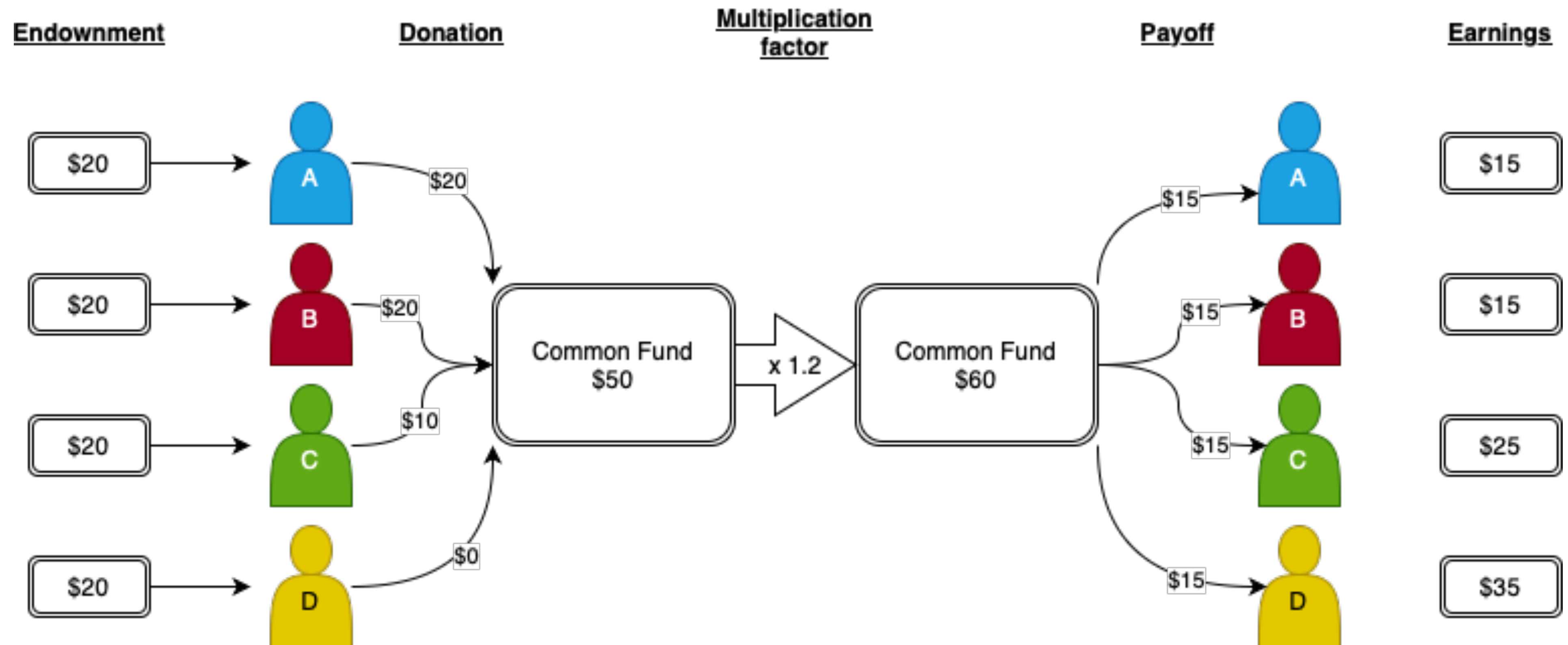


# **Empirical Evaluation of Overestimation Bias in Q-learning and Double Q-learning**

Lai Jien Weng (ID: 2104438)  
Supervised by: Dr. Tan Wei Lun

# Introduction



**“Freedom in a commons brings ruins to all.”**

**Hardin, G. (1968) ‘The Tragedy of the Commons’, 1, pp. 243–253. Available at: <https://doi.org/10.1126/science.162.3859.1243>.**

# Problem Statement

1. How does the multiplication factor affect cooperation in the Public Goods Game when modelled with Q-learning and Double Q-learning?
2. How do different endowments influence contribution strategies in the Public Goods Game using Q-learning and Double Q-learning?
3. How does a large action space impact fairness in the Public Goods Game when analysed with Q-learning and Double Q-learning?

# Objectives

1. To study how the multiplication factor affects cooperation in the Public Goods Game using Q-learning and Double Q-learning.
2. To explore the impact of different endowments on contribution strategies in the Public Goods Game with Q-learning and Double Q-learning.
3. To assess how a large action space influences fairness in the Public Goods Game using Q-learning and Double Q-learning.

# Significance

1. Shapes incentives for collective action.
2. Addresses inequities in resource allocation.
3. Promotes fairness in cooperative systems.

