Fairness and Cooperation between Independent Reinforcement Learners through Indirect Reciprocity

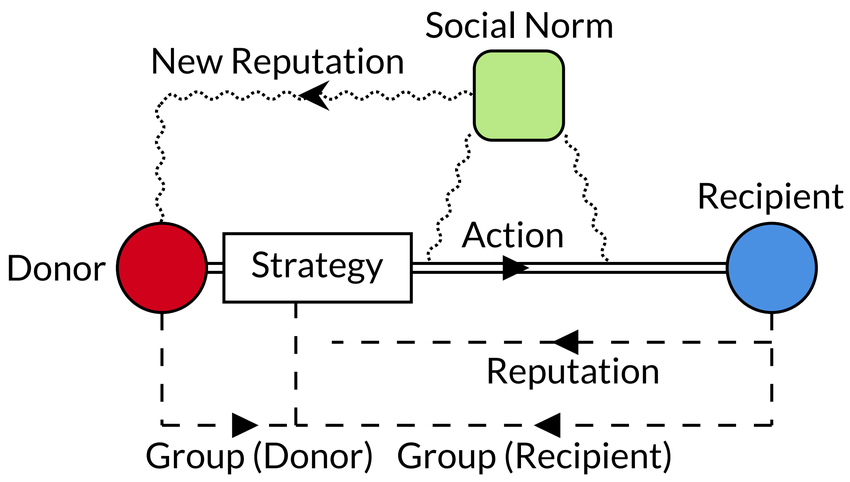
Jacobus Smit [1] Fernando P. Santos [1]

[1]Socially Intelligent Artificial Systems, Informatics Institute, University of Amsterdam

Model

1. Every agent has two visible characteristics: a **static tag** indicating membership to one of **two groups** and **dynamic reputation** updated following **social norms**.
2. Agents **interact in pairs** with randomly chosen roles: the **donor** can **choose to pay** cost c to confer benefit b to the **recipient**.
3. The donor's **policy** may discriminate based on the **relation between their and the recipient's tags** (same or different) and the **recipient's reputation** (good or bad).
4. After interacting, the **social norm determines the donor's new reputation** based on the action taken and the relation/reputation context.

There are **8 possible combinations of 3 bits of information** (action + context), hence a norm is an 8 bit vector, and there are therefore **256 possible norms**.



Multi-agent Reinforcement Learning

Two agents (i, j) who have group relation $x \in \{0, 1\}$ meet. j's reputation is $y \in \{0, 1\}$. With $Q_i \in \mathbb{R}^3$ as the **Q-table** of agent i, the action taken by i is

