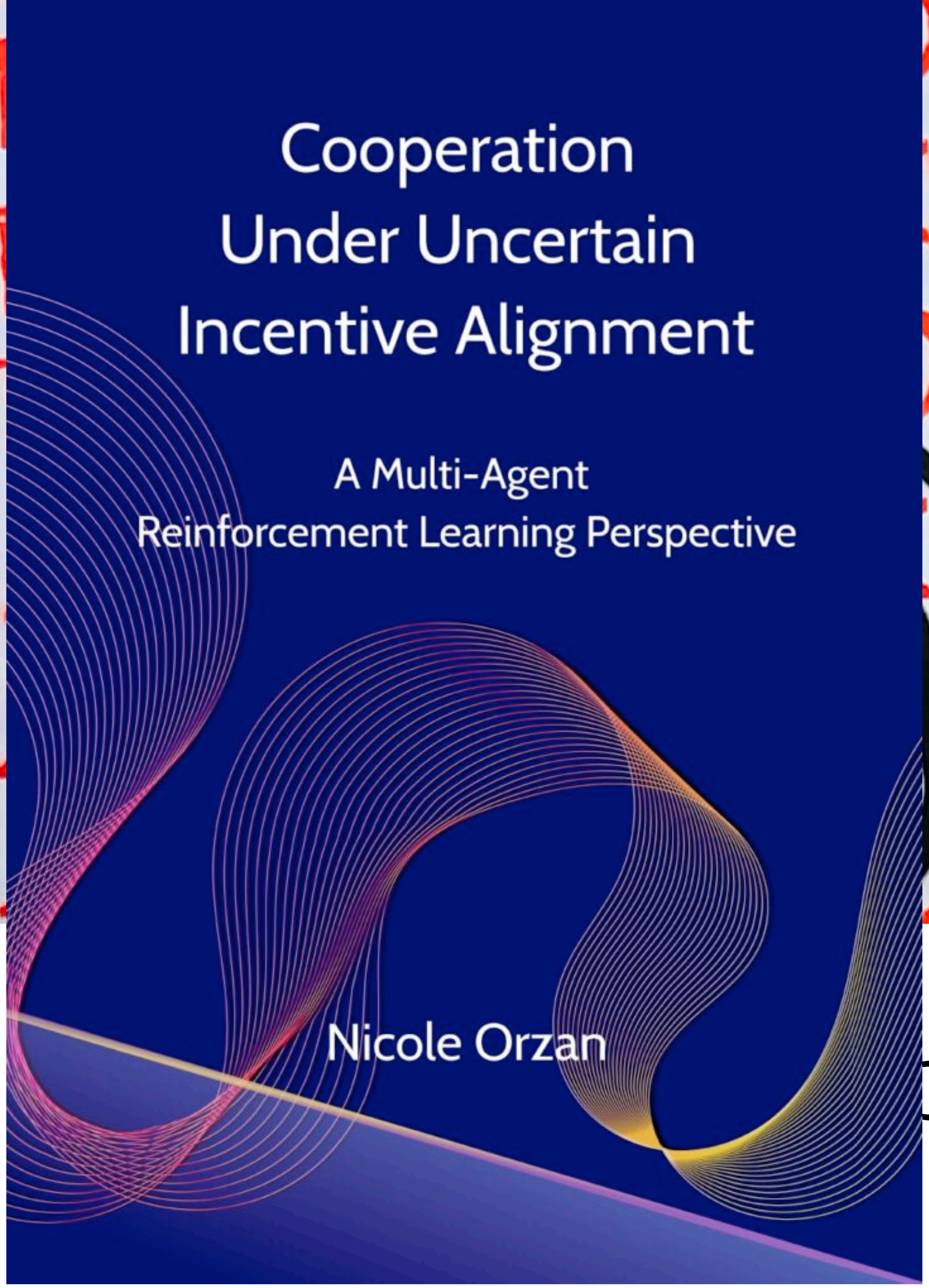
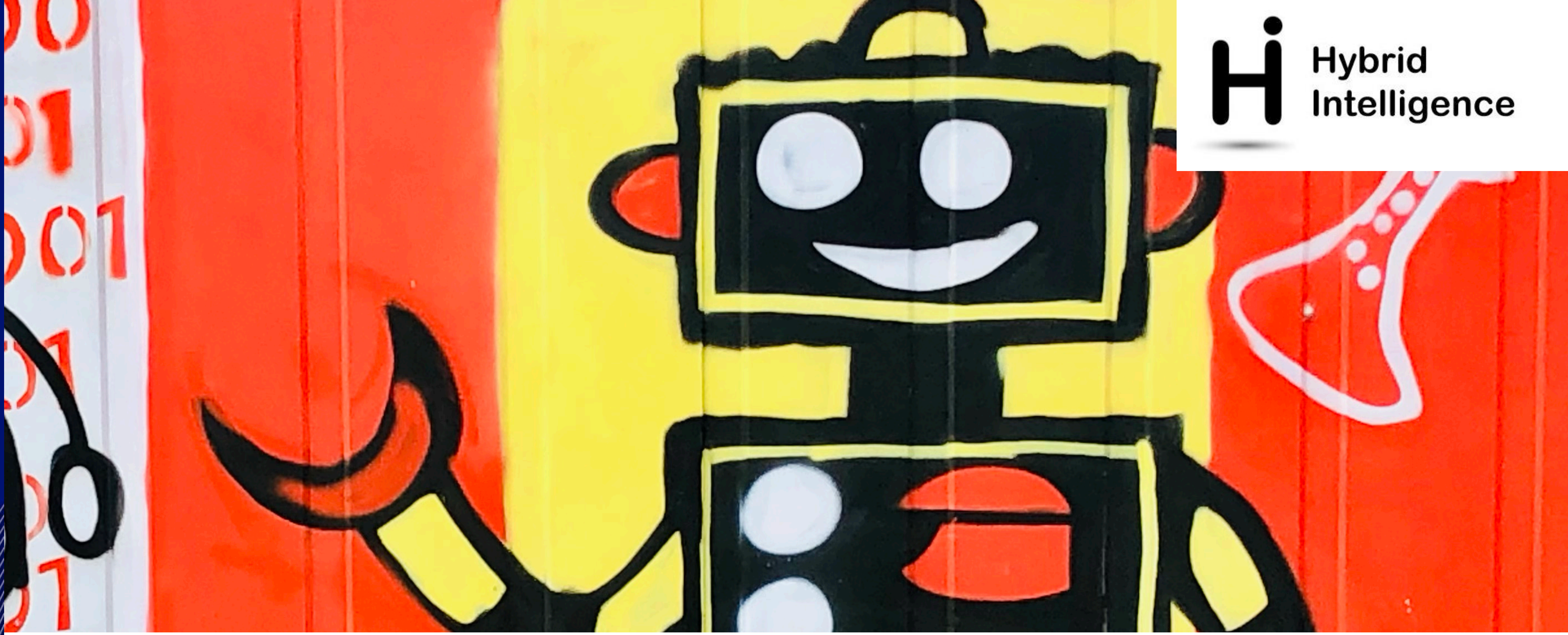




Hybrid
Intelligence

perate under Uncertain Incentive Alignment

Niki Orzan

Erman Acar

Davide Grossi

university of groningen

University of Amsterdam

Roxana Radulescu

Context

REVIEW

Science 2015

Economic reasoning and artificial intelligence

David C. Parkes[1*] and Michael P. Wellman[2*]

Economics is drawn to rational decision models because they directly connect choices and values in a mathematically precise manner. Critics argue that the field studies a mythical species, *homo economicus* ("economic man") and produces theories with limited applicability to how real humans behave. Defenders acknowledge that rationality is an idealization but counter that the abstraction supports powerful analysis, which is often quite predictive of people's behavior. […]

Artificial intelligence research is likewise drawn to rationality concepts, because they provide an ideal for the computational artifacts it seeks to create. Core to the modern conception of AI is the idea of designing agents: entities that perceive the world and act in it. The quality of an AI design is judged by how well the agent's actions advance specified goals, conditioned on the perceptions observed. This coherence among perceptions, actions, and goals is the essence of rationality. If we represent goals in terms of preference over outcomes, and conceive perception and action within the framework of decision-making under uncertainty, then the AI agent's situation aligns squarely with the standard economic paradigm of rational choice.

Thus, the AI designer's task is to build rational agents, or agents that best approximate rationality given the limits of their computational resources. In other words, AI strives to construct---out of silicon (or whatever) and information---a synthetic *homo economicus*, perhaps more accurately termed *machina economica*.

COOPERATION WITHOUT COMMUNICATION

Michael R. Genesereth, Matthew L. Ginsberg, and Jeffrey S. Rosenschein*

Logic Group, Knowledge Systems Laboratory,
Computer Science Department, Stanford University,
Stanford, California 94305

"Intelligent agents will inevitably need to interact flexibly with other entities. The existence of conflicting goals will need to be handled by these automated agents, just as it is routinely handled by humans."