# Literature Review

# Summary of Key Prior Works

Study	Approach	Focus	Limitations
ManChon U and Zhen Li (2010)	TD-learning	Homogeneous endowments, cooperation rates	Binary actions, no fairness metrics
Fehr and Gächter (2002)	Experimental PGG	Variable contributions, punishment	Non-RL, human subjects only
Fischbacher, Gächter and Fehr (2001)	Experimental PGG	Conditional cooperation	Non-RL, no MARL framework
Isaac, Walker and Thomas (1984)	Experimental PGG	Heterogeneous endowments	Non-RL, limited to small groups
Rashid <i>et al.</i> (2018)	Deep RL (QMIX)	Scalable cooperation in dilemmas	Limited interpretability, no fairness focus
Jaques <i>et al.</i> (2019)	Deep RL with communication	Coordination via social influence	Homogeneous agents, no endowment variation



# Methodology

# PGG: Nash vs. Pareto Outcomes

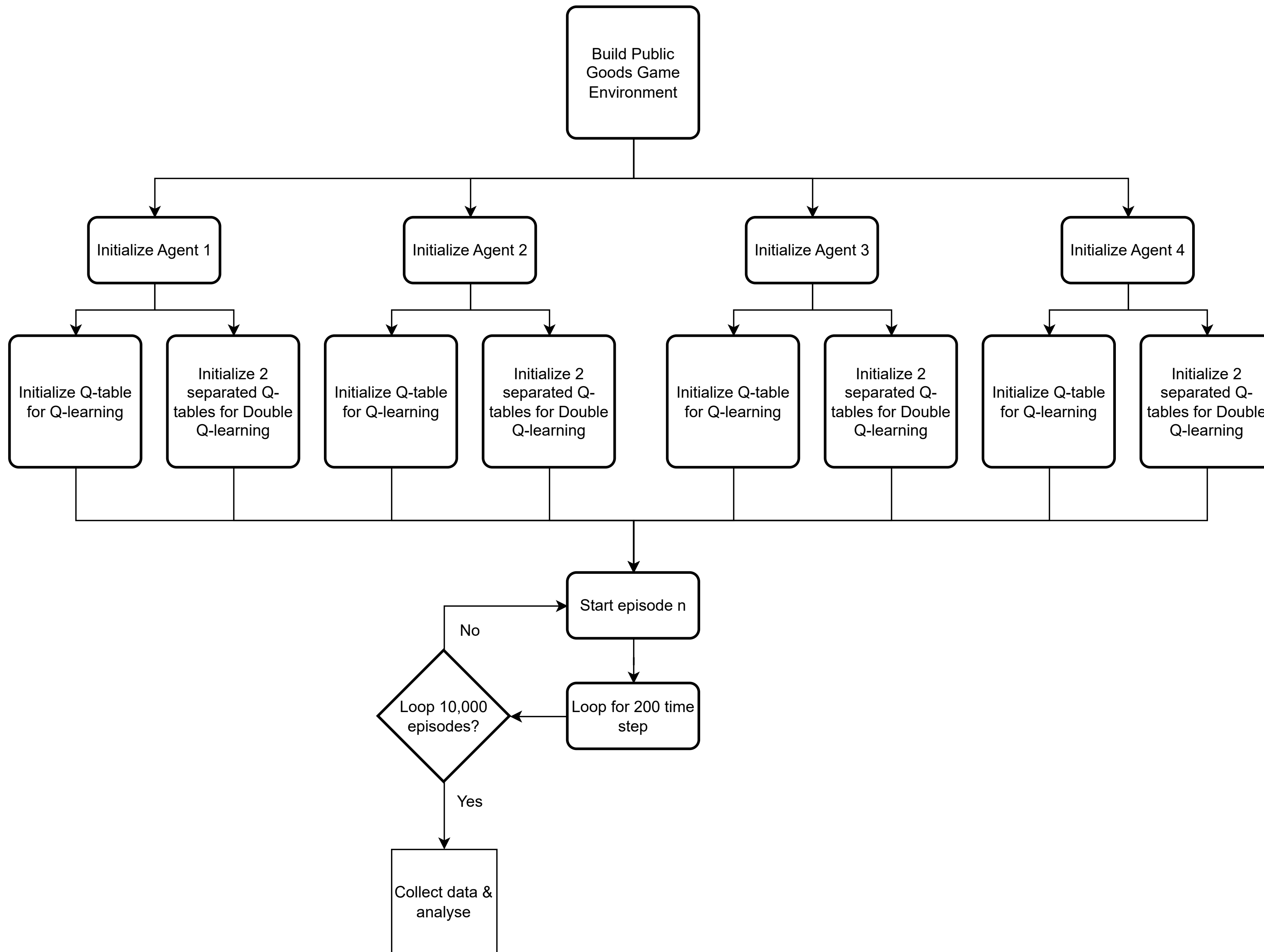
		Agent 2	
		Free-ride	Contribute
Agent 1	Free-ride	(1, 1)	(1.75, 0.75)
	Contribute	(0.75, 1.75)	(1.5, 1.5)

Note: Example shown with  $r = 1.5$ ,  $n = 2$ ,  $\mathcal{A}_i = \{0, e_i\}$ ,  $e_i = \{1, 1\}$

# PGG: Nash vs. Pareto Outcomes

		Agent 2	
		Free-ride	Contribute
Agent 1	Free-ride	<div>Nash Equilibrium (1, 1)</div>	(1.75, 0.75)
	Contribute	(0.75, 1.75)	<div>Pareto Optimality (1.5, 1.5)</div>

