# ESSAI-2024
# Self-Governing Multi-Agent Systems
# L7/10: Self-Regulation

**Jeremy Pitt** and **Asimina Mertzani**

Department of Electrical and Electronic Engineering

Imperial College London

# Aims and Objectives

- Aims
  - learn basic concepts of cybernetics, focusing on first order cybernetics, requisite influence and requisite social influence
  - learn basic concepts of psycho-acoustics
  - learn basic learning mechanisms
  - introduction to hybrid systems
- Objectives
  - identify the ways that concepts from cybernetics, psycho-acoustics and Q-learning can be brought together to achieve ethical self-regulation of a multi-agent system

# Self-Regulated Systems

A **self-regulated system** comprises a designated agent (or agency) acting as **regulator** by operating on some control variables, and a **regulated system** which applies changes in those variables (Ashby, 2020).

Note: You can perceive a self-regulated system as one unit, which corresponds to a system of systems.

# Cybernetics

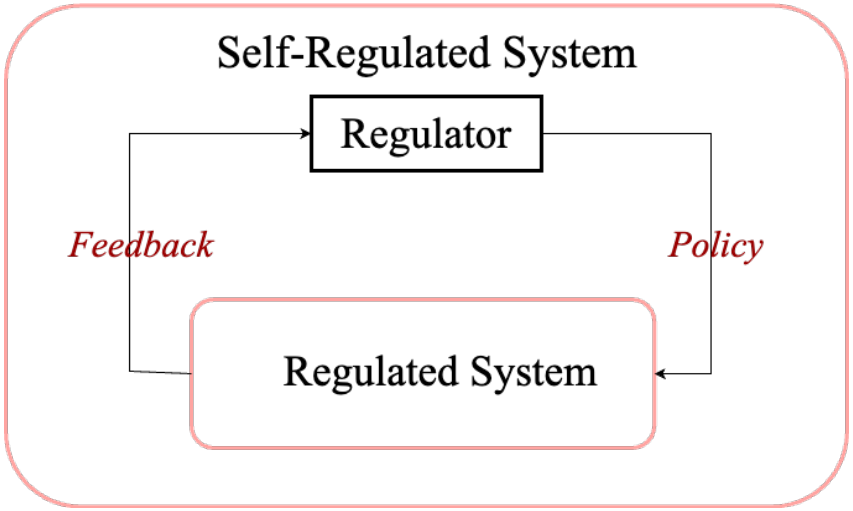
# Cybernetics

Cybernetics is the transdisciplinary study of **circular processes** such as feedback systems where outputs are also inputs.

# First Order Cybernetics

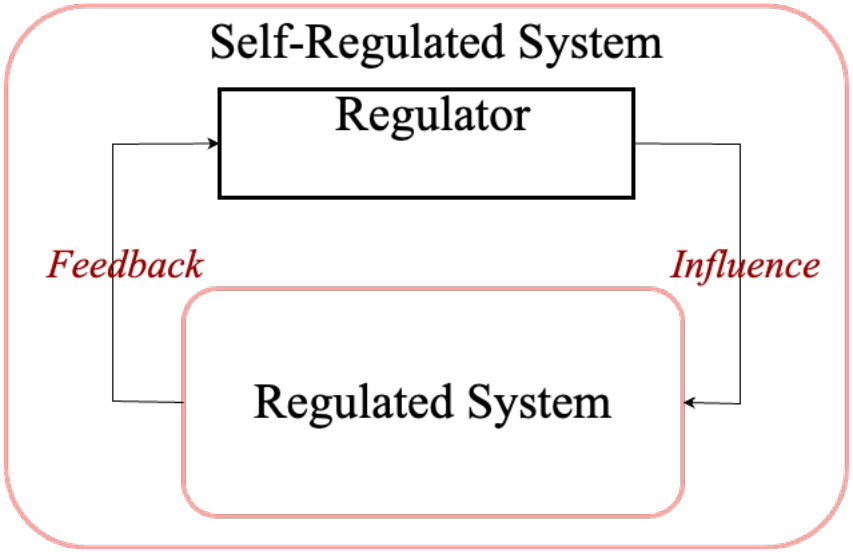
First order cybernetics are described as the **cybernetics of observed systems** (Chepin, 2021), and refer to self-regulated systems having an **input**, an **output**, and a **negative feedback** from the output back to the input (Ashby, 1952).

They are **closed systems** designed to be isolated from their environment, and they are considered as black-box mechanisms that ignore the role of the observer.



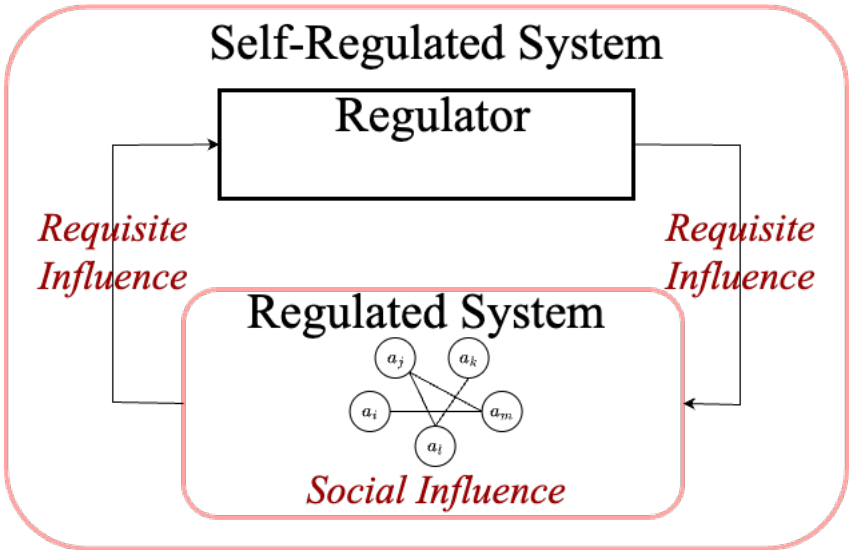
# Requisite Influence

To get the regulator with the regulated unit to **interoperate**, the system should have **requisite influence**, that is the existence of the **pathways to transmit the effects** of the selected actions to the **regulated system**.