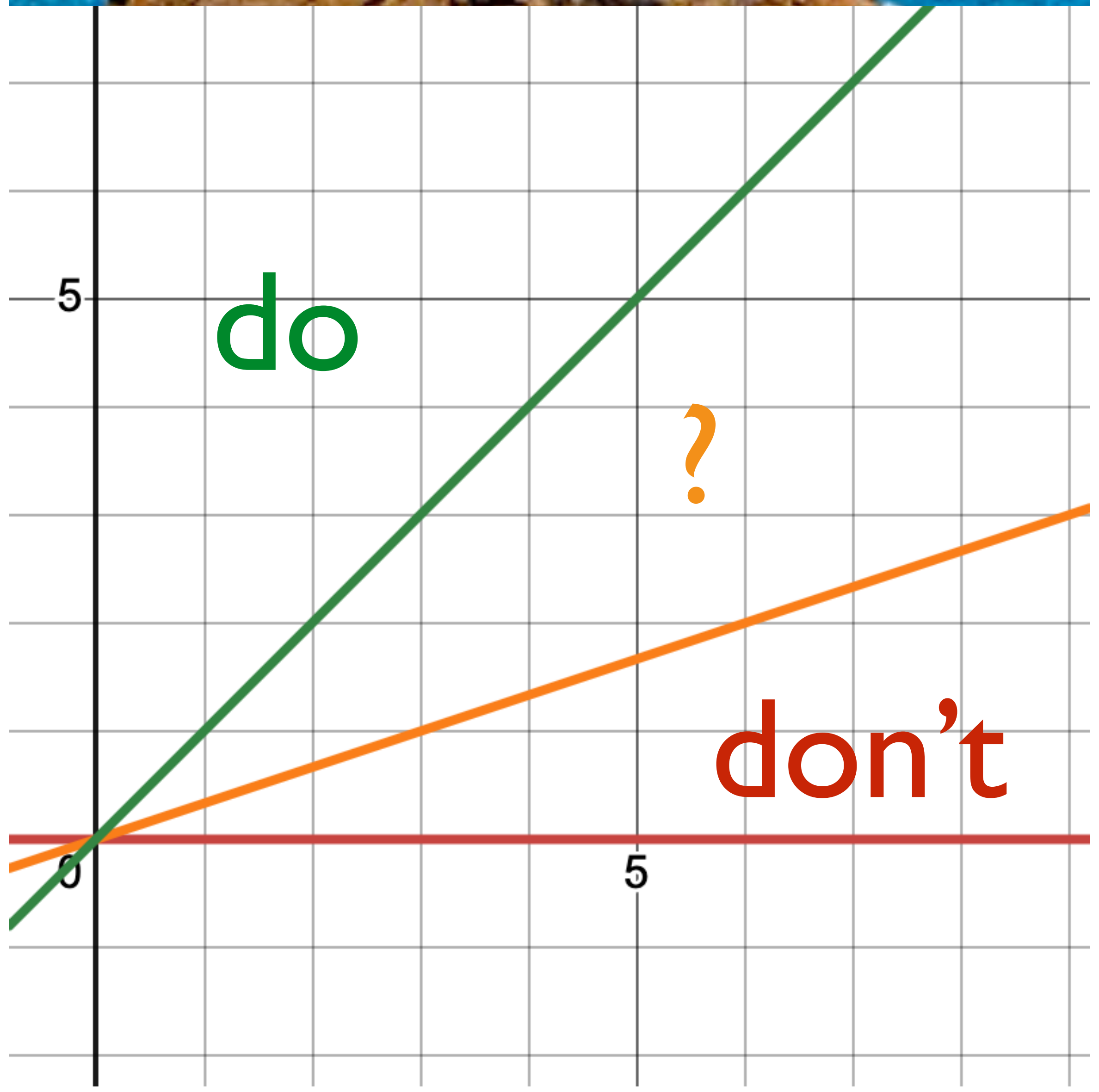
Open Problems in Cooperative AI :-(

Allan Dafoe[1], Edward Hughes[2], Yoram Bachrach[2], Tantum Collins[2], Kevin R. McKee[2], Joel Z. Leibo[2], Kate Larson[2, 3] and Thore Graepel[2]

[1]Centre for the Governance of AI, Future of Humanity Institute, University of Oxford, [2]DeepMind, [3]University of Waterloo

"Since machines powered by artificial intelligence are playing an ever greater role in our lives, it will be important to equip them with the capabilities necessary to cooperate and to foster cooperation."

Public Goods Problems



Public Goods Problems

- Each player owns **c** tokens, and decides whether to invest them (**C**ooperate) or not (**D**efect)
- The return to the total investment is a multipe **f** of the total investment and is evenly divided among players
- **f** modulates how attractive the investment is:
  - **f < 1** … obviously better not to invest
  - **n < f** … obviously better to invest
  - **1 < f < n** … better not to invest, but hopefully others do (dilemma)

Pareto dominated Nash equilibria

Multiplier Factor Games

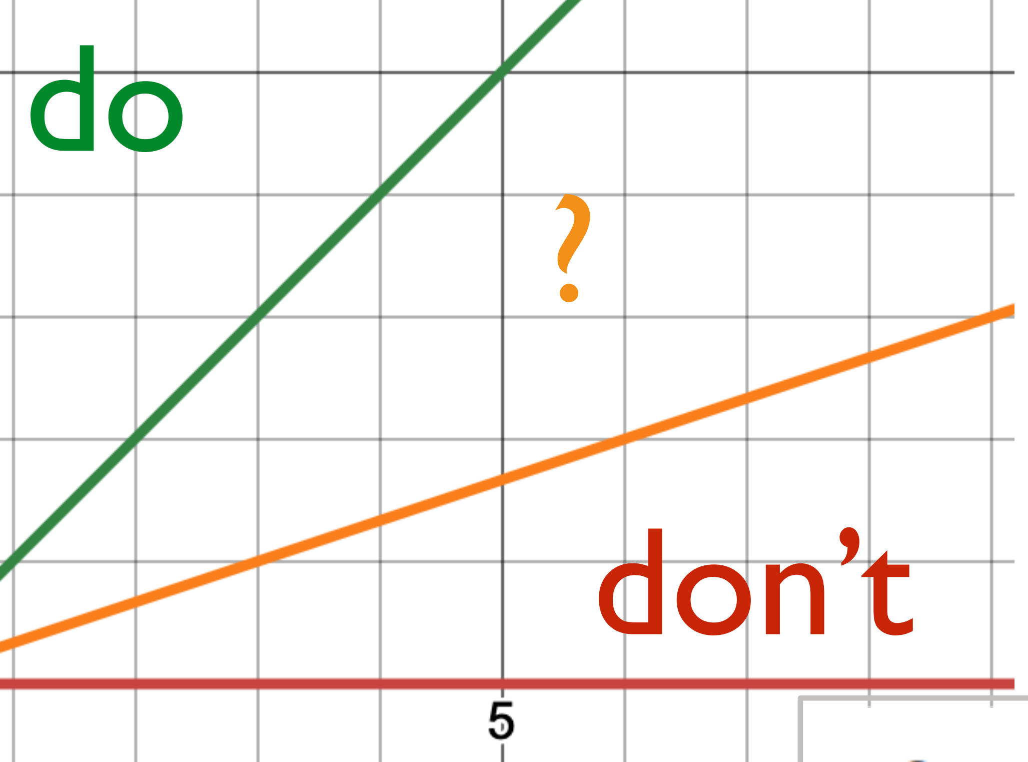
A multiplier factor game is a tuple $\langle N, \mathbf{c}, A, f, \mathbf{r} \rangle$, where:

- ☐ N is the set of players, with $|N| = $ n being the number of players
- ☐ $\boldsymbol{c} = (c_1, \ldots, c_n)$ *with* $c_i \in \mathbb{R}$ is the tuple of endowments
- ☐ $A = \{C, D\}$ is the action set of each player: cooperate (**C**) or defect (**D**)
- ☐ $f \in F \subseteq \mathbb{R}_{\geq 0}$ is the multiplier factor
- ☐ $\boldsymbol{r}$ is the tuple of agents' payoffs

all in! all out!

$$r_i(\boldsymbol{a}, f, \boldsymbol{c}) = \boxed{\frac{1}{n} \sum_{j=1}^{n} c_j I(a_j) \cdot \boxed{f}} + \boxed{c_i(1 - I(a_i))}$$

$$I(a_j) = \begin{cases} 1 & \textit{if } a_j = C \\ 0 & \textit{otherwise.} \end{cases}$$



Possible Games

Extended Public Goods Game

Imperfect knowledge of multiplication factor

$$r_i(\mathbf{a}, f, \mathbf{c}) = \frac{1}{n}\sum_{j=1}^{n} c_j I(a_j) \cdot \boxed{f} + c_i(1 - I(a_i))$$

Competitive

$f = 0.5$

Player Y \ Player X	C	D
C	2, 2	1, 5
D	5, 1	4, 4

Mixed-Motive

$f = 1.5$

Player Y \ Player X	C	D
C	6, 6	3, 7
D	7, 3	4, 4

Cooperative

$f = 2.5$

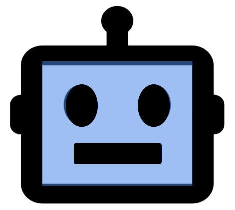
Player Y \ Player X	C	D
C	10, 10	5, 9
D	9, 5	4, 4