Note: Example shown with  $r = 1.5$ ,  $n = 2$ ,  $\mathcal{A}_i = \{0, e_i\}$ ,  $e_i = \{1, 1\}$



# Q-learning

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(s, a) \left[ u_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

where,

- $s_t \in S$  is the state at time  $t$ ,
- $a_t \in A$  is the action taken at time  $t$ ,
- $s_{t+1} \in S$  is the next state at time  $t$ ,
- $u_t$  is the reward at time  $t$ ,
- $\gamma \in [0,1]$  is the discount factor,
- $\alpha(s, a) \in (0,1]$  is the learning rate, and
- $\max_a Q(s_{t+1}, a)$  is the highest expected future rewards.

# Double Q-learning

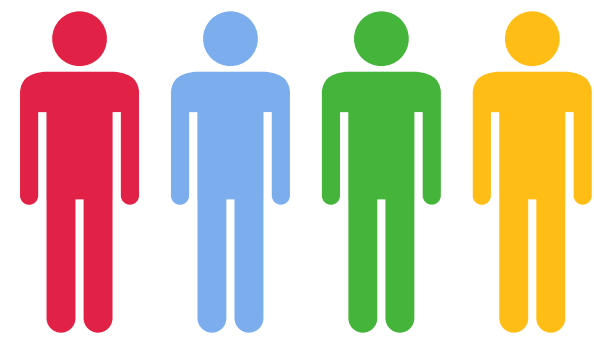
$$Q^A(s_t, a_t) \leftarrow Q^A(s_t, a_t) + \alpha \left[ u_t + \gamma Q^B[s_{t+1}, \arg \max_a Q^A(s_{t+1}, a)] \right]$$

$$Q^B(s_t, a_t) \leftarrow Q^B(s_t, a_t) + \alpha \left[ u_t + \gamma Q^A[s_{t+1}, \arg \max_a Q^B(s_{t+1}, a)] \right]$$

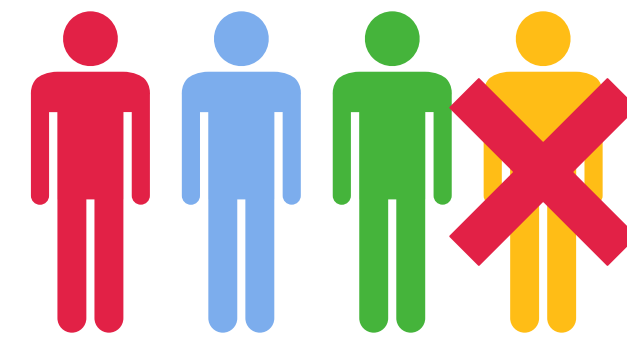
where all variables follow the definition in previous slide.

# Uncertainty in PGG

- Random exclusion: 75% all agent participates, 25% one random agent excluded.

