Reciprocity and Trust under an Evolutionary Context

Francisco Poggi

UdeSA

September 24
"The economist is usually not interested in choices or preferences in themselves; (...) There is little investigation of the utility function in itself, and what assumptions are made about it are of a very general nature.

The sociologist and the social psychologist, on the other hand, are interested as much in the choices themselves as in the consequences which flow from a general theory of decision-making."

Kenneth Arrow, 1958
Motivation

- **Typical economic problem**
  - Exogenous preferences + restriction + rationality concept $\rightarrow$ rational choice

- **Game Theory problem**
  - Exogenous preferences + solution concept $\rightarrow$ equilibrium

- **Mechanism design problem**
  - Exogenous possible preference profiles + solution concept $\rightarrow$ implementation game
Motivation

Virtually any behavior, no matter how bizarre, can be "explained" after the fact by simply assuming a taste for it. Thus the chief attraction of the present-aim model turns out also to be its biggest liability. Because it allows to explain everything, we end up explaining nothing. (...)

Our dilemma is how to expand our view of human motives without at the same time becoming vulnerable to Stingler’s crankcase oil objection.

Robert H. Frank, 1997
Microeconomics and Behavior

Solution: tastes are not arbitrarily given, but forged by the pressures of natural selection.
Some tastes are simple (useful independently of whether others share the same taste) but some other preferences are more complex (in the sense that the usefulness of having them depends on the fraction of other individuals who share them). Strategic preference.

- Hawks and doves
- Morality
- Love
- Anger and irrationality
Studying a particular social dilemma (a trust game) I wonder which preferences could survive under an evolutionary context.

I found that, under some circumstances, an heterogeneous equilibrium might arise with both reciprocal and materialistic individuals.

This allows for cooperation in equilibrium, but also generates situations of betrayal.
Where do preferences come from? At first sight: pure fitness. However, this does not hold true if we consider social interaction.

Two possible reasons that might lead to pro-social behavior:

- assortative matching
- observability of types
Introduction

Assortative matching

- People interact relatively more often with people of the same type.
The logic is similar to Vickers (1985) where a firm whose manager maximizes income has a greater benefit than if it was directly maximizing benefits.

If you could commit to an non-optimal behavior can lead you to a better outcome. Preferences, when somewhat observable, can serve as a commitment devise.

The Trust Game

Two players play a sequential game as follows:

- The first one has to decide whether to trust the other ($T$) or not ($-T$).
- When the first player chooses $-T$, the game ends and both players get the outcome $x$.
- When trusted, the second has to decide whether to betray ($B$) or to serve ($S$).
- If the second player betrays, he gets outcome $z$ and the first player gets outcome $l$.
- However, if the second player serves, both players get outcome $y$. 
The Trust Game

Expanded form

Diagram:

1
- T
  X, X
  T
  2
    B
    I, Z
    S
    Y, Y
Let's call $\Sigma$ the set of possible outcomes ($\Sigma = \{l, x, y, z\}$).

There is an objective profit function $q : \Sigma \rightarrow \mathbb{R}$

**Assumption:** $q(z) > q(y) > q(x) > q(l) = 0$

An subjective utility functions $u_i : \Sigma \rightarrow \mathbb{R}$

individuals maximize their expected utility and differ only by their utility functions. As usual, profit drives the evolutionary process while utility determines how agents behave.
Model

We start from a continuum of individuals with a measure normalized to one. Let $\Theta$ be the set of all possible types.

Definition

A state $(\theta, \tau, \pi)$ is a population with individuals of type $\theta, \tau \in \Theta$ where $\pi \in [0, 1]$ represents the proportion of individuals of type $\theta$.

A state contains all the relevant demographic data and it will always be public information.
It will be useful to distinguish types of individuals into two disjoint sets.

**Definition**

A type $\theta$ is **reliable** if and only if prefers the outcome $Y$ to $Z$. In other words, $\theta$ is reliable if never betrays.

**Definition**

A type $\tau$ is **opportunistic** if and only if $\tau$ is not reliable. In other words, $\tau$ is opportunistic if always betrays.
Model
Types: Reciprocal and materialistic individuals

Definition
individual $i$ is **reciprocal** if her type $R \in \Theta$ reliable and behaves as maximizing expected profit when plays first.

Definition
individual $i$ is **materialistic** if her type $M \in \Theta$ is opportunistic and behaves as maximizing expected profit when plays first.
Individuals will be randomly paired to play the trust game. First movers receive a imperfect signal $s \in \{R, O\}$ of the reliability of the type of the second mover. The signal tells the true with probability $p$ and with probability $(1 - p)$ takes a random value with the population ratio. Thus,

$$Pr(s = R|t_j = R) = p + (1 - p)\pi$$

$$Pr(s = R|t_j = O) = (1 - p)\pi$$

$p$ is the realization of a random variable with cumulative function $F(p)$ differentiable and a density function denoted by $f(p)$. 
individuals are randomly matched into ordered pairs.
for each couple \( p \) is realized and observed.
signal \( s \) is also realized and observed.
First player decides whether to trust or not.
If the first player trust, the second player decides if betray or serve.
Individuals receive their respective payoff.
Let’s $a = \frac{X}{Y}$

**Proposition**

When the proportion of reliable is low ($\pi \leq a$), first movers trust others when the signal is positive ($s = R$) and they know the other player enough ($p \geq \frac{a - \pi}{1 - \pi}$). When the proportion of reliable is high enough ($\pi > a$), first movers trust others when (A) the signal is positive ($s = R$), or when (B) the signal is negative ($s = O$) but they don’t know a lot about the other player ($p < \frac{\pi - a}{\pi}$).
M-R Equilibrium