# Requisite Social Influence

However, if the regulated system is a multi-agent system that is changing, the system requires **requisite social influence** (Mertzani, 2024).



# So what?

How can we use those concepts of cybernetics in SGMAS?

Or, how can we use those concepts to achieve the self-regulation of a socio-technical system?

Or, in other worlds, how can we achieve requisite social influence in a socio-technical system?

# Opinion Formation and Influence

From previous lecture we know that people **form judgements** and opinions based on the **opinion of experts**.

# Opinion Formation and Influence

However opinion of experts is **not the only one** that affects the opinions of individuals....

# Other Kinds of Influence

# Psycho-Acoustics

- 'expert voice': based on 'trusted' expertise (Horne, 2016)

# Psycho-Acoustics

- 'expert voice': based on 'trusted' expertise (Horne, 2016)
- 'own voice': based on direct personal experience (Fernyhough, 2017)

# Psycho-Acoustics

- 'expert voice': based on 'trusted' expertise (Horne, 2016)
- 'own voice': based on direct personal experience (Fernyhough, 2017)
- 'foreground noise': from social network (Nowak, 2019)

# Psycho-Acoustics: The 4 Voices

- 'expert voice': based on 'trusted' expertise (Horne, 2016)
- 'own voice': based on direct personal experience (Fernyhough, 2017)
- 'foreground noise': from social network (Nowak, 2019)
- 'background noise': from the community or culture in which the individual is embedded (Deutsch, 1955)

# Psycho-Acoustics: The 4 Voices and Their Use

- 'expert voice': based on 'trusted' expertise (Horne, 2016)
- 'own voice': based on direct personal experience (Fernyhough, 2017)
- 'foreground noise': from social network (Nowak, 2019)
- 'background noise': from the community or culture in which the individual is embedded (Deutsch, 1955)

So, we can use those concepts to inform the decision making of the individuals in the **regulated unit**.

And the collective decision might be used to **affect the regulator's** policy.

# Psycho-Acoustics and 4 Voices

**Which one to attend** to affect the behaviour of the regulator in the desired way?

# Psycho-Acoustics and 4 Voices

**Which one to attend** to affect the behaviour of the regulator in the desired way?

And how can the **regulator** identify how to change based on that expression of the collective (e.g. regulated unit)?

# The Problem

- the individuals in the **regulated unit** need to choose **which voice to attend** to form their individual expression
- the individual expressions are aggregated and form the **collective expression**
- the **regulator** should identify how to change based on that **collective expression**

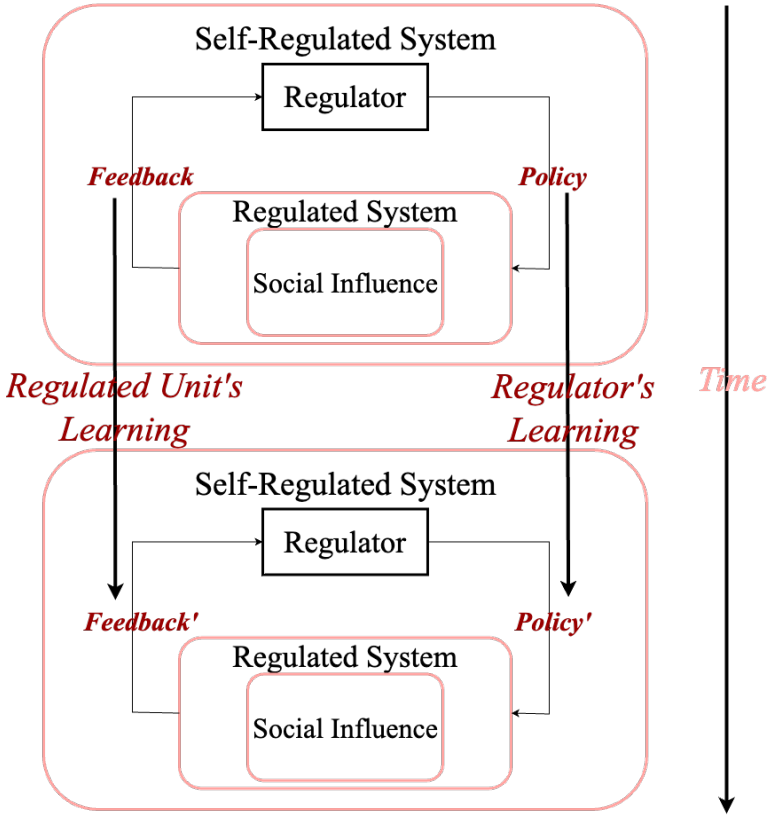
# How?

- the individuals in the **regulated unit** need to choose **which voice to attend** to form their individual expression
- the individual expressions are aggregated and form the **collective expression**
- the **regulator** should identify how to change based on that **collective expression**

Learning...?

