

Model

1. Every agent has two visible characteristics: a **static tag** indicating membership to one of **two groups** and **dynamic reputation** determined by the **social norm** of the society.
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 **strategy** 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.
5. Over time, the **average reputation** of a group **converges** which we use to find the **average payoffs** of each group.

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**.

Stability, Cooperation, and Fairness

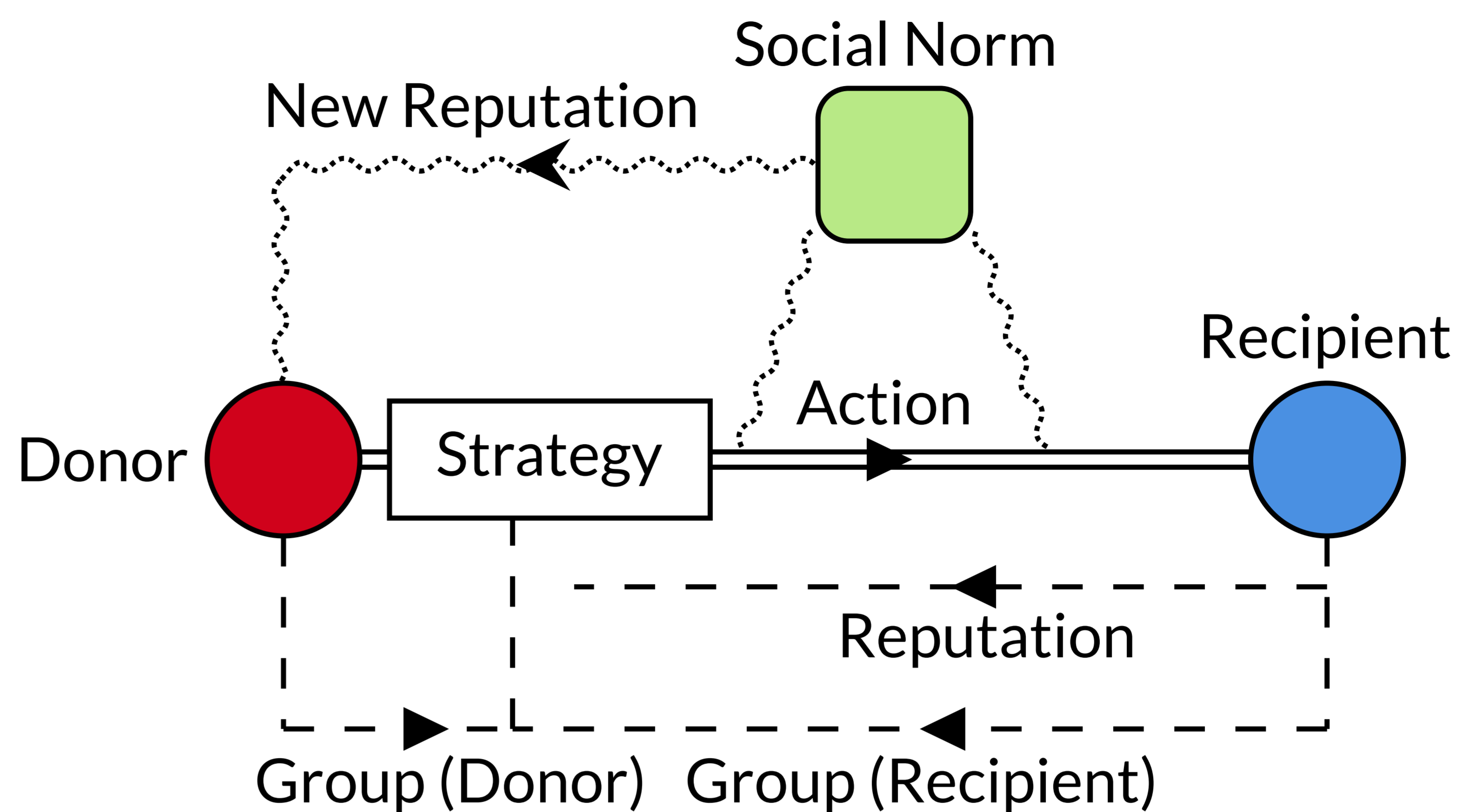
- We assume the **norm is fixed**, and each group all plays their own strategy. We borrow **stability** from **evolutionary game theory**, which means that **no other strategy can outperform** the incumbents of **either group**, and so strategic mutations are eradicated.
- The **cooperativeness** of a society is the probability of cooperation in an **arbitrary interaction**.
- We use **demographic parity ratio** as our definition of fairness: the ratio between the payoffs of the better and worse performing group.