$\tilde{f}_X = f + N(0, \sigma_X)$

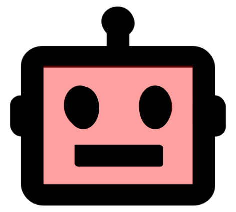
$\tilde{f}_Y = f + N(0, \sigma_Y)$

observed with **uncertainty**

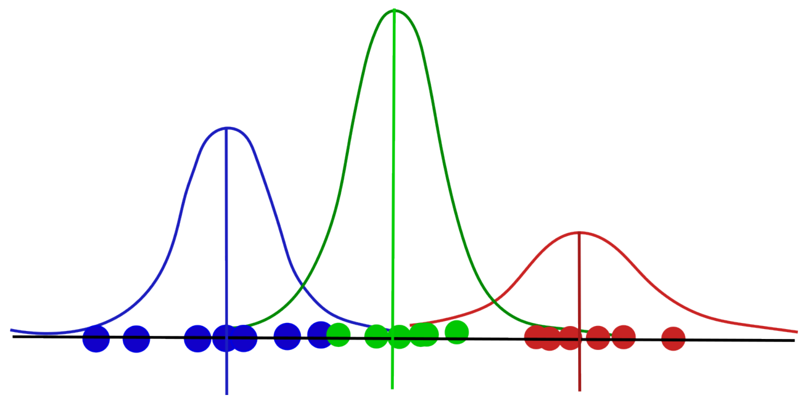
Player X



Player Y



possibly interpreted via a model ***Gaussian Mixture Model***



$a_X = C$

$a_Y = D$

	C	D
C	6, 6	3, 7
D	7, 3	4, 4

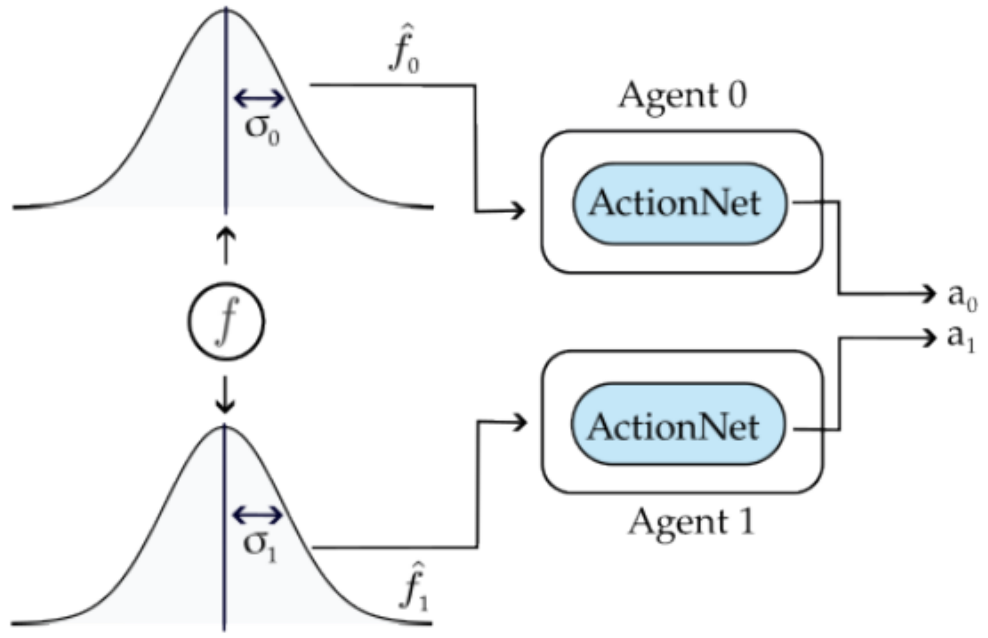
$r_X = 3$

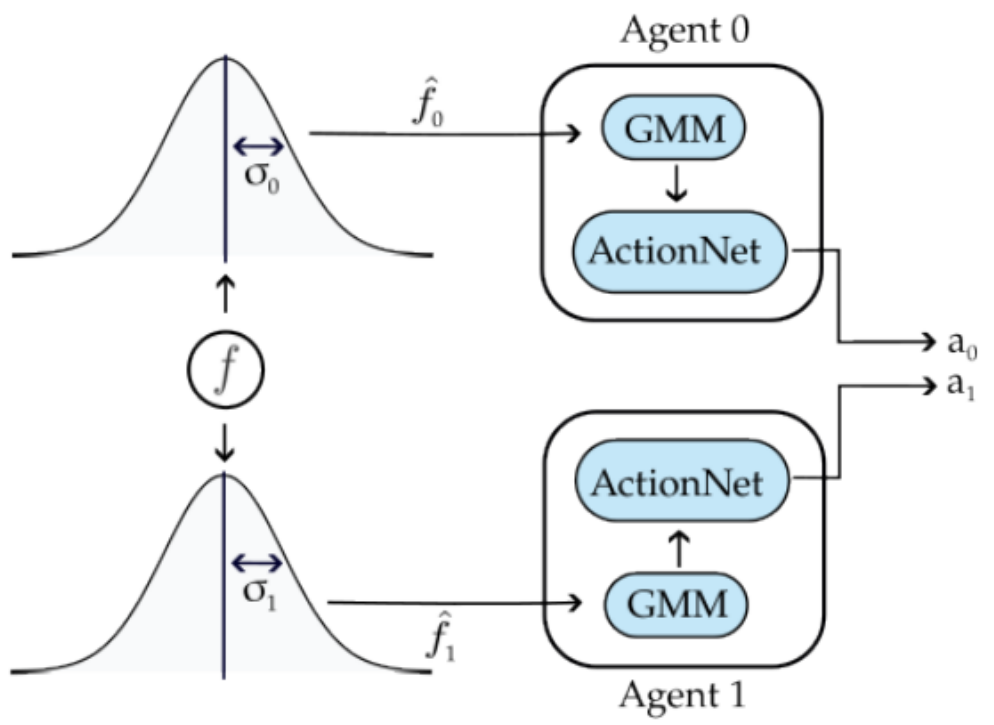
$r_Y = 7$

POSG!

Baseline results with n = 2

- independent learners
- REINFORCE
- 1-shot games

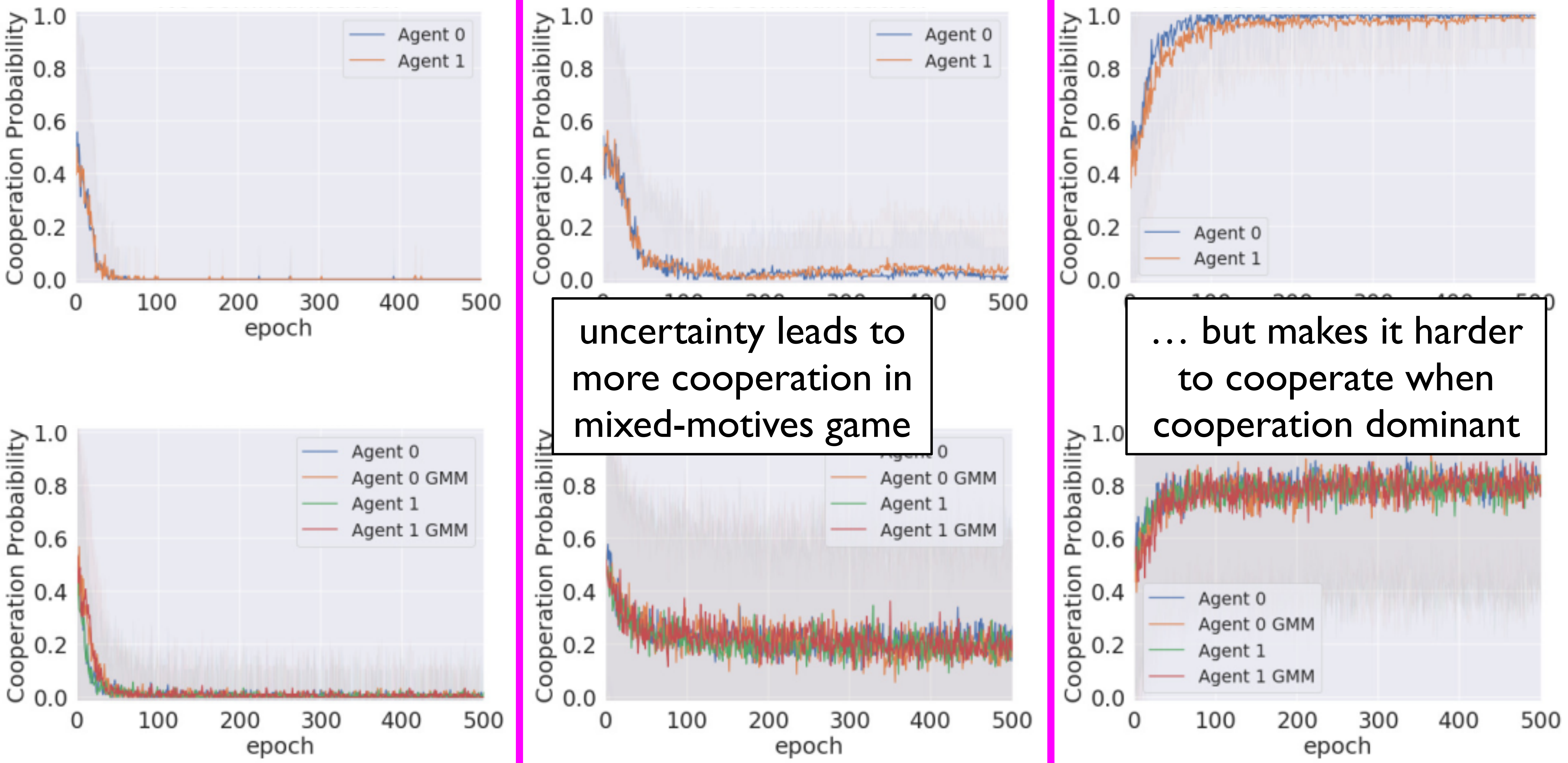




No Uncertainty

Uncertainty

$\sigma_i = 0.5$



uncertainty leads to more cooperation in mixed-motives game

… but makes it harder to cooperate when cooperation dominant

average cooperation rate of across 80 runs

(a) $f = 0.5$ (b) $f = 1.5$ (c) $f = 2.5$

no predefined protocol

no predefined meaning

Communication

uncertainty leads to more cooperation in mixed-motives game

Extended Public Goods Game + Communication

Possible Games

Competitive — $f = 0.5$

Player Y \ Player X	C	D
C	2, 2	1, 5
D	5, 1	4, 4

Mixed-Motive — $f = 1.5$

Player Y \ Player X	C	D
C	6, 6	3, 7
D	7, 3	4, 4

Cooperative — $f = 2.5$

Player Y \ Player X	C	D
C	10, 10	5, 9
D	9, 5	4, 4

$\tilde{f}_X = f + N(0, \sigma_X)$ $\qquad$ $\tilde{f}_Y = f + N(0, \sigma_Y)$

observed with **uncertainty**

$$\pi_{Ci} : O_i \times M_i \rightarrow [0, 1]$$

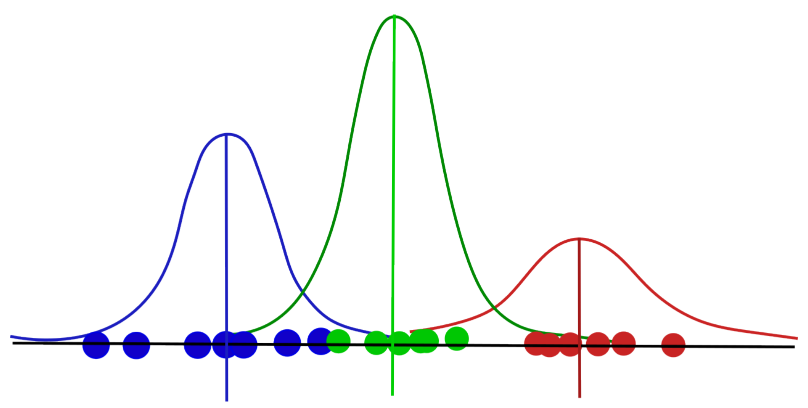
Player X $\xleftrightarrow[m_Y]{m_X}$ Player Y