# Learning Mechanisms in the Model

- Learning to attend the 4 voices (learning of the regulated unit)
- Learning to attend the regulated unit (learning of the regulator)



# Regulated Unit's Learning

**Aim:** distinguish between the for voices and choose the one that would change the regulator's policy in the desired way

**Method:** use **reinforcement coefficients** to incentivise the following behaviours (the three strategies that we investigate):

- attend the voice that maximises individual satisfaction in the short-term
- attend the voice that resembles the experts voice
- attend the voice that resembles the average noise from the network

**Reinforcement Coefficient's Update:** $cr_{voice} = cr_{voice} * (1 \pm c)$ evaluated with respect to the change they observed in the regulator's policy

# Regulator's Learning

**Aim:** enable the regulator to use the feedback from the regulated unit in a constructive way and change the policy in a way that it maximises the collective satisfaction

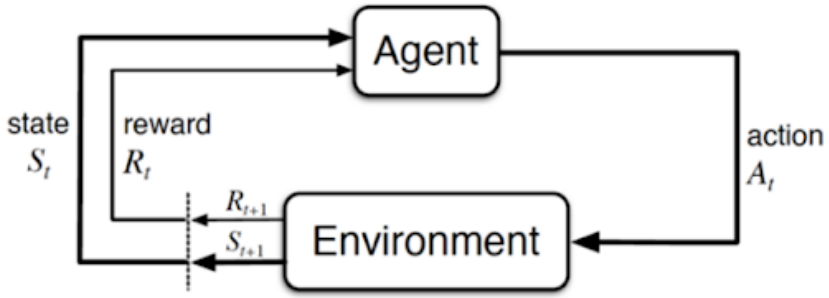
**Method:** Q-learning, is a Reinforcement Learning (RL) a learning process can be modeled as a Markov Decision Process (MDP).

# Regulator's Learning - Reinforcement Learning

The MDP uses the Markov Property, which states that **the future can be determined only from the present state** that encapsulates all the necessary information from the past.

An MDP is characterised by the following components:

- S: state $s \in S$
- A: action $a \in A$
- $P(s_{t+1}|s_t, a_t)$: transition probabilities
- $R(s)$: reward of the state
- $\gamma$: discount factor (to balance our short-term and long-term rewards)



# Regulator's Learning - Q-learning

The way that an agent chooses an action based on its current state is called policy $\pi$.
To learn to select the action that would **maximise the future reward** the agent uses an action-value function $Q(s_t, a_t)$ that is updated at each time step based on:

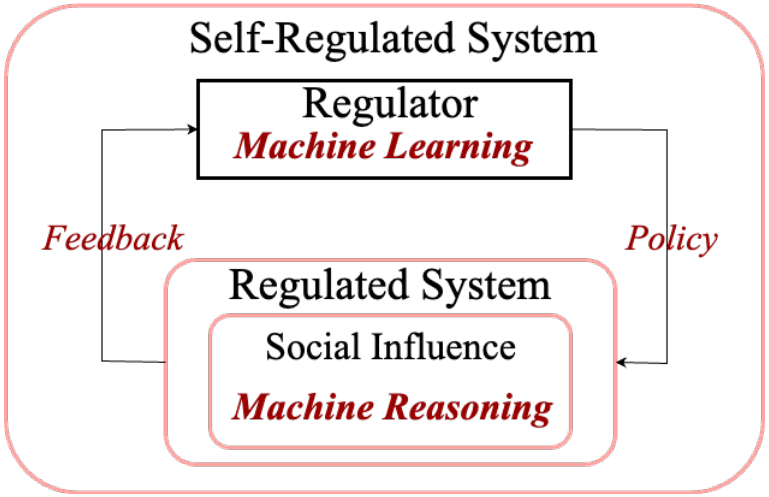
$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_t + \gamma max Q(s_{t+1}, a_t) - Q(s_t, a_t))$$

In this case:

- S: collective expression
- A: options of rule

# Hybrid System

- Learning of the regulated unit and 4 voices (machine reasoning)
- Learning of the regulator (machine learning)