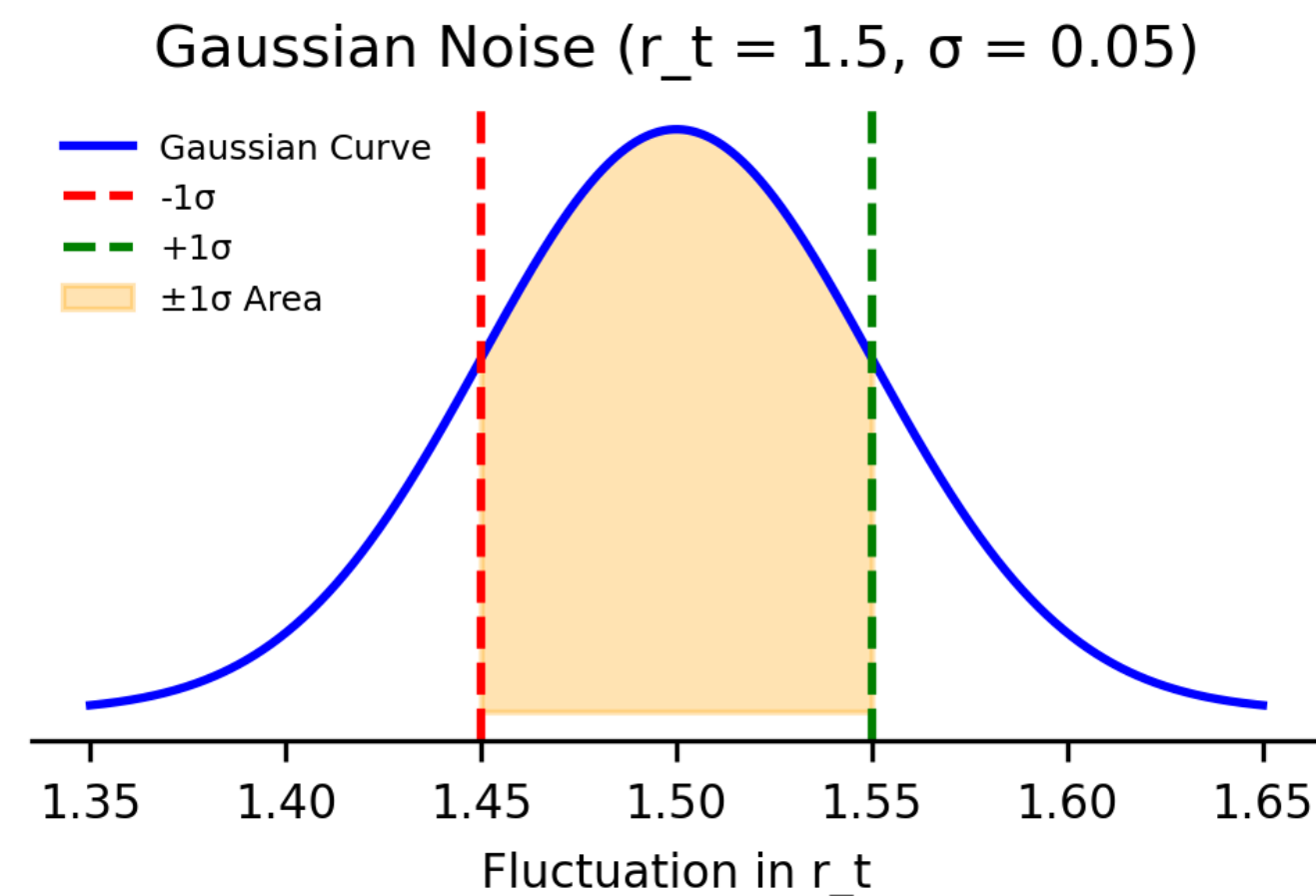


75%



25%

- Gaussian noise: Multiplication factor fluctuate within  $\sigma_r = 0.05$ .