$a_X = C$ $\qquad$ $a_Y = D$

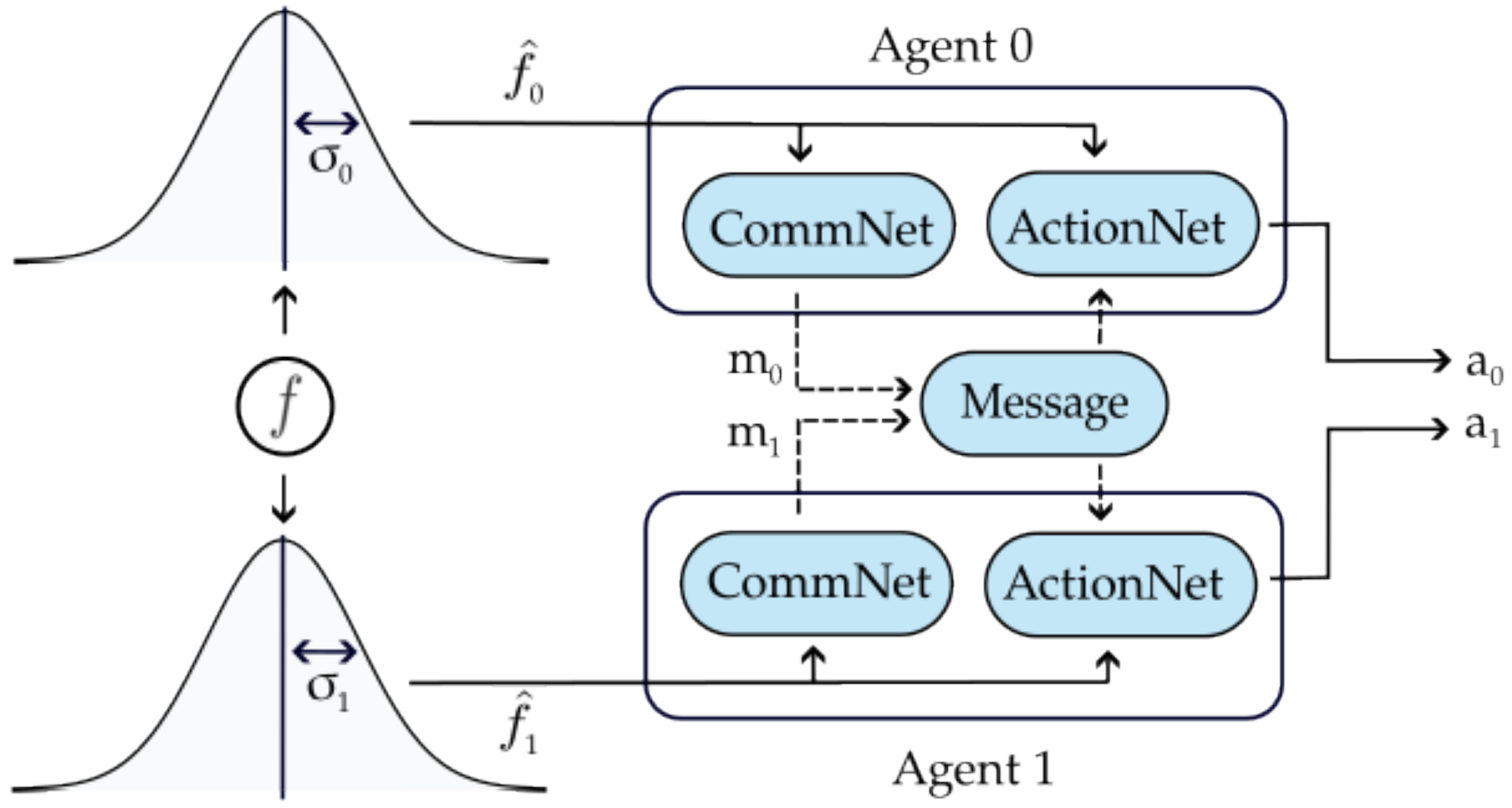
	C	D
C	6, 6	3, 7
D	7, 3	4, 4

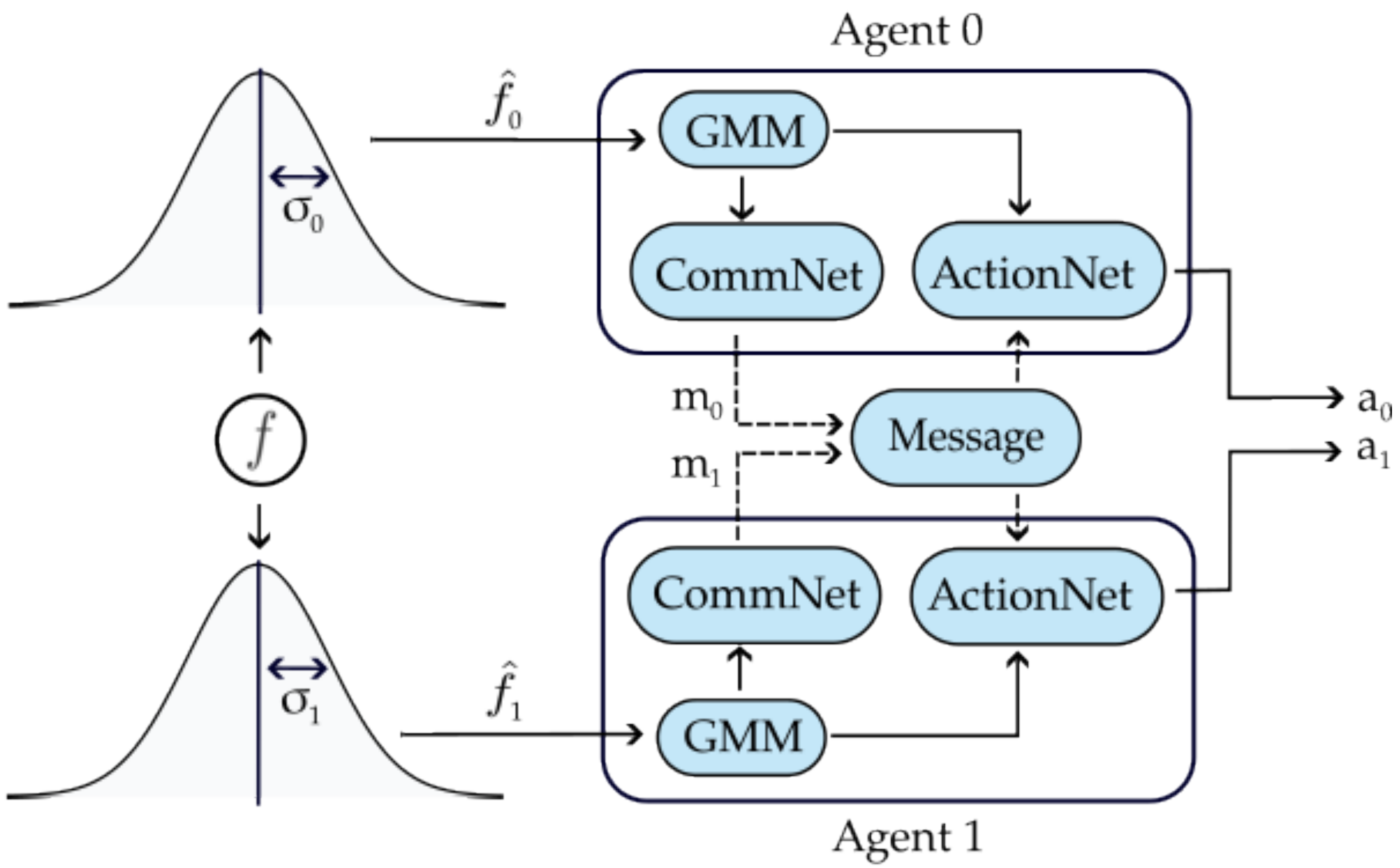
$r_X = 3$ $\qquad$ $r_Y = 7$

possibly interpreted via a model ***Gaussian Mixture Model***

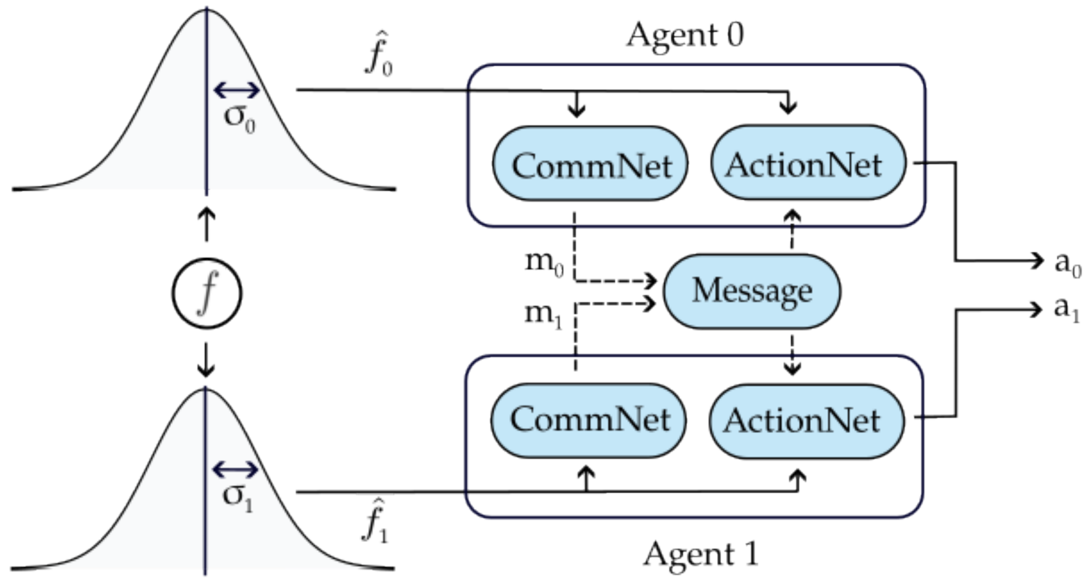


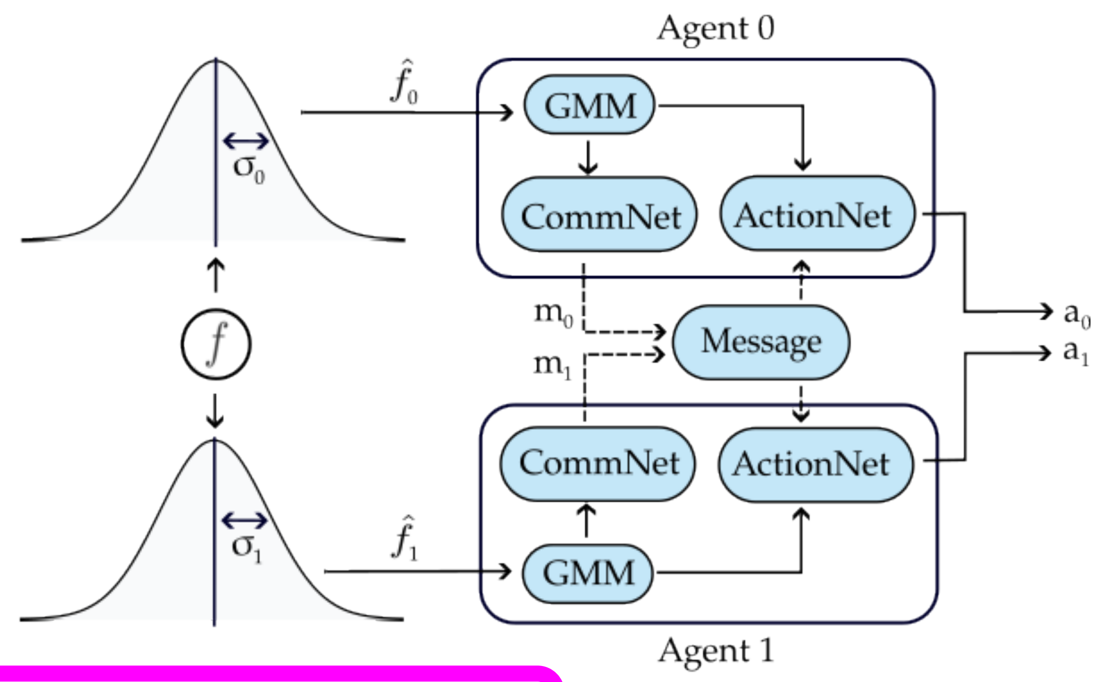
Two Uncertain Agents



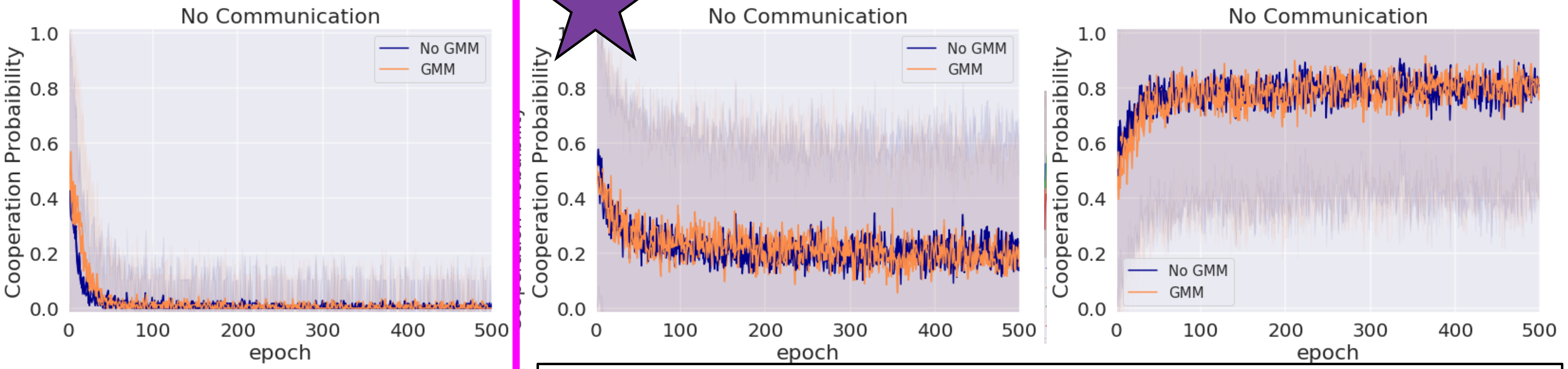


Two Uncertain Agents ($\sigma_i = 0.5$)



