Homo moralis

—

preference evolution under incomplete information and assortative matching

Ingela Alger (TSE, LERNA, CNRS, IAST, and Carleton University)

Jörgen W. Weibull (Stockholm School of Economics and IAST)

Introduction

- Economists traditionally assume selfishness

Introduction

- Economists traditionally assume selfishness
- However, sometimes "social" or "other-regarding" preferences are assumed: *altruism (Becker)*, *warm glow (Andreoni), inequity aversion (Fehr and Schmidt), preference for efficiency (Charness and Rabin), reciprocal altruism (Levine), esteem (Bénabou and Tirole)*

Introduction

- Economists traditionally assume selfishness
- However, sometimes "social" or "other-regarding" preferences are assumed: *altruism (Becker)*, *warm glow (Andreoni), inequity aversion (Fehr and Schmidt), preference for efficiency (Charness and Rabin), reciprocal altruism (Levine), esteem (Bénabou and Tirole)*
- Some classical economists included *moral values* in human motivation, see Smith (1759) and Edgeworth (1881)

Introduction

- Economists traditionally assume selfishness
- However, sometimes "social" or "other-regarding" preferences are assumed: *altruism (Becker)*, *warm glow (Andreoni), inequity aversion (Fehr and Schmidt), preference for efficiency (Charness and Rabin), reciprocal altruism (Levine), esteem (Bénabou and Tirole)*
- Some classical economists included *moral values* in human motivation, see Smith (1759) and Edgeworth (1881)
  - see also Arrow (1973), Laffont (1975), Sen (1977), Tabellini (2008)

Introduction

- What preferences and/or moral values should we expect humans to have, from first principles?

Introduction

- What preferences and/or moral values should we expect humans to have, from first principles?
- Study evolutionary foundations of human motivation!

Introduction

- What preferences and/or moral values should we expect humans to have, from first principles?
- Study evolutionary foundations of human motivation!
  - all our ancestors were successful at reproducing

Introduction

- What preferences and/or moral values should we expect humans to have, from first principles?
- Study evolutionary foundations of human motivation!
  - all our ancestors were successful at reproducing
  - suppose that we have inherited our ancestors' preferences (genetically, epigenetically, culturally)

Introduction

- What preferences and/or moral values should we expect humans to have, from first principles?
- Study evolutionary foundations of human motivation!
  - all our ancestors were successful at reproducing
  - suppose that we have inherited our ancestors' preferences (genetically, epigenetically, culturally)
  - then our preferences should direct us towards maximization of reproductive success

Introduction

- What preferences and/or moral values should we expect humans to have, from first principles?
- Study evolutionary foundations of human motivation!
  - all our ancestors were successful at reproducing
  - suppose that we have inherited our ancestors' preferences (genetically, epigenetically, culturally)
  - then our preferences should direct us towards maximization of reproductive success
- ...but theory suggests that this need not be the case!

Introduction

- Evolution of preferences in *decision problems*
- Counter-mechanism: imperfect perception and response systems
- Research by:
  - Gary Becker
  - Luis Rayo
  - Arthur Robson
  - Larry Samuelson

Introduction

- Preference evolution in *strategic interactions*
- Under *complete information*:
  - Counter-mechanism: commitment value of preferences
  - Example: the responder in an ultimatum game benefits from being known to be inequity averse
- Research:
  - Bester & Güth (1998)
  - Bolle (2000)
  - Possajennikov (2000)
  - Koçkesen, Ok & Sethi (2000)
  - Sethi & Somanathan (2001)
  - Heifetz, Shannon and Spiegel (2007)
  - Alger & Weibull (2010, 2012), Alger (2010)

Introduction

- Preference evolution in *strategic interactions*
- Under *incomplete information*:
  - preferences have no strategic commitment value: natural selection leads to preferences that maximize individual reproductive success
  - *homo oeconomicus* prevails!
- Research:
  - Ok & Vega-Redondo (2001)
  - Dekel, Ely & Yilankaya (2007)

Introduction

- Today's paper: preference evolution in *strategic interactions* under *incomplete information*

Introduction

- Today's paper: preference evolution in *strategic interactions* under *incomplete information*
- We impose few restrictions and yet...