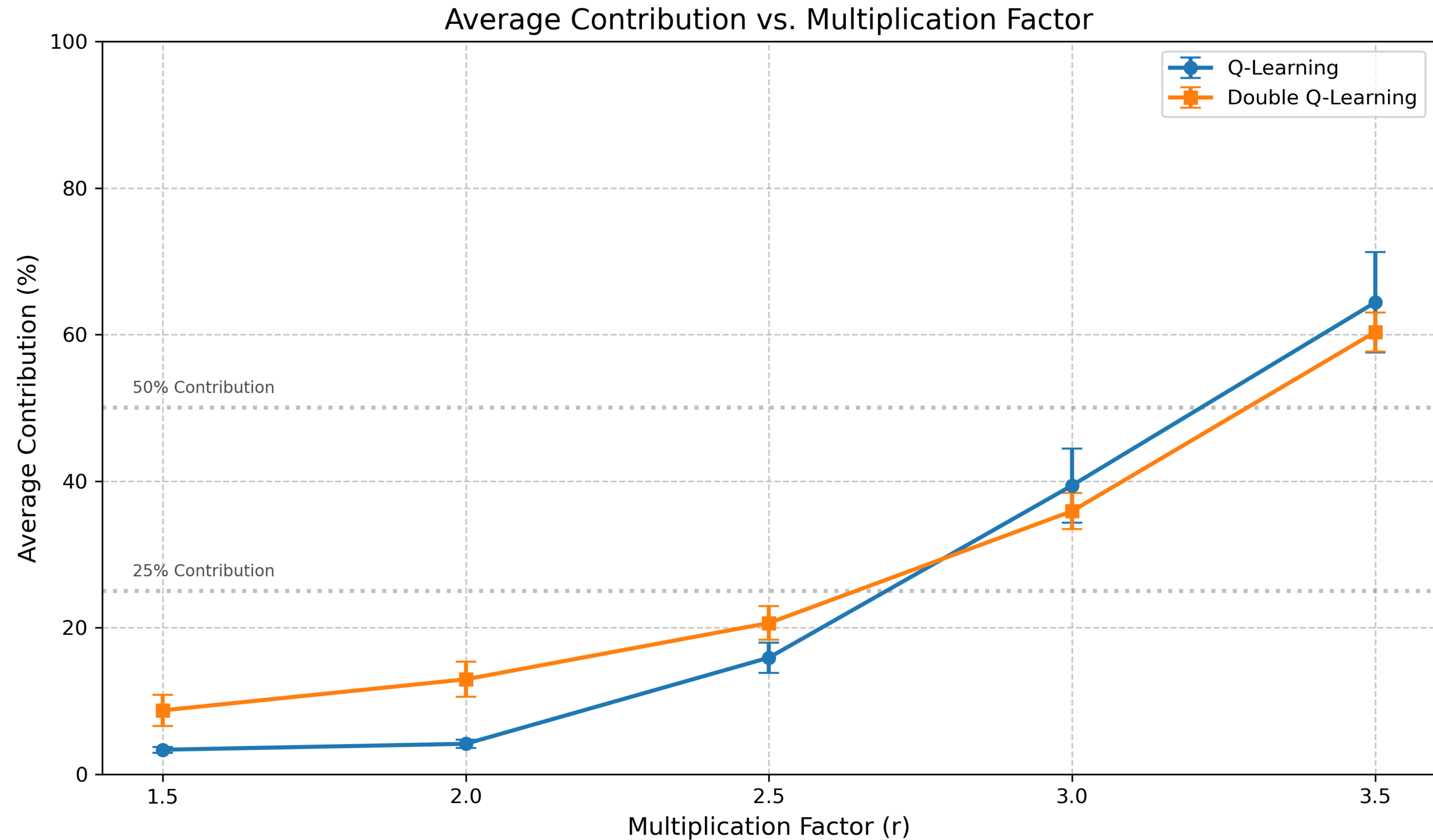


# Experiment 1: Multiplication Factor

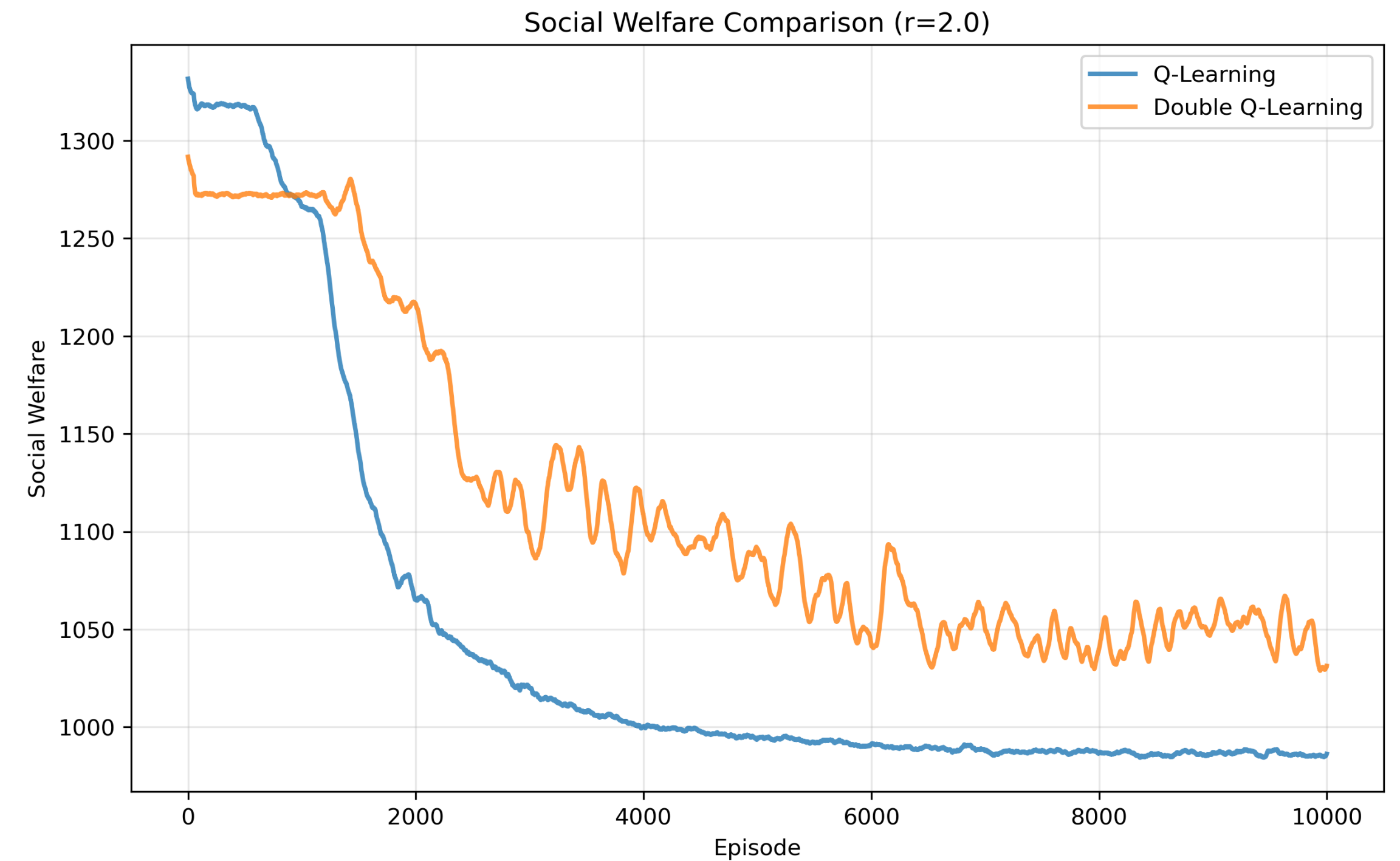
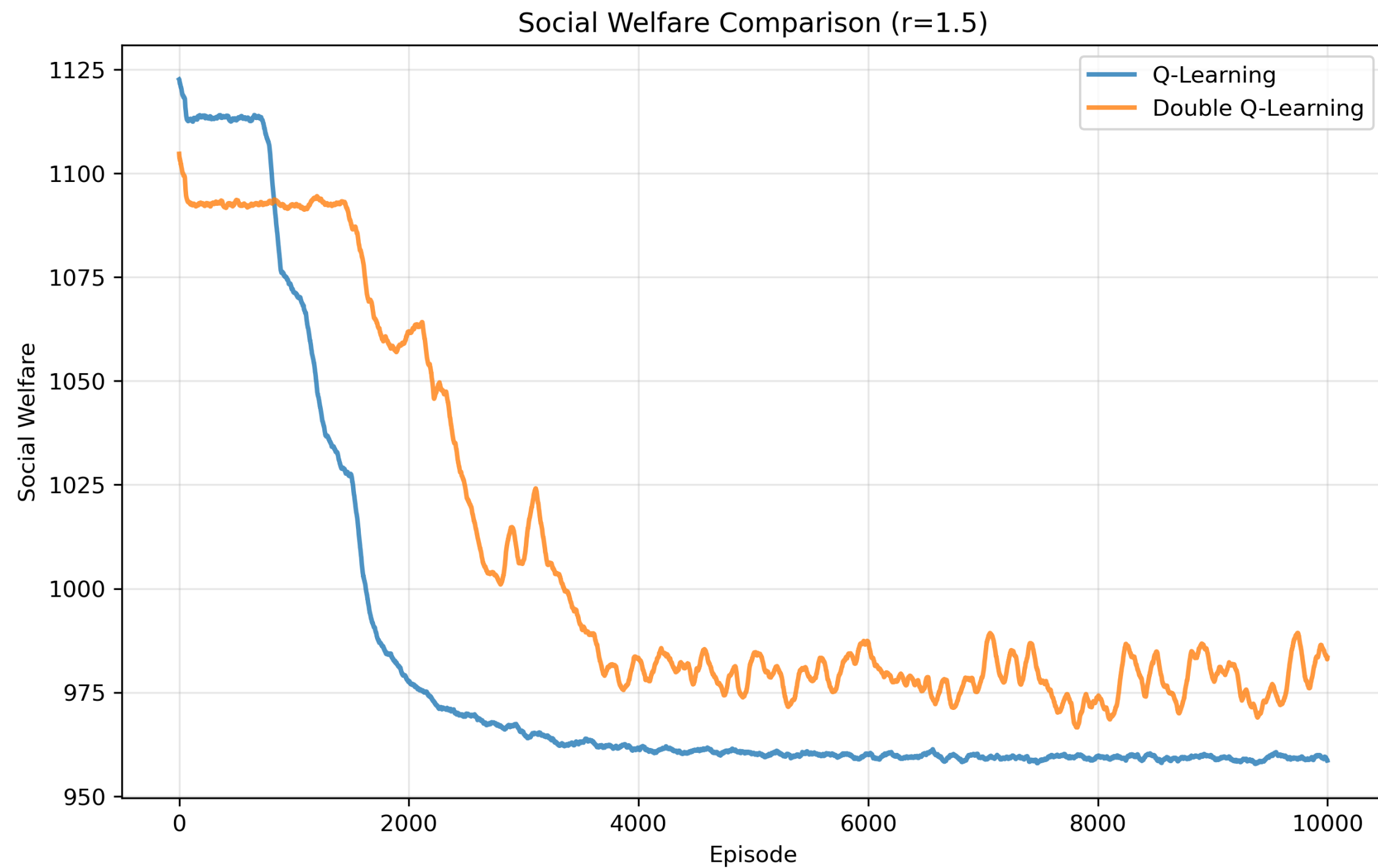
- Run experiment for  $r_t = \{1.5, 2.0, 2.5, 3.0, 3.5\}$ .
- Measure with:
  - Contribution rate across,  $\bar{a}_i$  against  $r_t$ ,
  - Social welfare,  $W_t$ ,
  - $\sigma_{contrib}$ .