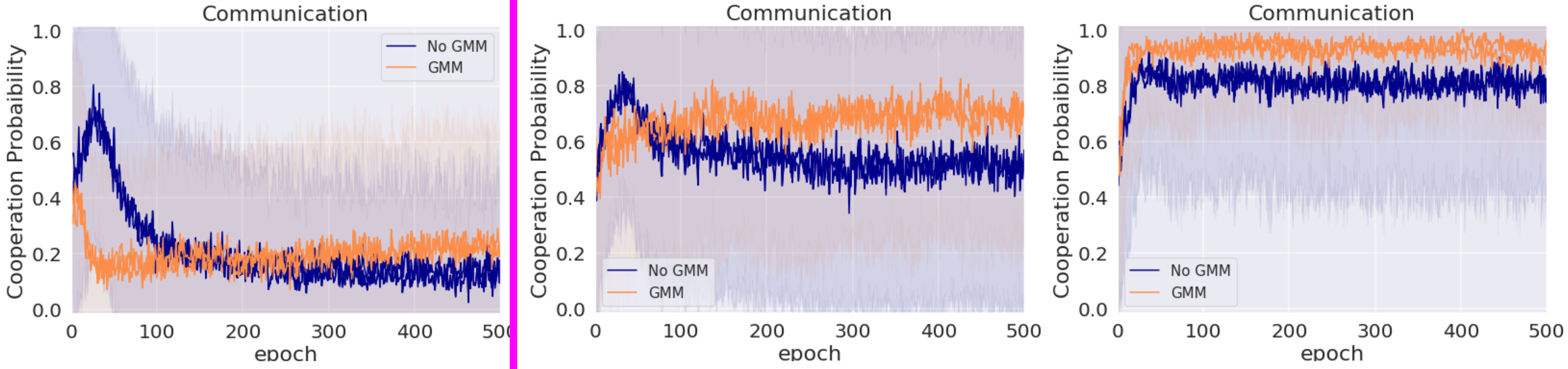


No Communication



uncertainty + communication leads to more cooperation across mixed-motives and cooperative games

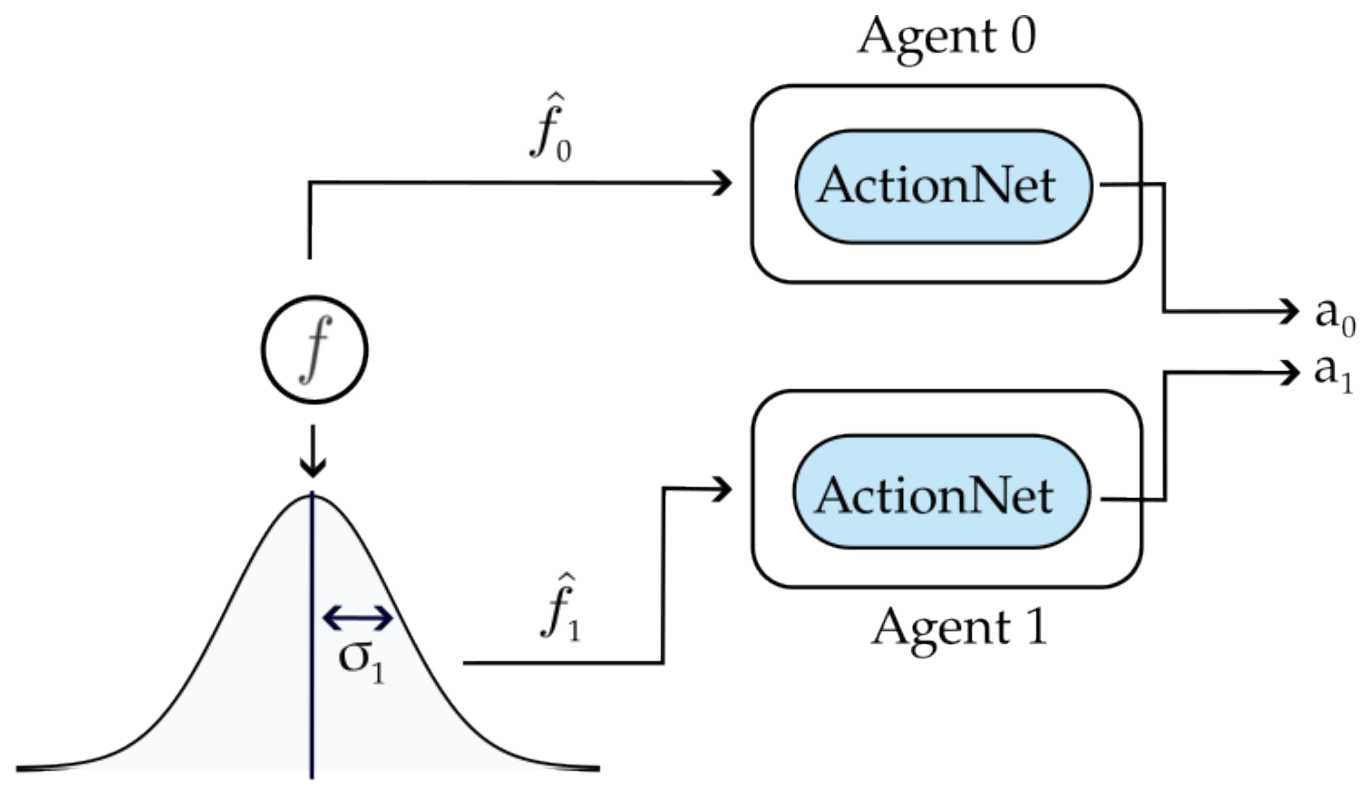
Communication



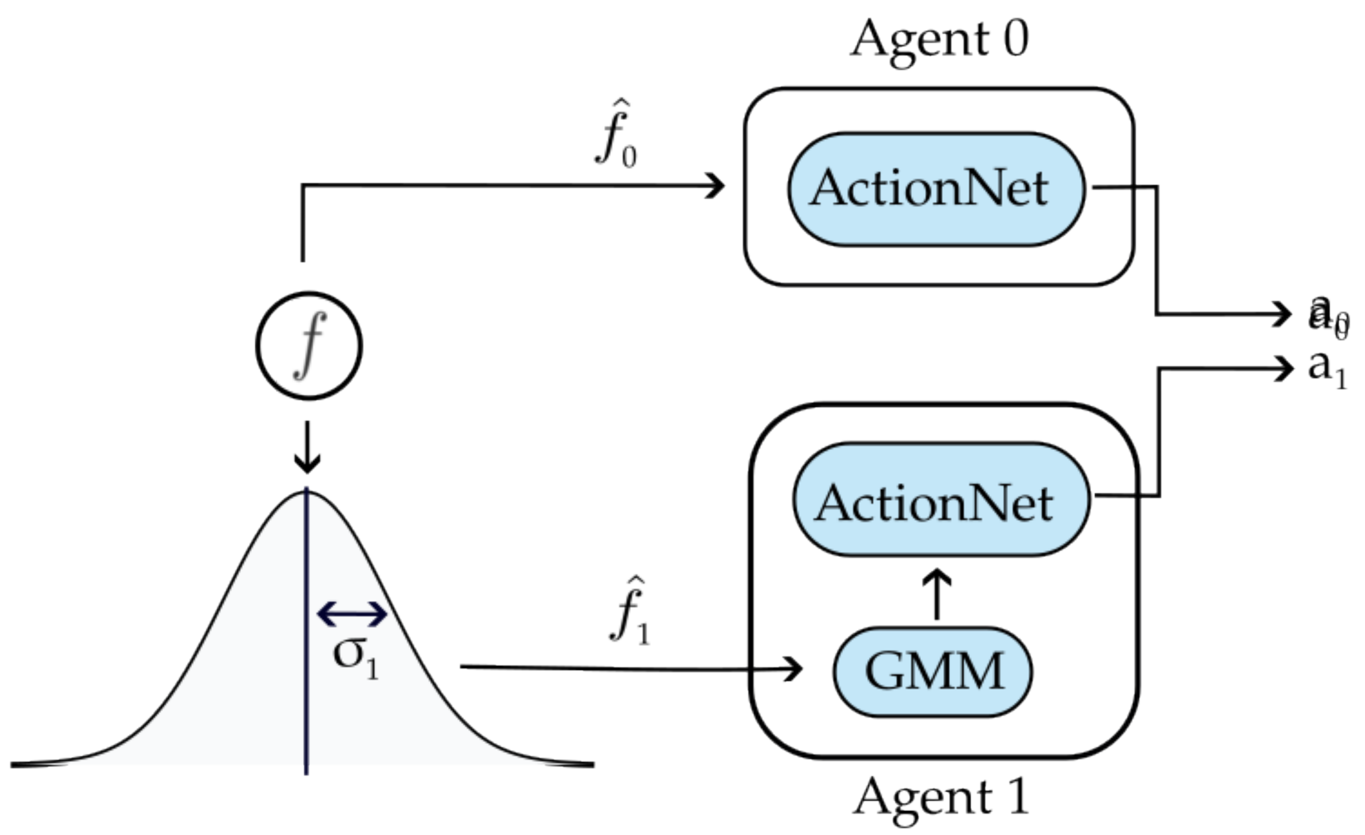
(a) $f = 0.5$ (b) $f = 1.5$ (c) $f = 2.5$