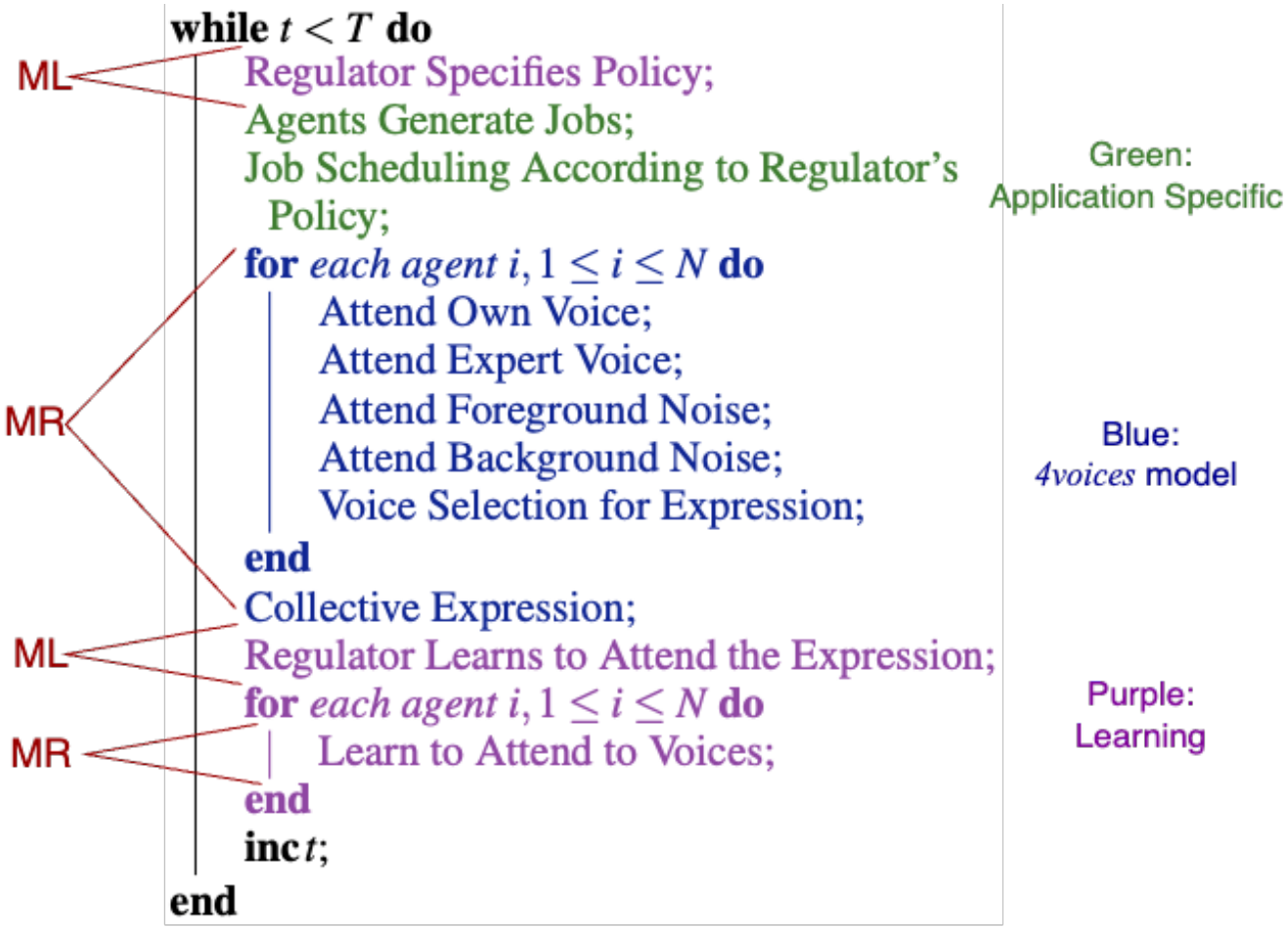
# Hybrid System - Our Definition

Hybrid system is a system comprising a **machine learning** (ML) and a **machine reasoning** (MR) component.

- The machine learning component can be a classification algorithm, a reinforcement learning algorithm, etc.
- The machine reasoning component might be an algorithm designed according to a theory of social or political sciences, or a method using reinforcement coefficients, or even a cybernetics mechanism.

In other words, it's a **cyclic two-phase process**, iterating over an ML and an MR component (or vice versa).

# The 4voices Algorithm



# Experimental Results - Baseline Behaviour

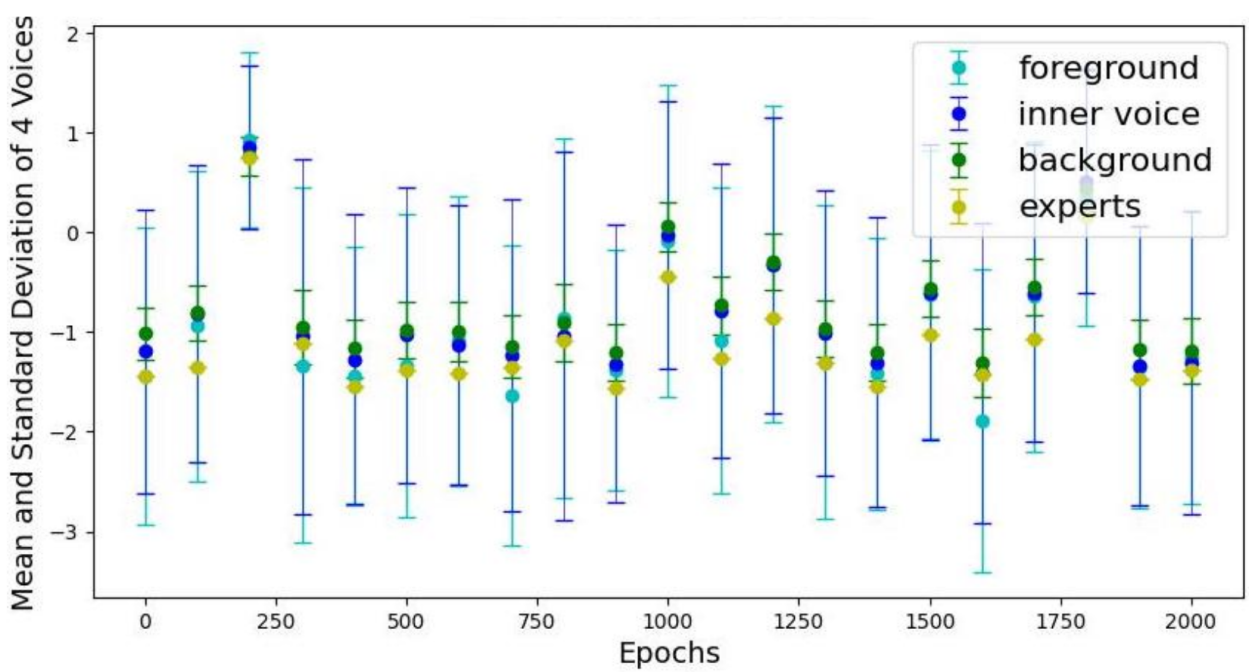



Figure: Diversity Expressed by Each Voice

- **mean value** of the voices is **similar** while the **standard deviation** of the voices is **not**
- **own voice's and foreground noise's standard deviation** is comparatively **higher** than the background noise, while the experts' voice standard deviation is zero since all experts agree on the same value
- trade-off between **diversity** of opinions and **congruence**

