Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms Presentation
Fehr, Ernst ; Fischbacher, Urs ; Gächter, Simon, 2002

Main theme:

Explaining the ubiquity of human cooperation despite the incentives for rational decision makers to break agreements or “cheat”.

Who is a strong reciprocator: a person willing to sacrifice resources
1. to be kind to those, who are kind (strong positive reciprocity)
2. to punish those who are being unkind (strong negative reciprocity)
even if it is costly and provides no present of future material rewards
The kindness of a strong reciprocator is conditional on the perceived kindness of the other player, which differs from a altruistic and reciprocally altruistic player.

The essential feature of **strong reciprocity** is “a powerful constraint for potential cheaters that can generate almost universal cooperation in situations in which purely selfish behavior would cause a complete breakdown of cooperation.”

Research question

Why is cooperation between people so ubiquitous even in one-time, anonymous, non-repeated interactions between non-kin? There are and have been incentives for self-interested individuals to cheat (not reciprocate cooperation) in any kind of social interaction or economic transaction since the payoff from cheating has been higher than honoring the interaction in the absence of “cooperative infrastructure”, especially in one-time interactions where costs are low.

The question then is that, despite these incentives and even in the absence of cooperative infrastructure why is cooperation so ubiquitous?

Cooperative infrastructure: laws, impartial courts, police etc. that alters people’s incentives with the threat of punishment in case of non-compliance with binding contracts.

Literature - what are the possible reasons for this:

Since cooperation regularly also takes place among non-kin, **genetic kinship theory** is not enough to account for the question at hand. Nor is the theory of reciprocal altruism, indirect reciprocity or costly signaling.

Evolution theorists have shown that natural selection can favor reciprocally cooperative behavior in bilateral interactions – when the chances to interact repeatedly with the same individual in the future are sufficiently high, since cheating can be deterred by the threat of withdrawing future cooperation. Therefore, in bilateral repeated interactions reciprocal cooperation can be an evolutionarily stable outcome.

In a similar vein, **game theorists** proved that, when the chances for repeated interactions are sufficiently high, rational decision makers (agents solely interested in their own payoffs/“material” well-being) can establish an equilibrium with full cooperation despite the

existence of short-run cheating incentives. Strategies that punish cheating in repeated games introduce long-term costs for cheating, against its short-term benefits. Hence, cooperation can also be sustained by self-interested, rational actors.

However, in multilateral n-person interactions, which are typical for human societies, the prospects for sustaining cooperation in an evolutionary equilibrium by individual threats of withdrawing future cooperation are quite limited. Boyd and Richardson (1988) have shown that, for reasonable group sizes this mechanism for sustaining cooperation does not work. In this paper, authors offer a distinct cooperation-generating force called “strong reciprocity”.

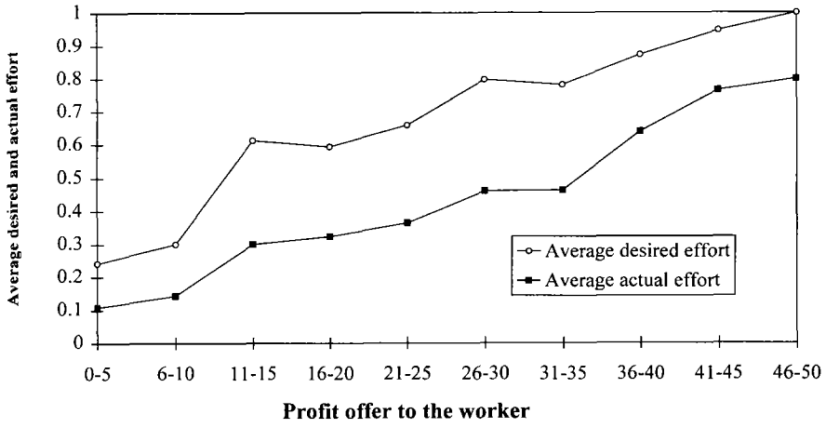
ONE SHOT ARGUMENT - unnatural habitat theory