average cooperation rate across 80 runs

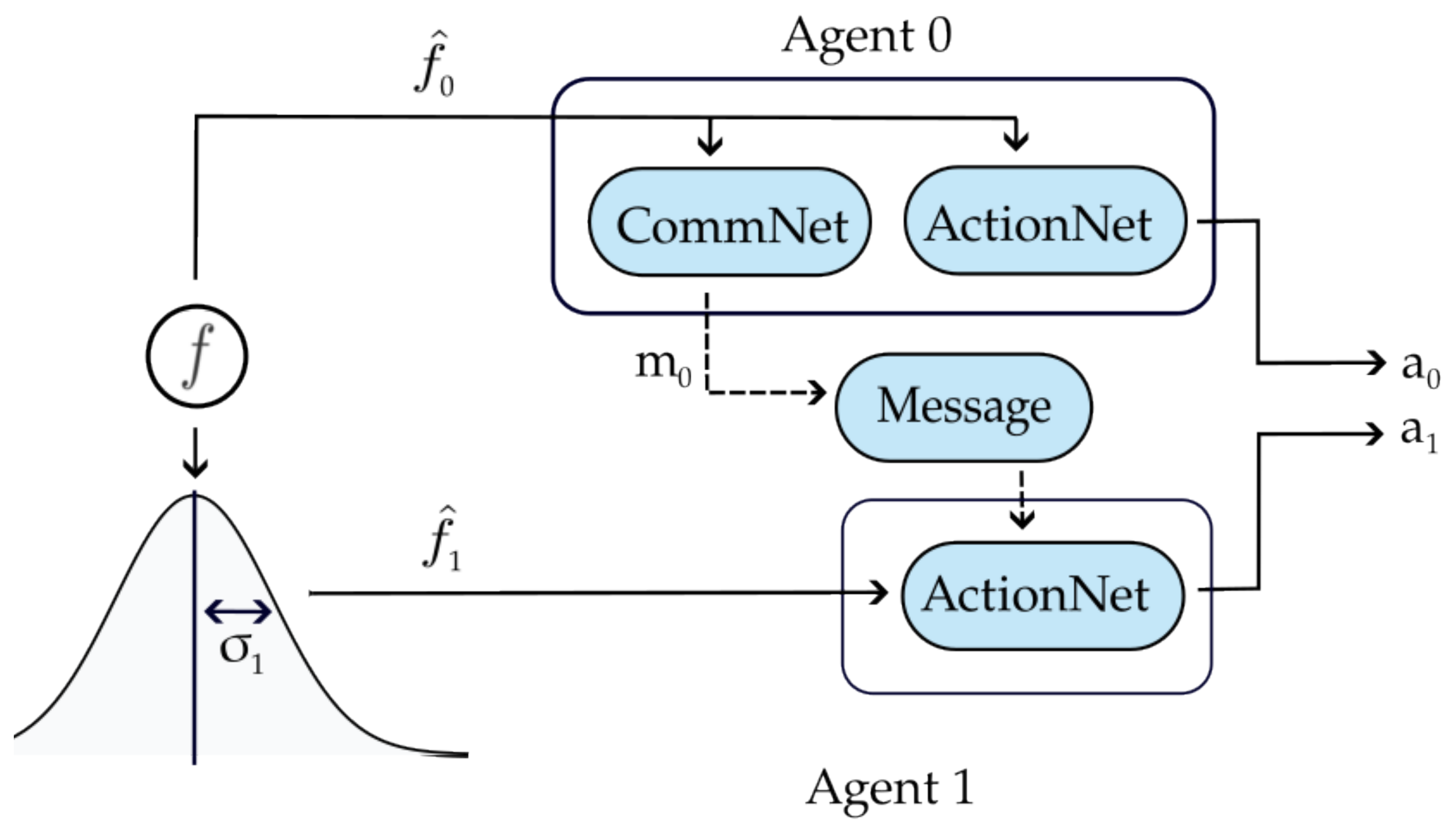
One Uncertain Agent



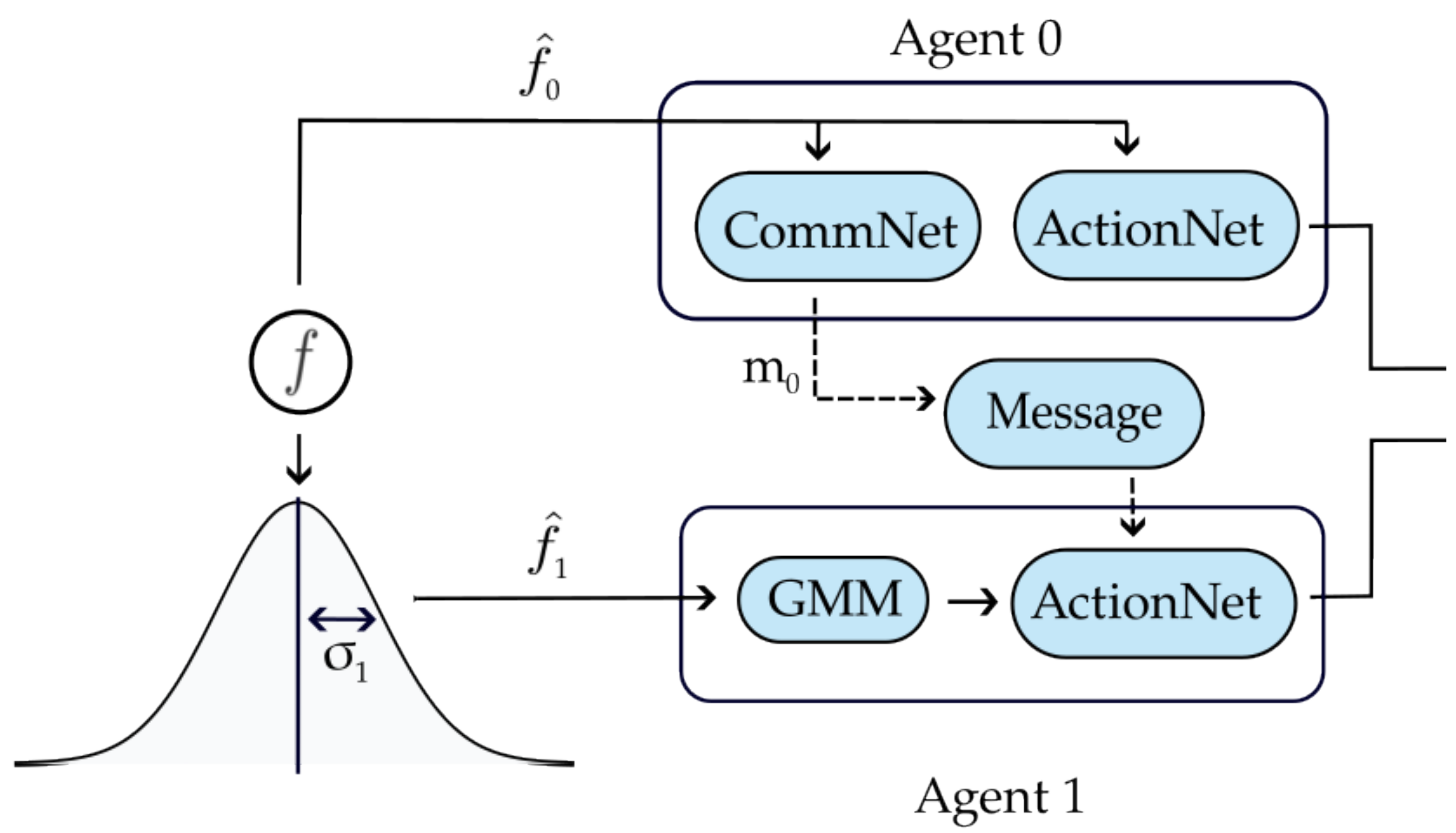
(a)



(b)



(c)



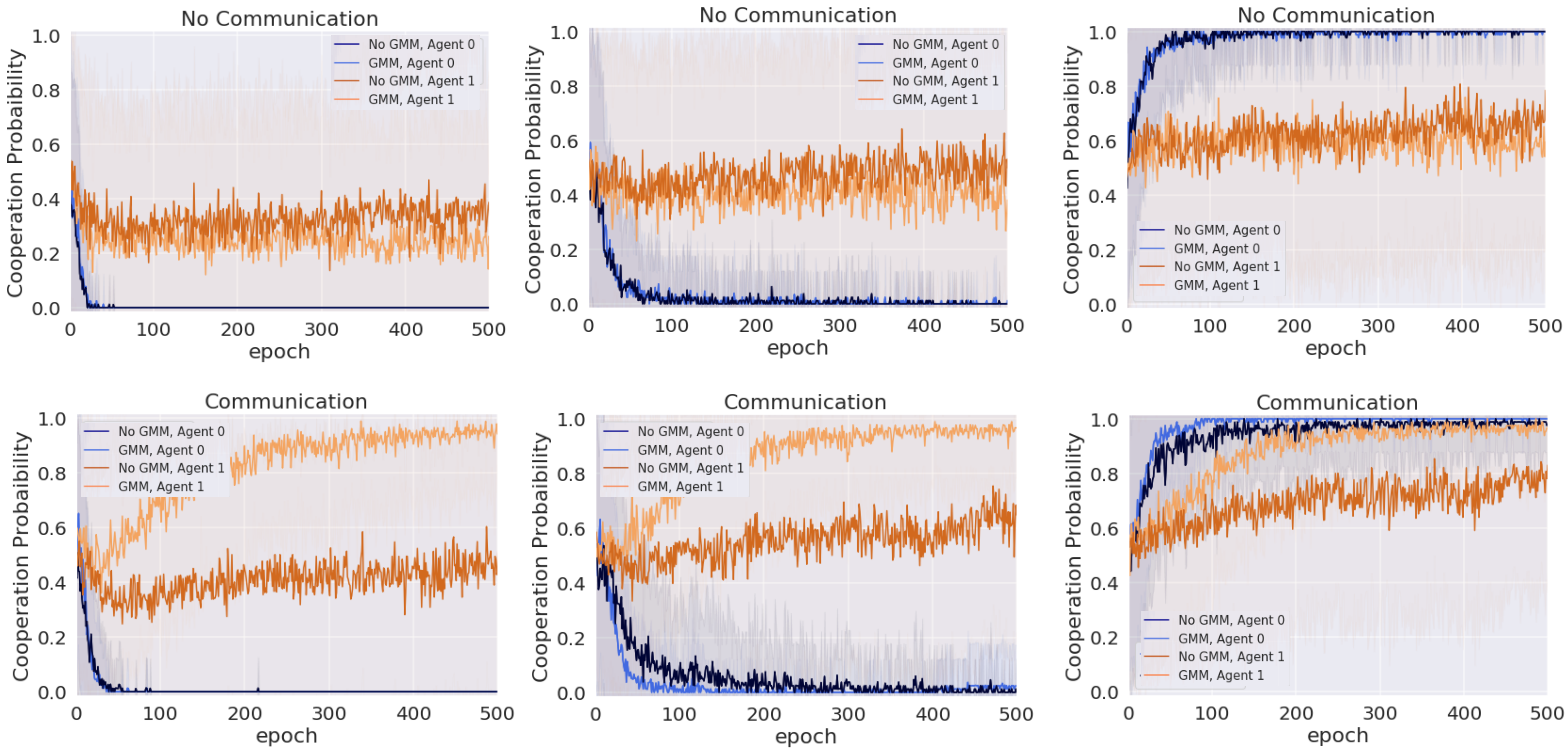
(d)

One Uncertain Agent ($\sigma_i = 2$)

The uncertain agent is brought to cooperate more across the board

No Communication

Communication



$= 0.5$

(b) $f = 1.5$

Communication increases the probability of cooperation of the uncertain agent (deception)

