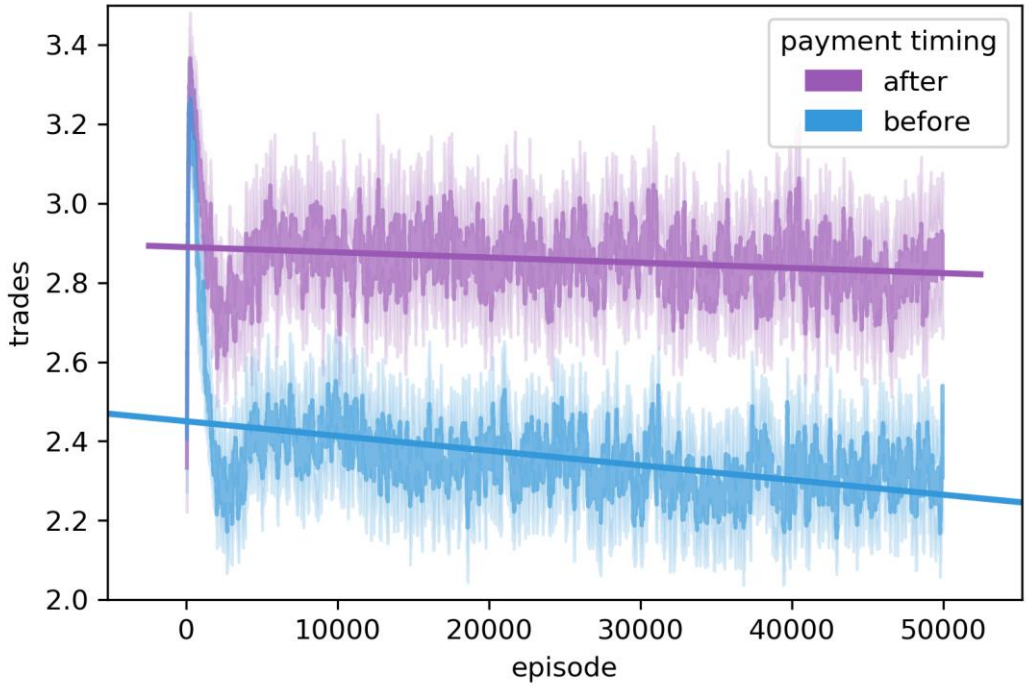
Experiments – Budget

Trading Mode: Original Action Trading

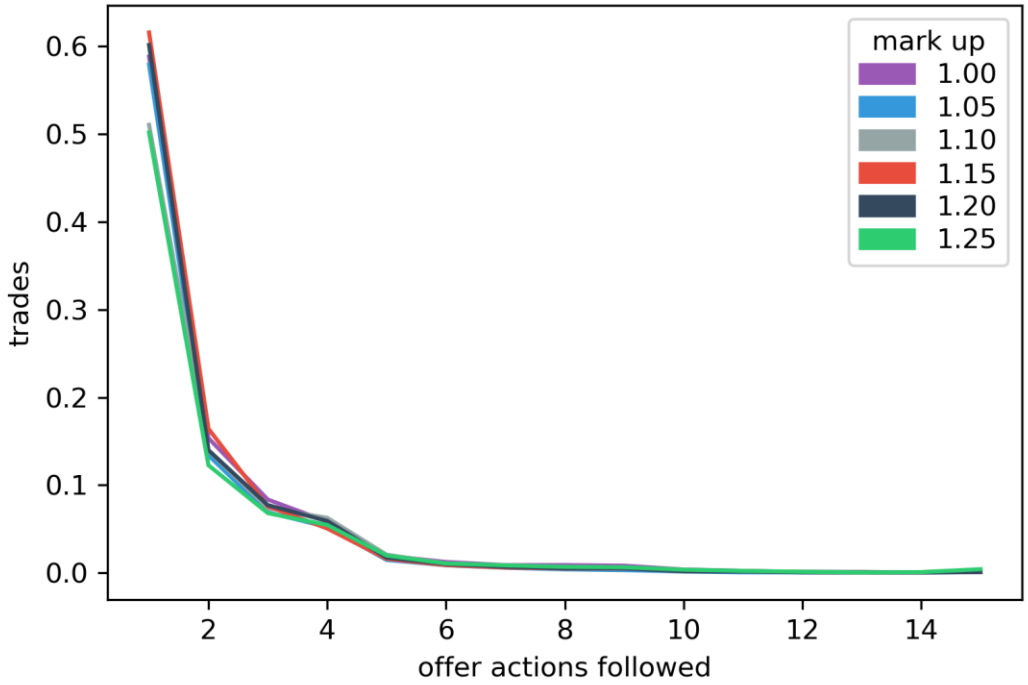


Experiments – Payment Timing

Trading Mode: Original Action Trading



Trading Mode: Separation Movement and Offer



Conclusion

Action Trading outperforms No Action Trading

Trading Modes have different advantages:

- Original Action Trading: more cooperation
- Separation Movement and Offer: stable across n

Existing optimal mark up

Budget has only negative impact

Less cooperation between agents who pay beforehand

Agents rather cooperate short- than long-term

Thank you for your attention