# Experiment 1: Multiplication Factor (cont.)

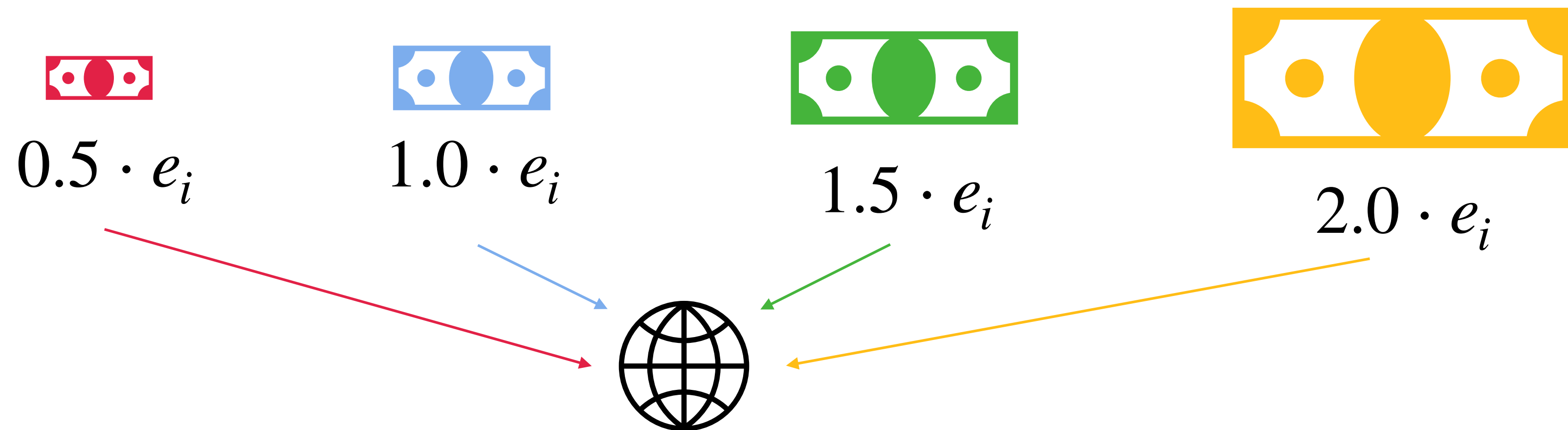


# Experiment 1: Multiplication Factor (cont.)



# Experiment 2: Heterogeneous Endowments

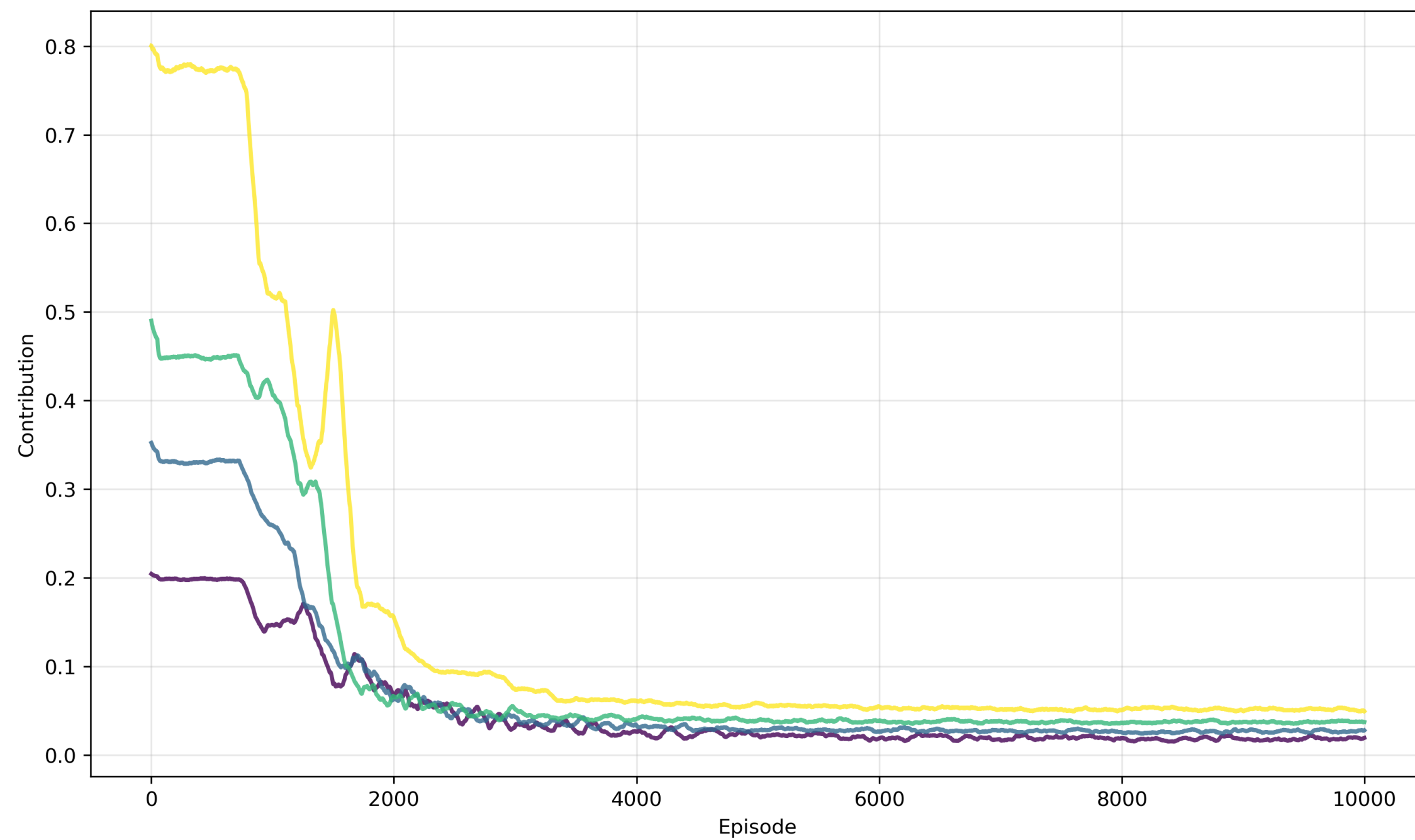
- Run experiment for  $e_i = \{0.5, 1.0, 1.5, 2.0\}$ .