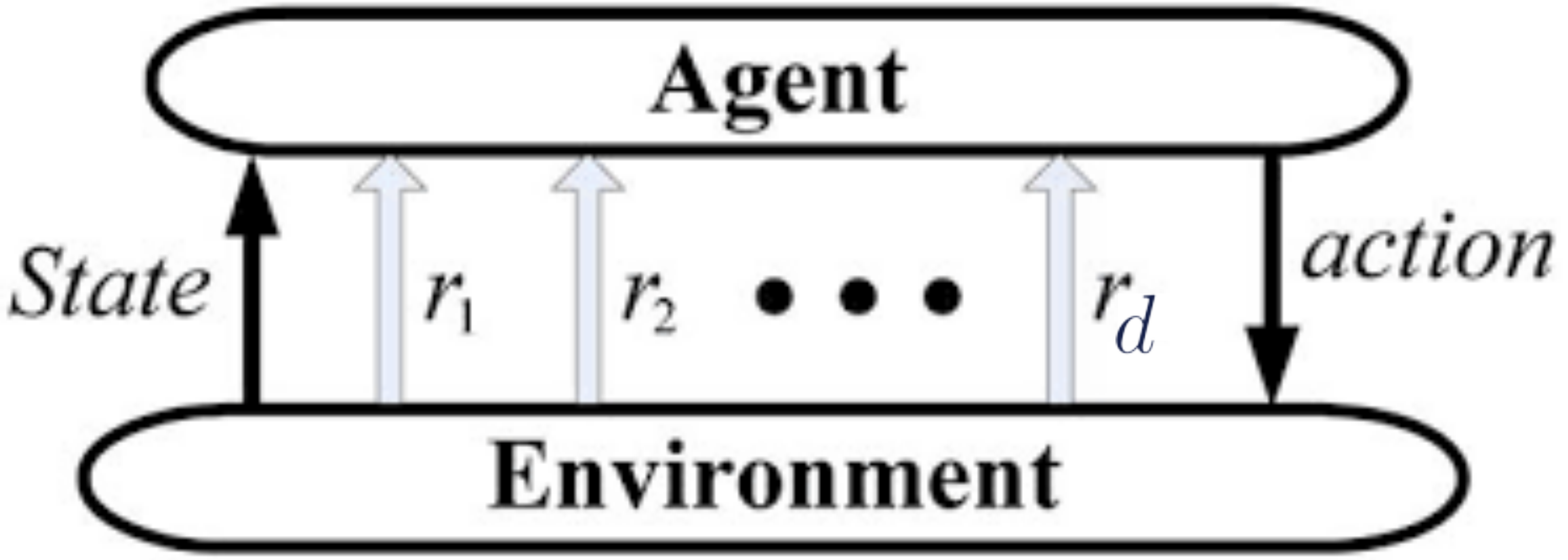
low speaker consistency

Modeling uncertainty helps the uncertain agent resist deception but not recovering enough information

Risk attitudes

uncertainty leads to
more cooperation in
mixed-motives game

Multi-Objective Reinforcement Learning



Defined by the tuple $< S, A, T, \gamma, \boldsymbol{R} >$ where:

- S set of states available to the agent
- A set of actions available to the agent
- T transition function (dynamics of the environment)
- $\boldsymbol{R}$ vectorial reward function $\mathbf{R}: S \times A \times S \rightarrow \mathbb{R}^d$ $\quad \mathbf{r}_{k+1} = \mathbf{R}(s_k, a_k, s_{k+1})$
- $\gamma \in [0, 1]$ discount factor

Multi-Objective Reinforcement Learning

Utility function maps a **vector reward to a scalar utility**:

$$V_u^\pi = u\left(\mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \gamma^t \boldsymbol{r}_t\right]\right)$$ **SER** Optimization Criterion

$$V_u^\pi = \mathbb{E}_\pi\left[u\left(\sum_{t=0}^{\infty} \gamma^t \boldsymbol{r}_t\right)\right]$$ **ESR** Optimization Criterion

Defined by the tuple $< S, A, T, \gamma, \boldsymbol{R} >$ where:

- S set of states available to the agent
- A set of actions available to the agent
- T transition function (dynamics of the environment)
- $\boldsymbol{R}$ vectorial reward function $\boxed{\mathbf{R} \colon S \times A \times S \to \mathbb{R}^d}$ $\boxed{\mathbf{r}_{k+1} = \mathbf{R}(s_k, a_k, s_{k+1})}$
- $\gamma \in [0, 1]$ discount factor

Flashback Multiplier Factor Games

A multiplier factor game is a tuple $\langle N, \boldsymbol{c}, A, f, \boldsymbol{r} \rangle$, where:

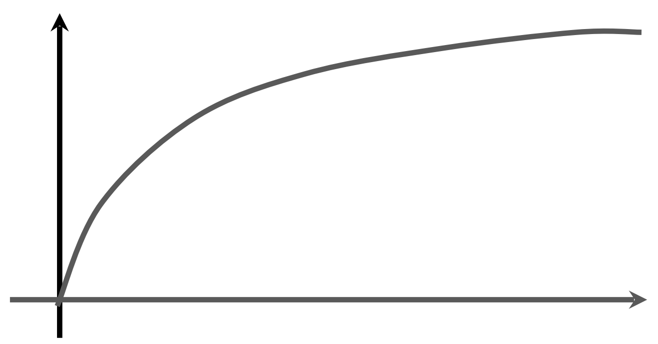
- ☐ N is the set of players, with $|N|$ = n being the number of players
- ☐ $\boldsymbol{c} = (c_1, \ldots, c_n)$ *with* $c_i \in \mathbb{R}$ is the tuple of endowments
- ☐ $A = \{C, D\}$ is the action set of each player: cooperate (**C**) or defect (**D**)
- ☐ $f \in \mathbb{R}_{\geq 0}$ is the (investment) multiplier factor
- ☐ $\boldsymbol{r}$ is the tuple of agents' payoffs

$$[\ r_i^C(\boldsymbol{a}, f, \boldsymbol{c})\ ,\ r_i^I(\boldsymbol{a}, \boldsymbol{c})\ ]$$

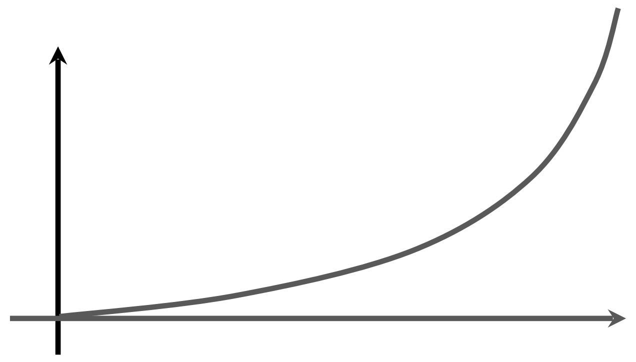
$$r_i(\boldsymbol{a}, f, \boldsymbol{c}) = \underbrace{\boxed{\frac{1}{n}\sum_{j=1}^{n} c_j I(a_j) \cdot \boxed{f}}}_{\textbf{collective return}} + \underbrace{\boxed{c_i(1 - I(a_i))}}_{\textbf{individual return}}$$

MO Multiplier Factor Games

- **Risk aversion**: prefer a sure outcome over a lottery whose expected payoff may be higher than the outcome payoff
- **Risk propensity**: prefer a lottery over a sure outcome whose outcome may be higher than the expected payoff of the lottery.
- **Risk neutrality**: Indifference between lotteries and sure outcomes



- $0 < \beta < 1$, the function is concave (risk-aversion)
- $\beta > 1$, the function is convex (risk-propension)



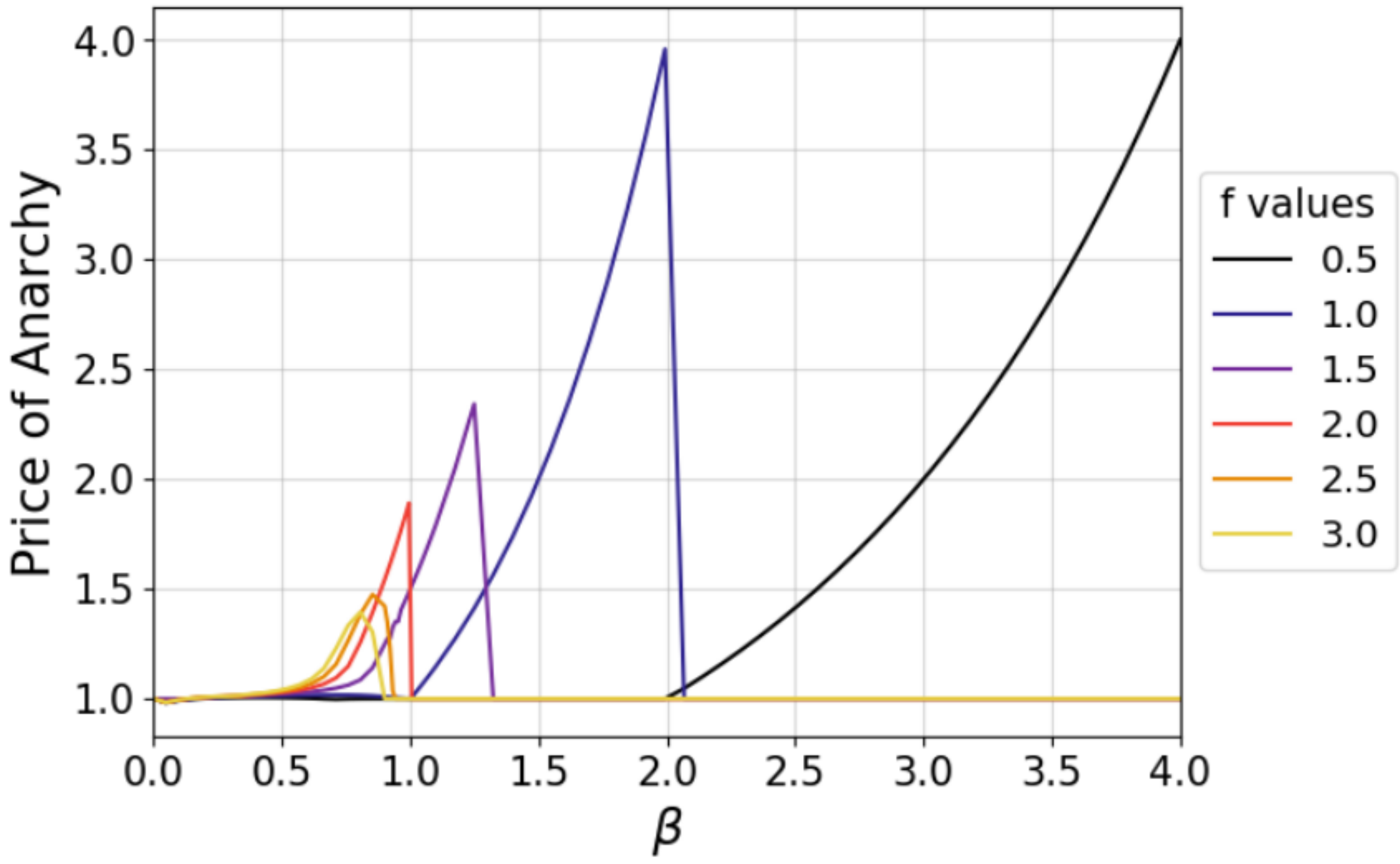
$$\left[\ r_i^C(\boldsymbol{a}, f, \mathbf{c})^{\beta},\ r_i^I(\boldsymbol{a}, \mathbf{c})\ \right]$$

$$\text{SER}\left(r_i(\boldsymbol{a}, f, \mathbf{c}) = \underbrace{\boxed{\frac{1}{n}\sum_{j=1}^{n} c_j I(a_j) \cdot \boxed{f}}}_{\text{collective return}} + \underbrace{\boxed{c_i(1 - I(a_i))}}_{\text{individual return}} \right)$$