# Reminder of Learning...

**Regulated Unit's Learning:**

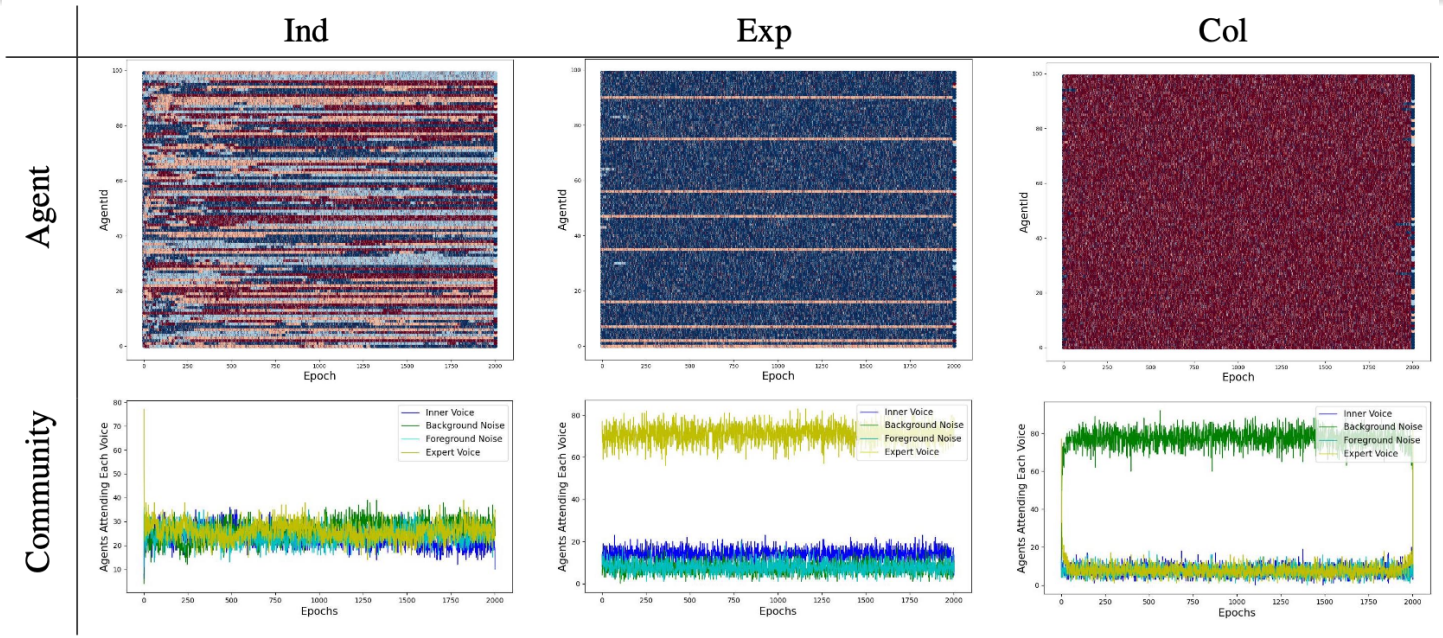
The three strategies that we investigate:

- attend the voice that maximises individual satisfaction in the short-term (Ind)
- attend the voice that resembles the experts voice (Exp)
- attend the voice that resembles the average noise from the network (Col)

**Regulator's Learning:**

- No learning **(Random)**
- Reinforcement Learning **(RL)**

# Experimental Results - Type of Truth




Voice Selected Individually and Collectively for Different Experimental Conditions

- they are divided between different voices ⇒ individual truth based on preferences (Ind)
- they learn to listen to the experts ⇒ informational/ground truth (Exp)
- they learn to listen to the background noise (formed by selecting a random sample of the population) ⇒ community truth (Col)
- Overall, the 4voices model enables the agents to identify **different forms of expertise** and 'types of truth'