Probability of Being Trusted

\[ b_1 = \frac{a - \pi}{1 - \pi} \text{ and } b_2 = \frac{\pi - a}{\pi} \]

**Proposition**

When \( \pi < a \),

\[
P(\text{trust} | t_j = R, \pi) = \pi [1 - F(b_1)] + (1 - \pi) \int_{b_1}^{1} f(p) p \, dp
\]

\[
P(\text{trust} | t_j = M, \pi) = \pi [1 - F(b_1)] - \pi \int_{b_1}^{1} f(p) p \, dp
\]
Proposition

When $\pi > a$,

$$P(\text{trust} | t_j = R, \pi) = \pi + (1 - \pi)[E(p) + F(b_2) - \int_0^{b_2} f(p)p \, dp]$$

$$P(\text{trust} | t_j = M, \pi) = \pi - \pi E(p) + (1 - \pi)F(b_2) + \pi \int_0^{b_2} f(p)p \, dp$$
Proposition

\[ P(\text{trust}|t_j = R, \pi) \geq P(\text{trust}|t_j = M, \pi) \text{ for every } \pi \in (0, 1), \text{ being equal only when (a) } \pi < a \text{ and } Pr(p > b_1) = 0, \text{ or (b) } \pi > a \text{ and } Pr(p > b_2) = 0. \]

For these cases,

\[ P(\text{trust}|t_j = R, \pi) = P(\text{trust}|t_j = M, \pi) = 0. \]

Proposition

\[ P(\text{trust}|t_j, \pi) \text{ is continuous in } \pi \text{ for } t_j = R, M. \]
M-R Equilibrium

Example: probability of getting each possible outcome
M-R Equilibrium
Expected material payoff

\[ P_T(\pi) = \sum_{w \in W} wP_{wT}(\pi) \]
A proportion $\pi^*$ is an evolutionary equilibrium of context $(\theta, \tau) \in \Omega^2$ if

$$P_\theta(\pi^*) = P_\tau(\pi^*)$$

if

$$\pi^* = 0 \text{ and } P_\theta(0) \leq P_\tau(0)$$

or if

$$\pi^* = 1 \text{ and } P_\theta(1) \geq P_\tau(1)$$
An evolutionary equilibrium $\pi^*$ is evolutionary stable in context $(\theta, \tau) \in \Omega^2$ if

$$\frac{dP_\theta(\pi^*)}{d\pi} > \frac{dP_\tau(\pi^*)}{d\pi}.$$
Results

Heterogeneity

Proposition

If $Pr(p > a) > 0$ then $\pi = 0$ is not an evolutionary equilibrium of context $(R, M)$.

Proposition

If $\int_{1-a}^{1} pf(p) \, dp < \frac{Z-Y}{Z-X}$ then $\pi = 1$ is not an evolutionary equilibrium of context $(R, M)$. 
Results
Existence, Unicity and Stability

Theorem (existence)

If \( Pr(p > a) > 0 \) and \( \int_{1-a}^{1} pf(p) \, dp < \frac{Z-Y}{Z-X} \) exists at least one \( \pi^* \in (0, 1) \) such that \( P_R(\pi^*) = P_O(\pi^*) \).

Theorem (unicity and stability)

If \( p \sim U(0, 1) \) and \( \int_{1-a}^{1} pf(p) \, dp < \frac{Z-Y}{Z-X} \) exists only one evolutionary equilibrium \( \pi^* \). This equilibrium is stable.
Example

- Parameters:
  - $q(x) = 0.66$
  - $q(y) = 1$
  - $q(z) = 1.09$
  - $q(l) = 0$
  - $p \sim U[0, 1]$

- $\pi^* = 0.83$

- Probability of betrayal $= 0.09$
M-R justification
Evolutionary dominance

Definition
We say that a preference relationship $\theta \in \Theta$ evolutionary dominates $\theta' \in \Theta$ if and only if $P_\theta(\pi, \theta, \theta'') \geq P_{\theta'}(\pi, \theta', \theta'') \forall \pi \in [0, 1] \forall \theta'' \in \Theta$ being the relationship strict for some $(\pi, \theta'') \in [0, 1] \times \Theta$. We denote this relationship with an arrow $\leftarrow$.

Theorem
preference $\theta$ is maximal in the partially ordered set $< \Theta, \leftarrow>$ if and only if individuals with preferences $\theta$ behave always as reciprocal or materialistic.
Conclusions

- Payoff-driven evolution is a strong theory and a powerful tool that economy should borrow in order to understand preference formation and explain certain phenomena. Of course the use of these tools does not replace other possible mechanisms to explain cooperation such as repeated games, but it is a different approach that might help us understand endogenously preference formation and some stylized behavior.

- I have shown that, when types are imperfectly observable, an heterogeneous situation with a proportion of reciprocal individual might be evolutionary stable, even under random matching. I have also illustrated the formal conditions about observability and payoffs that guarantees an heterogeneous equilibrium.
The intuition is quite simple: reciprocal individuals are reliable, and this characteristic is somewhat observable. This gives them a greater probability to achieve cooperation, and thus a relative advantage against materialistic individuals. However, if the proportion of reciprocal individuals is relatively high rational agents are more confident and this gives a beneficial opportunity to materialistic individuals.


Agradecimientos

Muchas gracias!