MO Multiplier Factor Games

$$\sum_{i=0}^{n} u_i(\boldsymbol{V}_i^{\boldsymbol{\pi}})$$

$$PoA = \frac{\max_s W(\boldsymbol{\pi})}{\min_{\boldsymbol{\pi} \in \text{Nash}} W(\boldsymbol{\pi})}$$



$$\left[\; r_i^C(\boldsymbol{a}, f, \boldsymbol{c})^{\beta}, \;\; r_i^I(\boldsymbol{a}, \boldsymbol{c}) \;\right]$$

$$\text{SER}\left(r_i(\boldsymbol{a}, f, \boldsymbol{c}) = \underbrace{\boxed{\frac{1}{n}\sum_{j=1}^{n} c_j I(a_j) \cdot \boxed{f}}}_{\text{collective return}} + \underbrace{\boxed{c_i(1 - I(a_i))}}_{\text{individual return}} \right)$$

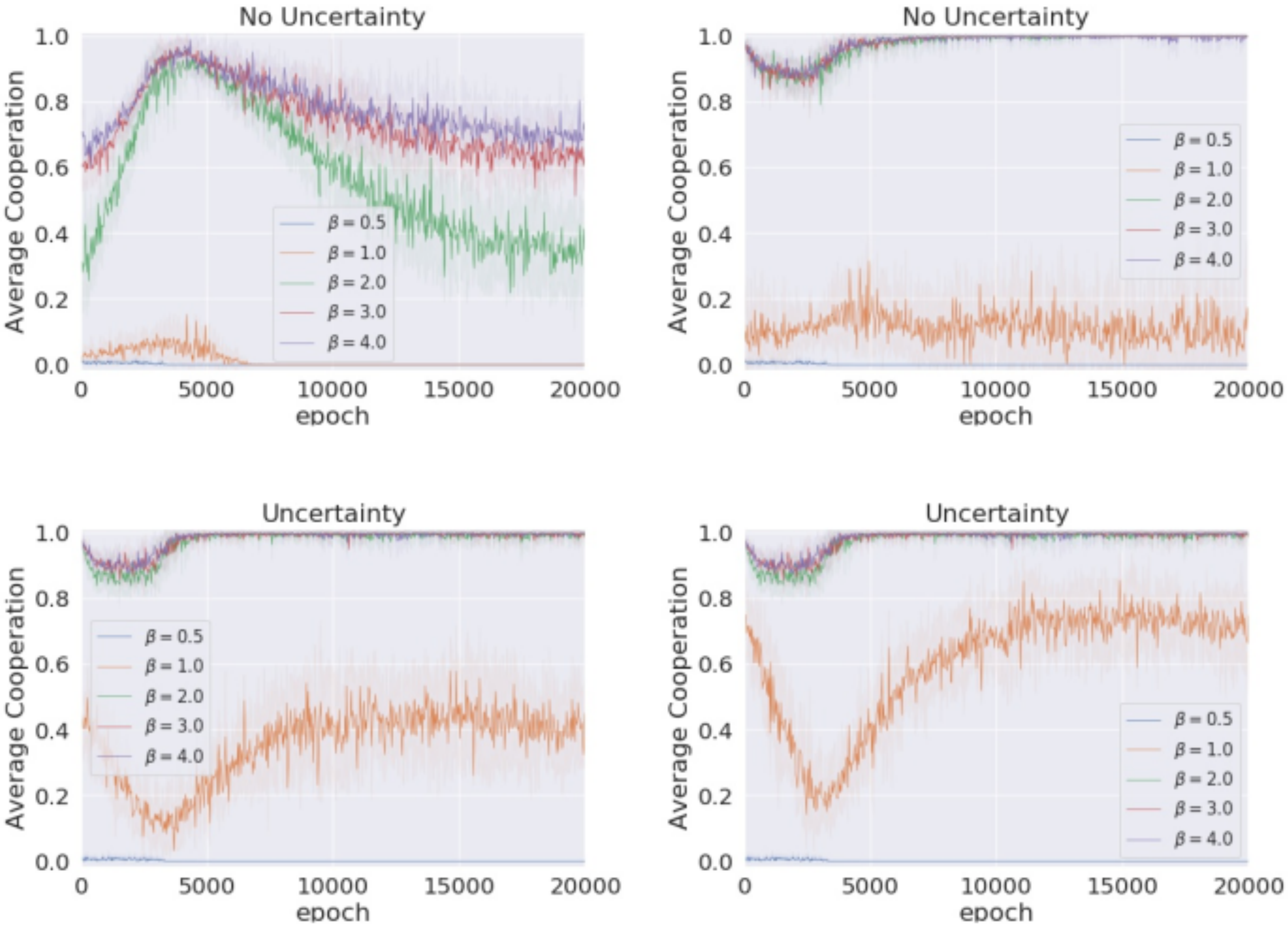
Experimental Setup

$$V_u^{\pi} = u\left(\mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t \boldsymbol{r}_t\right]\right)$$

SER Optimization Criterion

- Population of N=20 agents
- M=4 agents sampled at each epoch, playing for 10 rounds
- f in $[f_{\text{min}}, f_{\text{max}}]$, with $f_{\text{min}} = 0.5$ and $f_{\text{max}} = 6.5$
- Agents implemented as MO-DQN
- 2000 epochs

Learning with homogeneous preferences



(a) $f = 0.5$ (b) $f = 1.5$

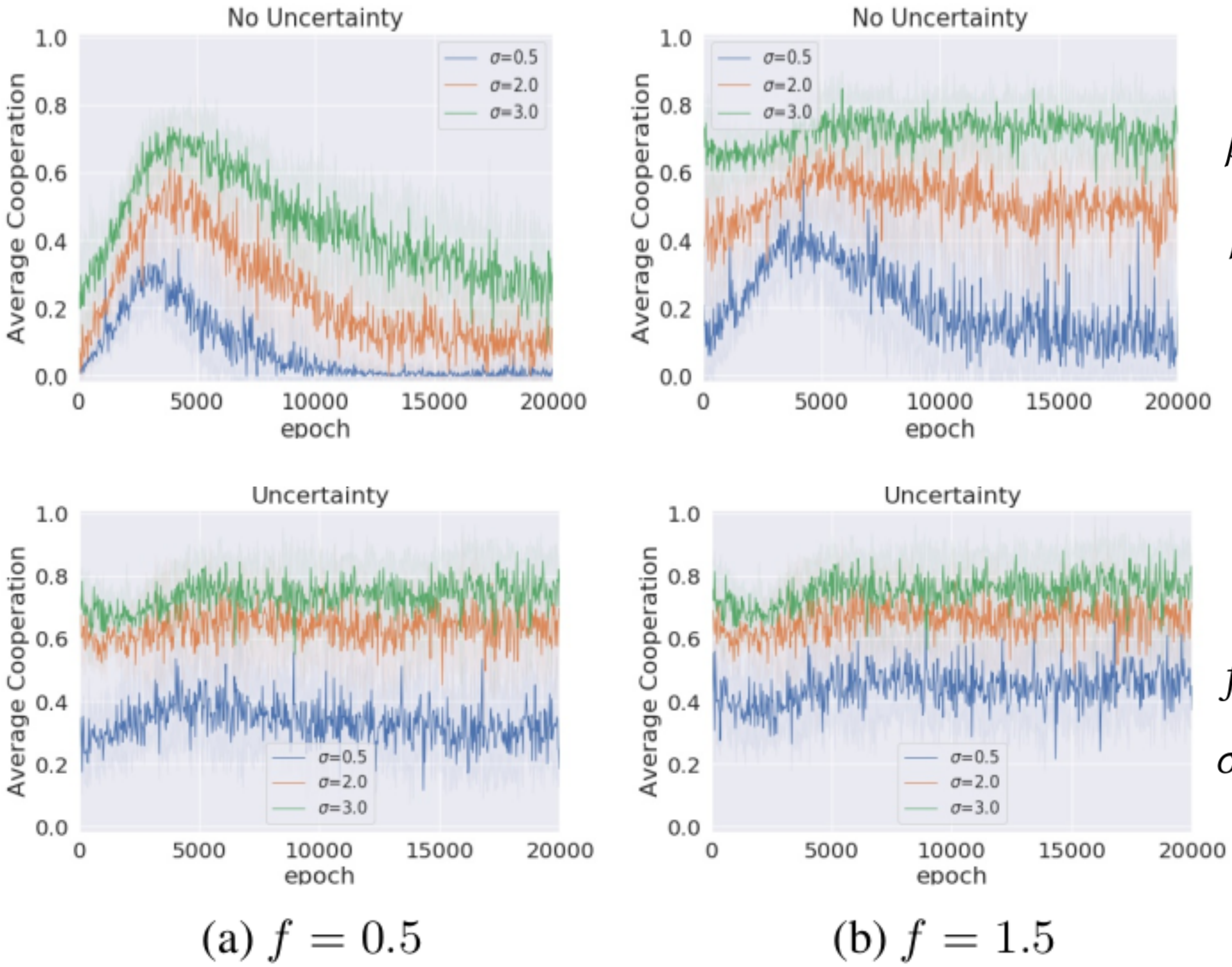
$$f^i_{obs} = f + \mathcal{N}(0, \sigma_i^2)$$

$$\sigma_i = 2 \ \forall \ i \in N$$

uncertainty leads to more cooperation across the board for risk-neutral and risk-seeking agents

for active agents, over last 50 epochs, over 20 runs

Learning with heterogeneous preferences



$\beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2) \ \forall\ i \in N$

$\mu_\beta = 1$

uncertainty leads to more cooperation across the board

the higher the risk propensity the better

$f_{obs}^i = f + \mathcal{N}(0, \sigma_i^2)$

$\sigma_i = 2 \ \forall\ i \in N$

but dominated choices are played with fairly high probability in the competitive game

for active agents, over last 50 epochs, over 20 runs

Conclusions

- ☐ Simple environment for MARL when agents are uncertain about the game they play
- ☐ Uncertainty appears to have a positive impact on levels of cooperation
- ☐ … with potential drawbacks (deception, dominated choices)

- ☐ All results are empirical, it would be desirable to have a more fundamental understanding of the exact conditions under which uncertainty positively impact cooperation, the extent of such improvement, and its tradeoffs

Nash equilibria under **SER**