Forgiveness evolves to ensure cooperation in long-term agreements

Luis A. Martinez Vaquero, The Anh Han, Luís Moniz Pereira and Tom Lenaerts

Commitments for enhancing cooperation are widespread in human societies. They offer an alternative to punishment and rewards. Commitments are defined within the context of social dilemmas as agreements to cooperate, with posterior compensations when any of the parties involved defects. It has been shown to be an evolutionarily viable strategy in one-shot social dilemmas. However, in many situations agreements aim to establish long-term mutually beneficial interactions.

Our analytical and numerical results reveal under which conditions revenge, apology, forgiveness and ostracism can evolve and deal with mistakes within ongoing agreements in the context of the Iterated Prisoner's Dilemma.

We show that, when the agreement fails, participants prefer to take revenge by defecting in the subsisting encounters. Incorporating costly apology and forgiveness reveals that, even when mistakes are frequent, there exists a sincerity threshold for which mistakes will not lead to the destruction of the agreement, inducing even higher levels of cooperation.

We also show that when interactions are taking place among group of individuals, reinserting individuals that where expelled after defecting is more efficient than maintaining them ostracized with the cost that may come from it.

Forgiveness is, in its different ways, an evolutionarily viable strategy, which plays a fundamental role in inducing cooperation in repeated dilemmas.