# Experimental Results - Pathways for Requisite Influence

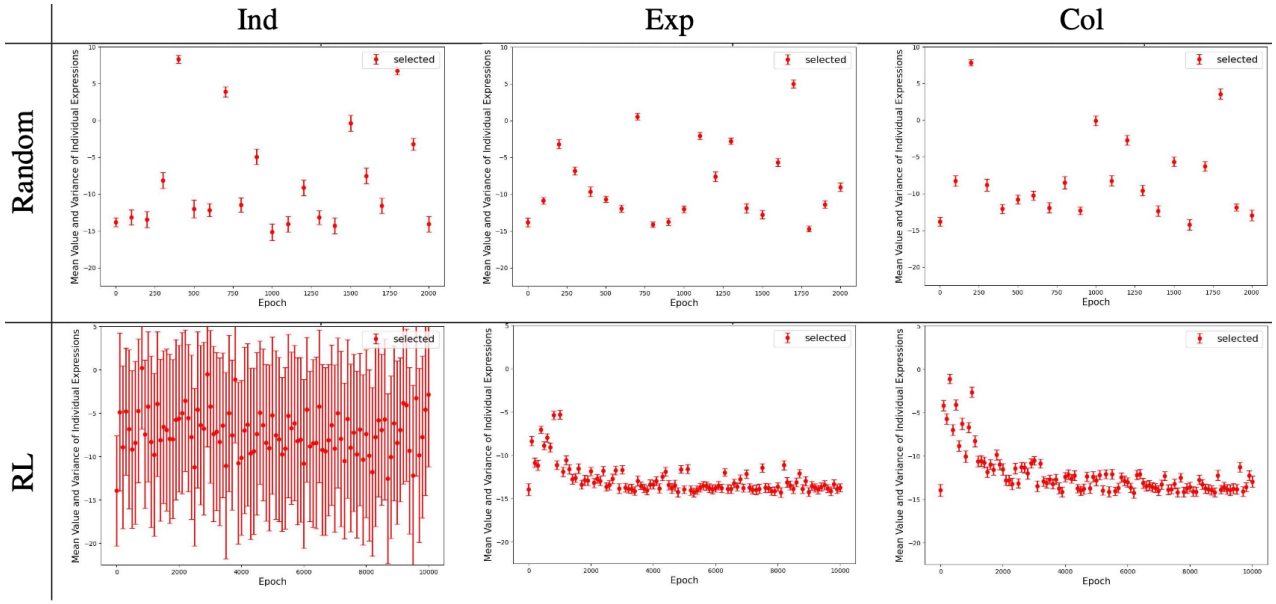



Figure: Mean and st.deviation of individual expressions (Ind. Perspective)

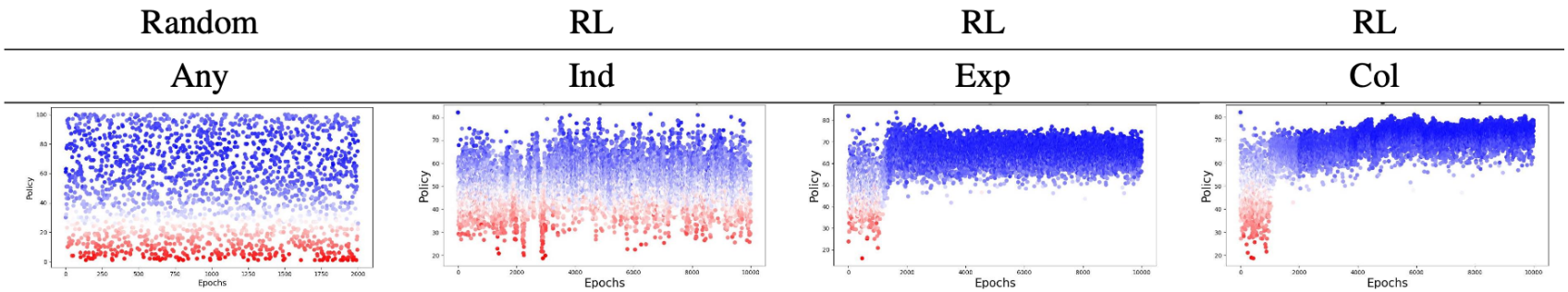



Figure: Policy Selected & Collective Expression (Collective Perspective)

# Experimental Results - Pathways for Requisite Influence

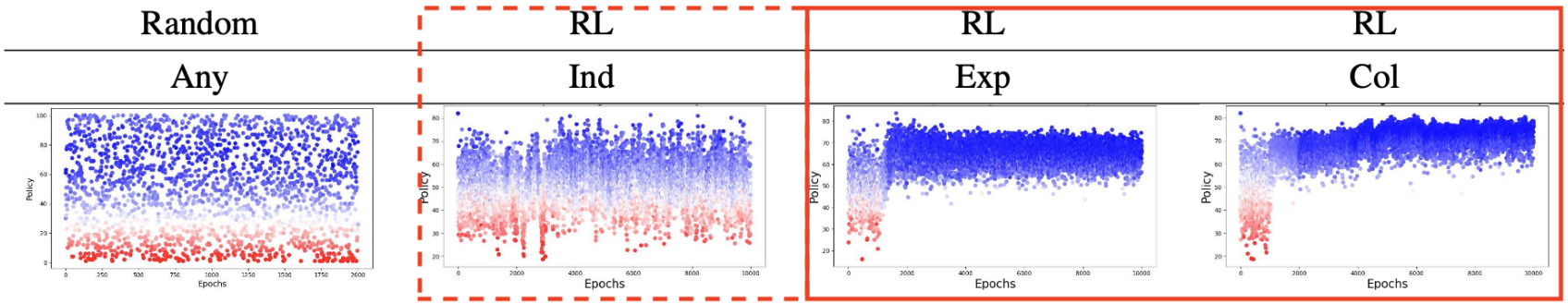



Figure: Policy Selected & Collective Expression (Collective Perspective)

- The pathways for requisite influence from the regulator to the regulated unit are established when regulator is using RL.