Introduction

- Today's paper: preference evolution in *strategic interactions* under *incomplete information*
- We impose few restrictions and yet...
- The math leads to a general class of moral preferences: *homo moralis*

Introduction

- Today's paper: preference evolution in *strategic interactions* under *incomplete information*
- We impose few restrictions and yet...
- The math leads to a general class of moral preferences: *homo moralis*
- A *homo moralis* gives some weight to own reproductive success and some weight to "what is the right thing to do". Torn between
  - selfishness and
  - morality in line with Immanuel Kant's categorical imperative

Introduction

Kant's categorical imperative

"Act only according to that maxim whereby you can,
at the same time, will that it should become a universal law"

Introduction

- Driving force: assortativity in the matching process
  - Hamilton (1964), Hines and Maynard Smith (1979), Grafen (1979, 2006), Bergstrom (1995, 2003, 2009), Rousset (2004)

Outline

- Model
- Results
- Three points:
  - assortativity is common
  - the behavior of *homo moralis* is compatible with experimental evidence
  - morality is different from altruism
- Conclusion

Model

- A large (continuum) population
- Individuals are randomly matched into pairs
- Each pair has a symmetric interaction, with strategy set X
- $\pi(x, y)$: fitness increment from using strategy $x \in X$ against $y \in X$

Model

- Each individual has a *type* θ, which defines a *goal function* u_θ: $X^2 \to \mathbb{R}$
- Type set: Θ
- u_θ is continuous $(\forall\theta \in \Theta)$
- *Homo oeconomicus*: $u = \pi$
- Each individual's type is his/her *private information*

Model

- At most two types present, θ and τ, in proportions $1-\varepsilon$ and ε
- If ε is small, θ is the *resident* type and τ the *mutant* type
- $\Pr[\theta|\tau,\varepsilon]$: *conditional match probability*
- $\Pr[\theta|\tau,\varepsilon]$ is continuous in ε
- Write σ for $\lim_{\varepsilon\to 0}\Pr[\tau|\tau,\varepsilon]$; the *index of assortativity* of the matching process (Bergstrom, 2003)
    - Uniform random matching $\Rightarrow \sigma = 0$
    - Interactions between siblings who inherited their types from their common parents $\Rightarrow \sigma = 1/2$

Model

Definition

A strategy pair (x^*, y^*) is a **(Bayesian) Nash Equilibrium (BNE)** in state $s = (\theta, \tau, \varepsilon)$ if

$$\begin{cases} x^* \in \arg\max_{x \in X} & \Pr[\theta|\theta, \varepsilon] \cdot u_\theta(x, x^*) + \Pr[\tau|\theta, \varepsilon] \cdot u_\theta(x, y^*) \\ y^* \in \arg\max_{y \in X} & \Pr[\theta|\tau, \varepsilon] \cdot u_\tau(y, x^*) + \Pr[\tau|\tau, \varepsilon] \cdot u_\tau(y, y^*) . \end{cases}$$

Model

- Average fitnesses in state $s = (\theta, \tau, \varepsilon)$ at strategy profile (x^*, y^*):

$$\Pi_\theta(x^*, y^*, \varepsilon) = \Pr[\theta|\theta, \varepsilon] \cdot \pi(x^*, x^*) + \Pr[\tau|\theta, \varepsilon] \cdot \pi(x^*, y^*)$$

$$\Pi_\tau(x^*, y^*, \varepsilon) = \Pr[\theta|\tau, \varepsilon] \cdot \pi(y^*, x^*) + \Pr[\tau|\tau, \varepsilon] \cdot \pi(y^*, y^*)$$

Model

Definition

A type $\theta \in \Theta$ is **evolutionarily stable against a type** $\tau \in \Theta$ if there exists an $\bar{\varepsilon} > 0$ such that $\Pi_\theta(x^*, y^*, \varepsilon) > \Pi_\tau(x^*, y^*, \varepsilon)$ in all Nash equilibria (x^*, y^*) in all states $s = (\theta, \tau, \varepsilon)$ with $\varepsilon \in (0, \bar{\varepsilon})$.