- Measure with:
  - Individual contribution per episode,  $\tilde{a}_{i,k}$ .

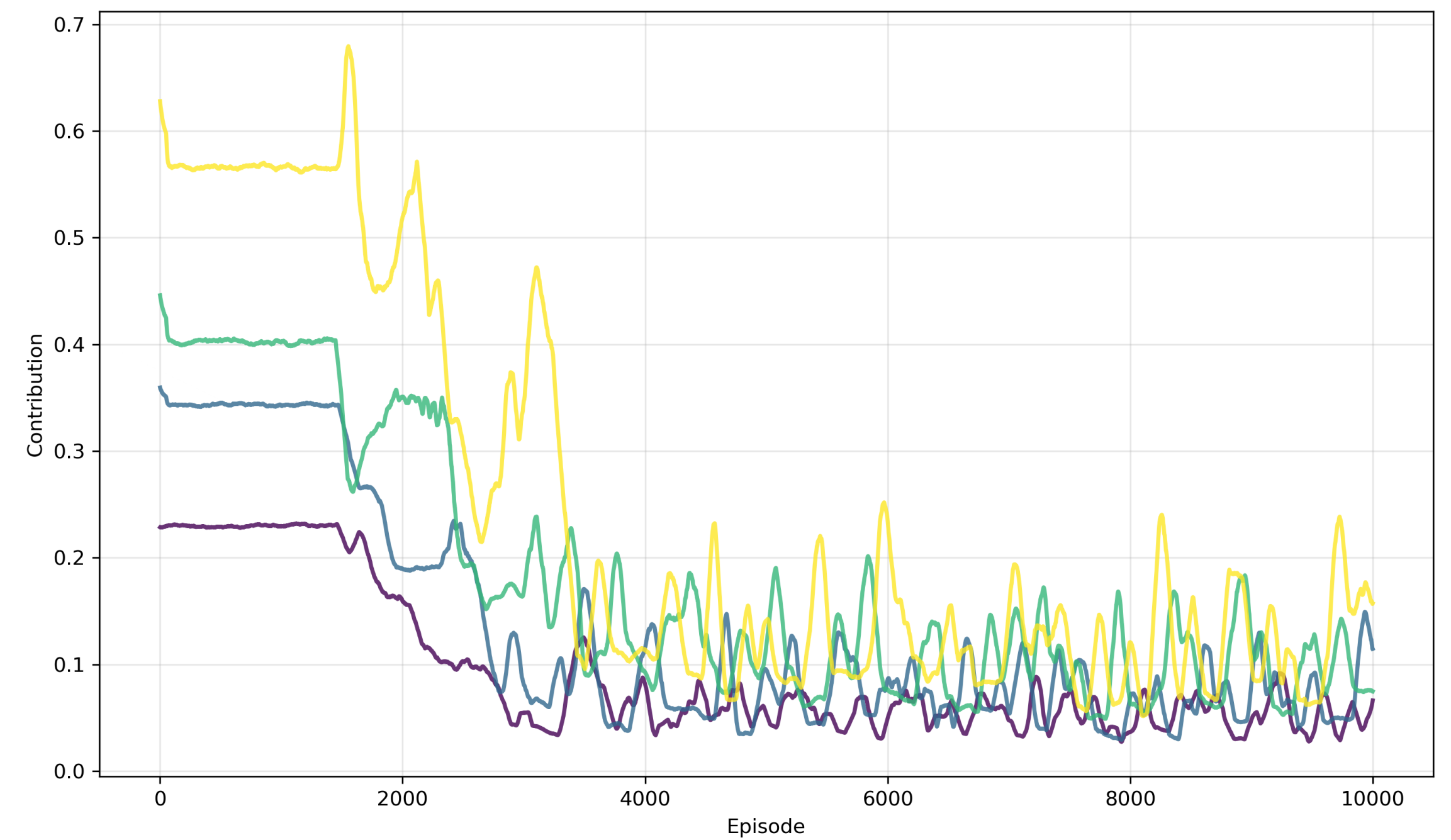
# Experiment 2: Heterogeneous Endowment (cont.)

Individual Contribution of Q-learning



With  $r_t = 1.5$

Individual Contribution of Double Q-learning



With  $r_t = 1.5$

# Experiment 3: Fairness

- Run experiment for 25-level discrete contribution options,  $\mathcal{A}_i = \{0, 0.04e_i, \dots, e_i\}$ .
- Test with a control set with 4-level discrete contribution options.
- Measure with:
  - Shapley value,  $\phi_i$

# Statistical Testing

- All tests will be tested with paired t-test.

# Summary

- This study investigated cooperative behaviour in PGG within a MARL framework of:
  1. Multiplication factor
  2. Heterogeneous endowments
  3. Fairness in reward distribution
- The experiment employed:
  - A. Tabular Q-learning
  - B. Double Q-learning

**Thank you**