$$a^* = \arg\max_{a \in \{0,1\}} Q_i[x, y, a],$$

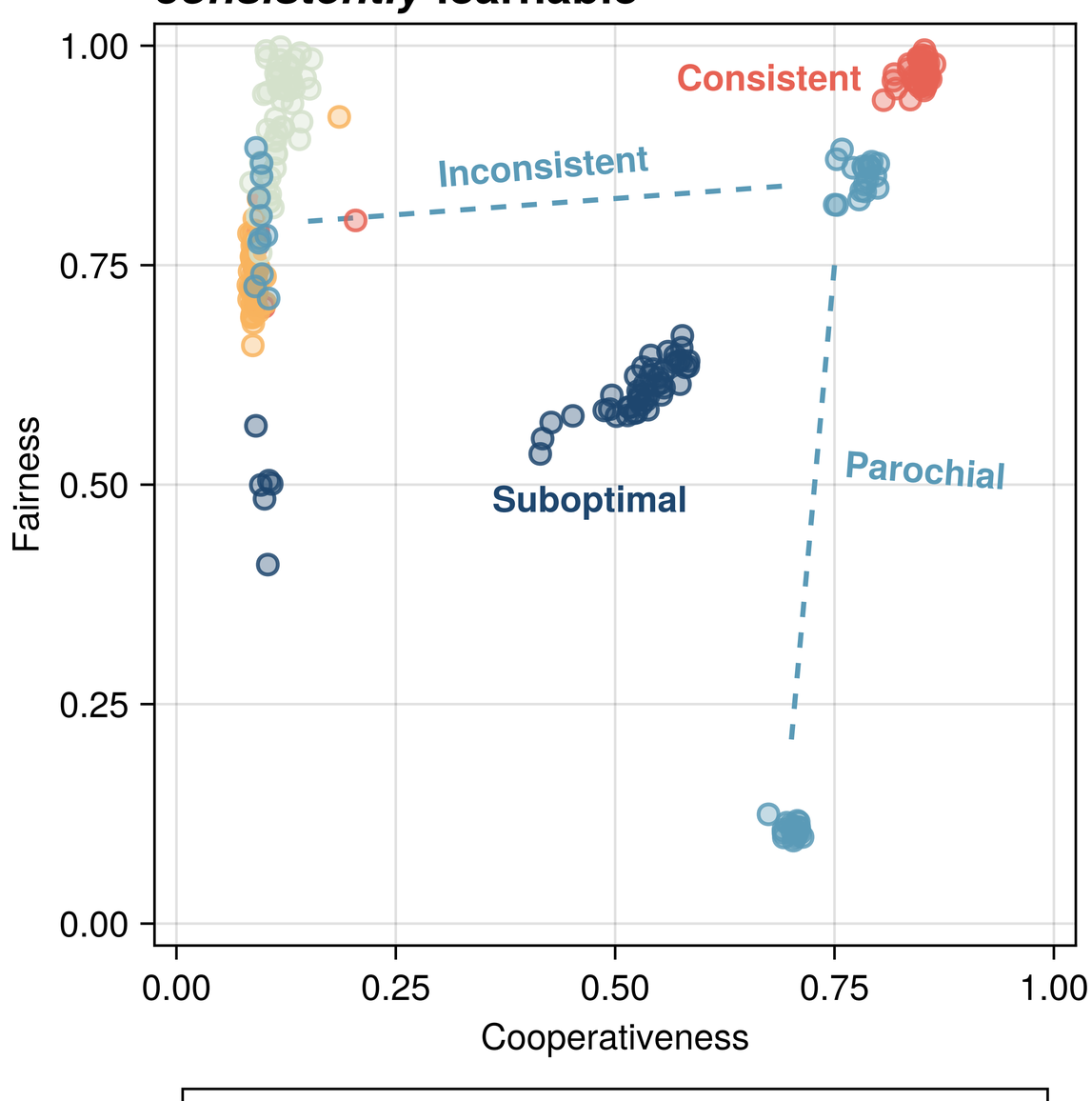
In doing so, agent j **will receive payoff** a^*b and agent i **will pay cost** a^*c. Agent j will attribute this payoff to the **last action they took** ($\hat{a}$) in context $\hat{x}$ and $\hat{y}$:

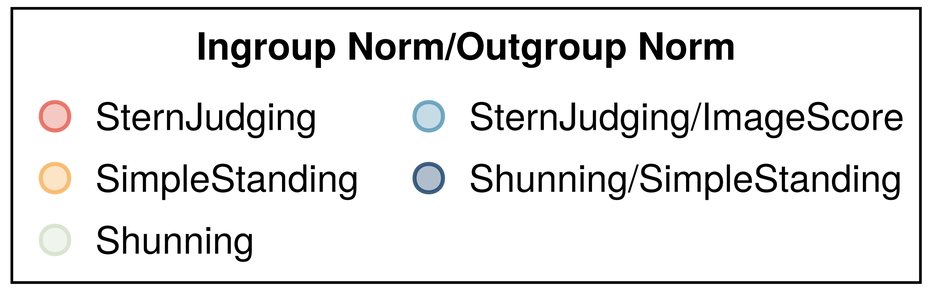
$$Q_i^{t+1}[x, y, a^*] \leftarrow (1 - \alpha)Q_i^t[x, y, a^*] - \alpha a^* c$$
$$Q_j^{t+1}[\hat{x}, \hat{y}, \hat{a}] \leftarrow (1 - \alpha)Q_j^t[\hat{x}, \hat{y}, \hat{a}] + \alpha \hat{a} b$$

Corresponding Q-values are **decayed in every interaction**, regardless of an agent's role. The relevant Q-value is the **last action taken as a donor**.

Paper, Code, and Poster

Very few norms make *fair* cooperation *consistently* learnable





Average path of Q-values under *SternJudging*