# Experimental Results - Pathways for Requisite Influence

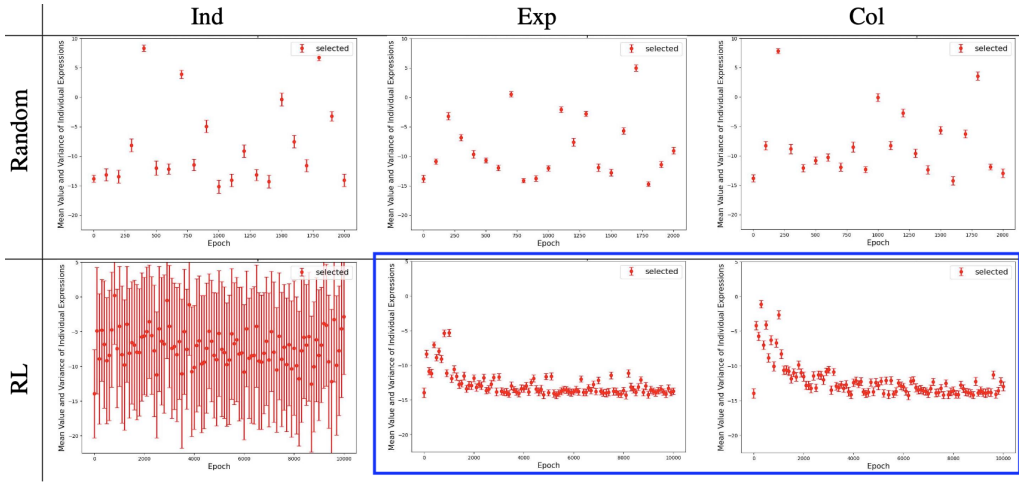



Figure: Mean and st.deviation of individual expressions (Ind. Perspective)

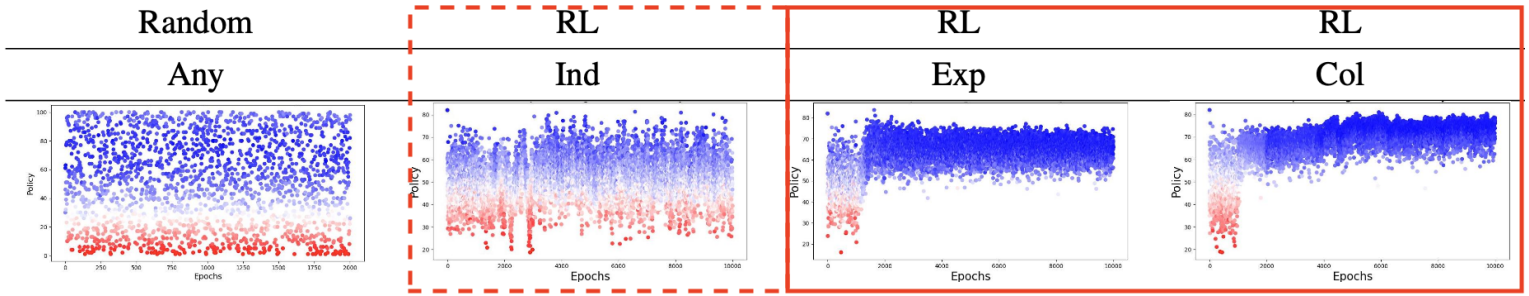
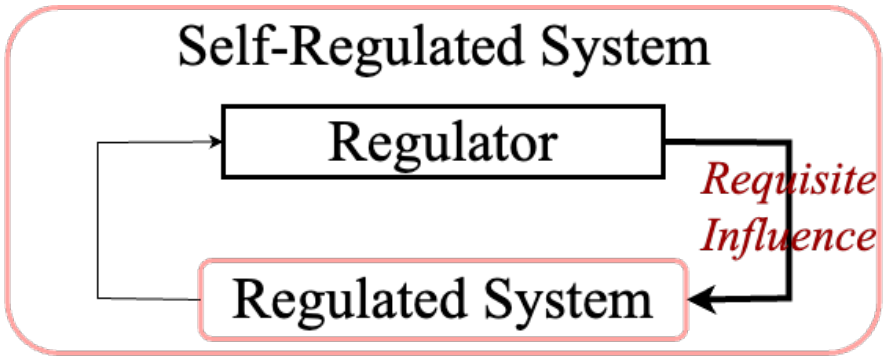



Figure: Collective Expression (Collective Perspective)

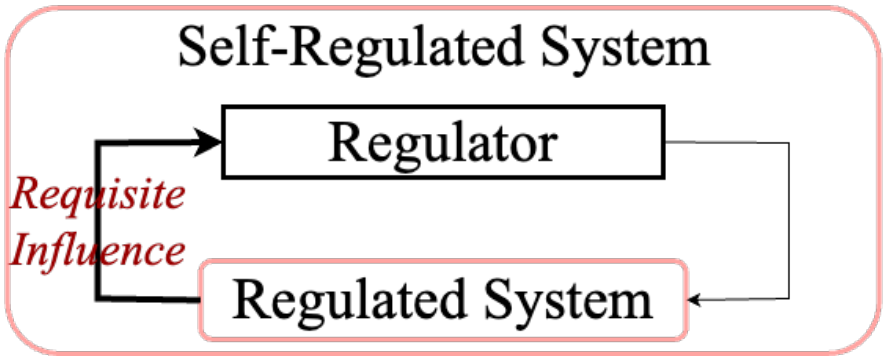
- The pathways for requisite influence from the regulated unit to the regulator are established only when agents aim to learn to attend the voice of the experts (i.e. the ground truth) or the opinion of the collective (i.e. the community truth).

# Experimental Results - Pathways for Requisite Influence

- The pathways for requisite influence from the regulator to the regulated unit are established when regulator is using RL.
- The pathways for requisite influence from the regulated unit to the regulator are established only when agents aim to learn to attend the voice of the experts (i.e. the ground truth) or the opinion of the collective (i.e. the community truth).