Model

Definition

A type $\theta \in \Theta$ is **evolutionarily unstable** if there exists a type $\tau \in \Theta$ such that for each $\bar{\varepsilon} > 0$ there exists an $\varepsilon \in (0, \bar{\varepsilon})$ with $\Pi_\theta(x^*, y^*, \varepsilon) < \Pi_\tau(x^*, y^*, \varepsilon)$ in all Nash equilibria (x^*, y^*) in state $s = (\theta, \tau, \varepsilon)$.

Results

Definition

An individual is a *homo moralis* with degree of morality $\kappa \in [0, 1]$ if her utility function is of the form

$$u_\kappa(x, y) = (1 - \kappa) \cdot \pi(x, y) + \kappa \cdot \pi(x, x)$$

Homo moralis is torn between selfishness and morality:
- $\pi(x, y)$: maximizing own fitness
- $\pi(x, x)$: doing what would be "right for both", in terms of fitness, if the other party did the same

Results

Definition

A *homo hamiltoniensis* (a homage to the late evolutionary biologist William Hamilton) is a *homo moralis* with degree of morality $\kappa = \sigma$:

$$u_\sigma (x, y) = (1 - \sigma) \cdot \pi (x, y) + \sigma \cdot \pi (x, x)$$

Results

- Let

$$\beta_\sigma(y) = \arg\max_{x \in X} u_\sigma(x, y)$$

- What *HH* does when resident:

$$X_\sigma = \{x \in X : x \in \beta_\sigma(x)\}$$

- Θ_σ^m: set of types τ that, as vanishingly rare mutants, when residents play some $x_\sigma \in X_\sigma$, also play x_σ

Results

Theorem

(Part 1) If $\beta_\sigma(x)$ is a singleton for all $x \in X_\sigma$, then homo hamiltoniensis is evolutionarily stable against all types $\tau \notin \Theta_\sigma^m$.

Results

- Intuition: *HH* preempts mutants

Results

- Intuition: *HH* preempts mutants
- A resident population of *HH* play some x_σ:

$$x_\sigma \in \arg\max_{x \in X} (1-\sigma) \cdot \pi(x, x_\sigma) + \sigma \cdot \pi(x, x)$$

Results

- Intuition: *HH* preempts mutants
- A resident population of *HH* play some x_σ:

$$x_\sigma \in \arg\max_{x \in X} (1-\sigma) \cdot \pi(x, x_\sigma) + \sigma \cdot \pi(x, x)$$

- A vanishingly rare mutant type, who plays some $z \in X$, obtains average fitness

$$(1-\sigma) \cdot \pi(z, x_\sigma) + \sigma \cdot \pi(z, z)$$

Results

- The type space Θ is *rich* if for every strategy $x \in X$ there exists a type for which x is strictly dominant.

Theorem

(Part 2) If Θ is rich, $X_\theta \cap X_\sigma = \varnothing$ and X_θ is a singleton, then θ is evolutionarily unstable.

Results

Intuition

- Consider any resident type θ who plays some x_θ where $x_\theta \notin X_\sigma$

Results

Intuition

- Consider any resident type θ who plays some x_θ where $x_\theta \notin X_\sigma$
- Θ rich $\Rightarrow \exists$ type $\hat{\tau}$ committed to a best reply $\hat{x}$ to x_θ in terms of average mutant fitness (in the limit as $\varepsilon = 0$)

$$\hat{x} \in \arg\max_{x \in X} \ (1-\sigma) \cdot \pi(x, x_\theta) + \sigma \cdot \pi(x, x)$$

Results

- *Homo oeconomicus* thrives in non-strategic environments (decision problems)

Results

- *Homo oeconomicus* thrives in non-strategic environments (decision problems)
- For *homo oeconomicus* to thrive in strategic interactions, it is necessary that the index of assortativity be zero.