Multi-agent Reinforcement Learning

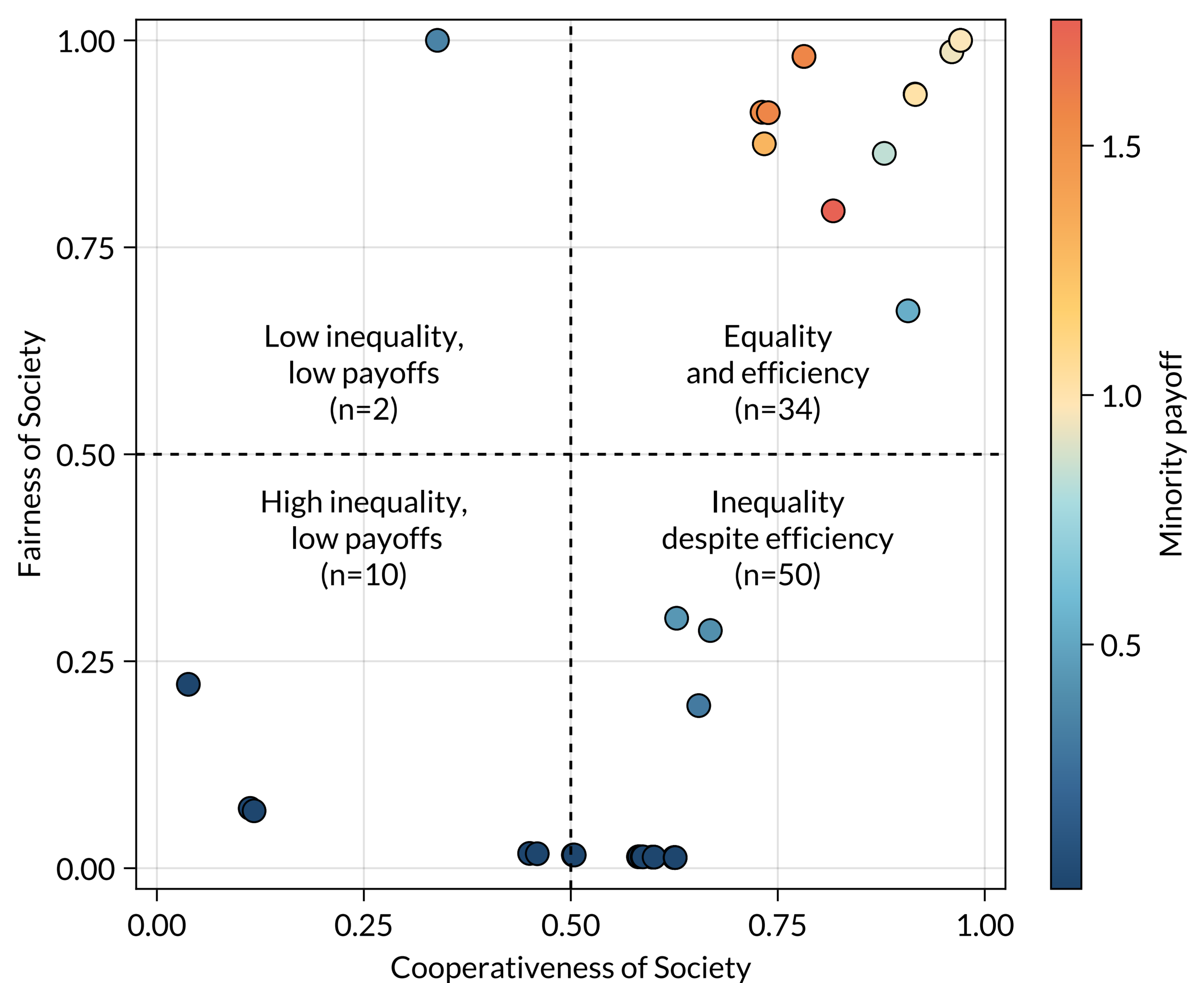
The Q-table is a 4×2 matrix. Agents learn policies with **modified Q-learning**:

- **No future rewards term** as agents require too much information to accurately predict them.
- 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 and code



Fairness and cooperation are achievable given the right combination of social norm and strategies.



The norm decides which policies are learnable in practice with multi-agent RL (majority-group strategies plotted).