Particularly problematic is the argument that experimental subjects are not capable of distinguishing between repeated interactions and one-shot interactions. As a consequence, subjects tend to inappropriately apply heuristics and habits in experimental one-shot interactions that are only adaptive in a repeated interaction context but not in a one-shot context**. The evidence suggests that subjects are well aware of the difference between one-shot and repeated interactions because they behave quite differently in these two conditions.**

Taxi driver example
Real world explanation:
Taxi driver + passenger exchange (in the age before Uber)

There also is evidence for strong reciprocity between anonymously interacting trading partners from well-controlled laboratory experiments

***Employer - worker experiment*: A one-shot game where players are anonymous so players cannot build reputation or threaten each other with future withdrawal of cooperation.**
wage offer (w), desired effort level (é), actual effort level (e)
Material payoff for employer: π(f) = 100e - w
Material payoff for worker: π(w) = w - c(e) where c(e) is cost of effort to worker



For example, even purely selfish employers have an incentive to make a cooperative first move (i.e., to make a generous job offer) if they expect a sufficient number of workers to behave in a strongly reciprocal manner. Similarly, even purely selfish workers have an

incentive to provide a high level of effort in case of a reward/punishment mechanism if they expect employers to be strong reciprocators. The existence of this costly reward/punishment mechanism also increases total payoffs generated in the game on net by 40%, in other words making the pie larger!

Multilateral cooperation and punishment opportunities
Public goods game
- 1 + 0.4 = -0.6 tokens
-1 + 4(0.4) = +0.6 tokens
Punishment as a tool of strong reciprocators to force selfish types to cooperate.
Absence of punishment: selfish types induce reciprocators to defect.

Experiment (Fehr and Schmidt (1999))
Prediction - stable groups: When the subject is punished for free riding, they know the punisher is part of the group and has a stronger incentive to coordinate. Thus cooperation rates should be lower in random group design.
Results: subjects punish heavily in both group designs.
Pattern: large majority of punishments is cooperators punishing the defectors.
Punishment of below-average investments prevails in period 10 too.

Impact of punishment pattern on investment behavior:

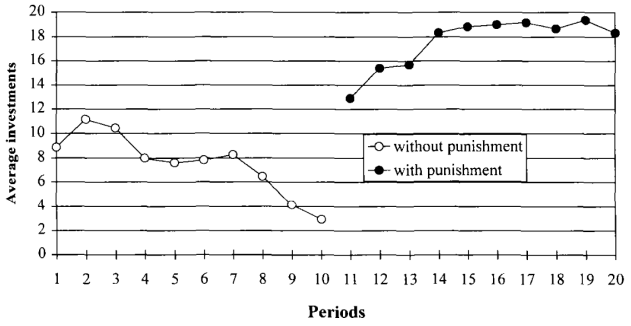


Figure 3. Average investments over time in public good games with stable groups (10 groups).

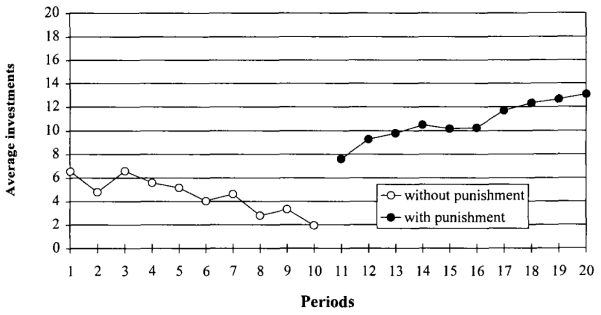


Figure 4. Average investments over time in public good games with random groups (18 groups).

SOCIAL PREFERENCE

Social preference is a term to describe situations where selfishly acting rational decision makers take into account payoffs and intentions i.e., wellbeing of other agents in their utility functions. So in this context, an individual may still be completely selfish, since the benefit of another person is engraved in his utility function. For instance, giving my mom a flower makes her happy and in turn makes me happy too. This might be due to a sense of empathy, a religious imperative of taking care of family, a cultural habit and so on about which social preference theories are silent and give only a proximate cause instead of the ultimate root causes.

Main results

If there **is punishment**, strong reciprocators can force self-interested people to cooperate, whereas when there is **no punishment**, self-interested people's behavior will affect the strong reciprocators in a way that makes them choose non-cooperation.
Social structure is hence important in achieving stable cooperation.