*when R is using RL*




*when RU attends ground or community truth*

# Experimental Results -Initial Conditions for Stability

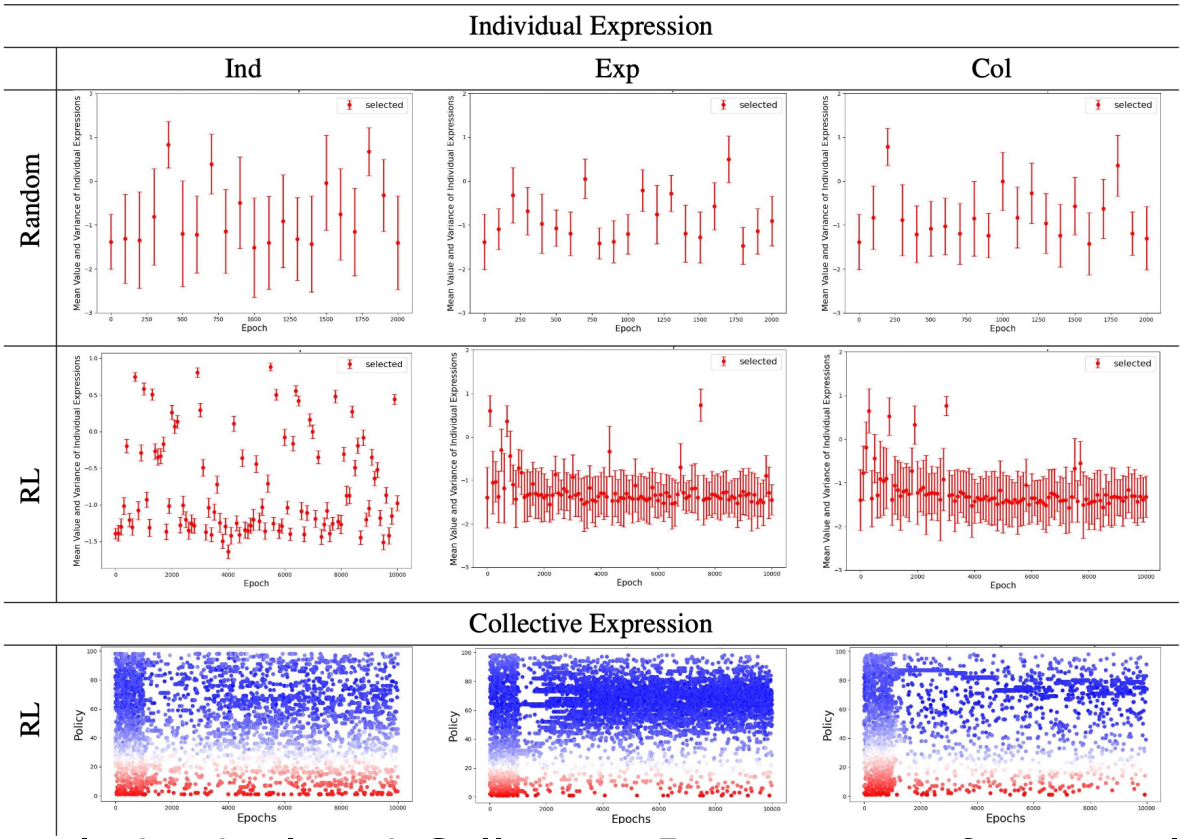
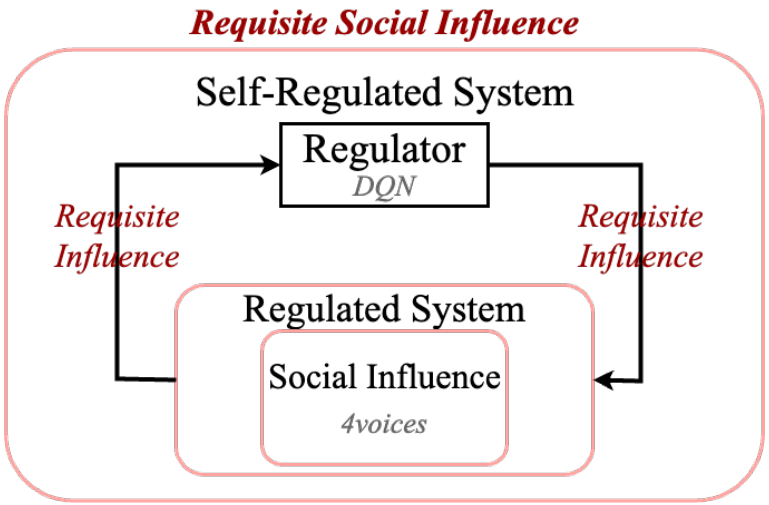



Figure: Individual and Collective Perspectives for a single run

**Systemic stability** is maintained when the regulator selects a policy using RL and the agents learn to attend the voice of the experts (i.e. the ground truth) or the opinion of the collective (i.e. the community truth).

# Summary

- Based on ideas from opinion formation, dynamic social psychology and psycho-acoustics, we introduce the novel 4voices model.
- We provide a hybrid system comprising ML in the regulator (in the form of Deep Q-learning) with MR in the regulated unit (in the form of the 4voices model).
- The experimental results show that the proposed hybrid architecture establishes both the required relational complexity to maintain systemic stability and the pathways for requisite social influence.



# Takeaway

Therefore, the synthesis of Deep Q-learning in the regulator and 4voices in the regulated system **establishes requisite social influence** and points the way towards ethical regulators.

"The implementation of "super-ethical" systems is identified as an urgent imperative for humanity to avoid the danger that super-intelligent machines might lead to a technological dystopia."