Matching processes

- Assortativity is positive as soon as there is a positive probability that both parties in an interaction have inherited their preferences (or moral values) from a common "ancestor" (genetic or cultural)
- A long tradition in biology...
- In social science: culture, education, ethnicity, geography, networks, customs and habits

Matching processes

Interactions between kin: vertical transmission

- Pairwise interactions between siblings, for which strategies are not gender specific
- A population of grown-ups where a proportion $1-\varepsilon$ have type $\theta \in \Theta$ and the residual proportion has strategy $\tau \in \Theta$
- Suppose that couples form randomly
- Assume that each child is equally likely to inherit each parent's type

Matching processes

Interactions between kin: vertical transmission

Proposition

Under random mating and monogamy, $\sigma = 1/2$.

Matching processes

Interactions between kin: oblique transmission

Proposition

- *Assume monogamy, and suppose that each child inherits:*

$\diamond$ *a parent's type with probability* $\rho \in [0, 1]$

$\diamond$ *the type of a uniformly randomly drawn grown-up in the population otherwise*

$\diamond$ *the siblings' choices of role model are statistically independent.*

- *Then* $\sigma = \rho^2/2$.

Matching processes

Interactions between non-kin: education

Proposition

- *Each individual:*

◇ *acquires her business strategies in school*

◇ *enters a new two-person business partnership upon finishing school: with a former schoolmate with probability* $\upsilon \in [0,1]$*, with a graduate uniformly randomly drawn from the whole pool of newly minted graduates in society at large otherwise.*

- *Then* $\sigma = \upsilon$*.*

Matching processes

Interactions between non-kin: migration

Proposition

- *A hunter gatherer society in which each community has a hunting team consisting of two men.*
- *Hunting techniques taught to youngsters.*
- *A fraction $\gamma \in [0, 1]$ of the young men migrate from their native community to a uniformly randomly drawn community in society at large, while the others remain in their native community.*
- *Then $\sigma = 1 - \gamma$.*

Homo moralis in action: dictator game

- Two individuals. Hand money to one of the two, the *dictator*, with equal probability for both

Homo moralis in action: dictator game

- Two individuals. Hand money to one of the two, the *dictator*, with equal probability for both
- The dictator decides, unilaterally, how to split the money

Homo moralis in action: dictator game

- Two individuals. Hand money to one of the two, the *dictator*, with equal probability for both
- The dictator decides, unilaterally, how to split the money
- A strategy $x \in [0,1]$ is the share to give, if dictator, to the other party

$$\pi(x,y) = \frac{1}{2}\left[v(1-x) + v(y)\right]$$

Homo moralis in action: dictator game

- Two individuals. Hand money to one of the two, the *dictator*, with equal probability for both
- The dictator decides, unilaterally, how to split the money
- A strategy $x \in [0, 1]$ is the share to give, if dictator, to the other party

$$\pi(x, y) = \frac{1}{2}\left[v(1 - x) + v(y)\right]$$

- *Homo moralis* gives a positive amount to the other if κ is large enough

Homo moralis in action: ultimatum game

- Two individuals. Hand money to one of the two, the *proposer*, with equal probability for both

Homo moralis in action: ultimatum game

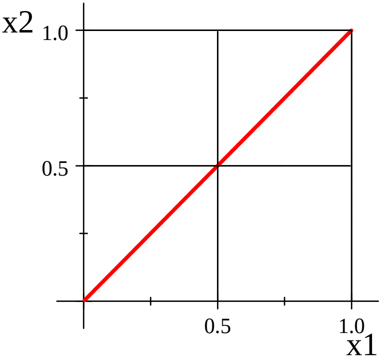
- Two individuals. Hand money to one of the two, the *proposer*, with equal probability for both
- The proposer suggests a split. The other party, the *responder*, may reject, and then all money is lost.

Homo moralis in action: ultimatum game

- Two individuals. Hand money to one of the two, the *proposer*, with equal probability for both
- The proposer suggests a split. The other party, the *responder*, may reject, and then all money is lost.
- A strategy $x = (x_1, x_2) \in [0, 1]^2$ is
  - the share to suggest if proposer, x_1
  - the acceptance threshold if responder, x_2

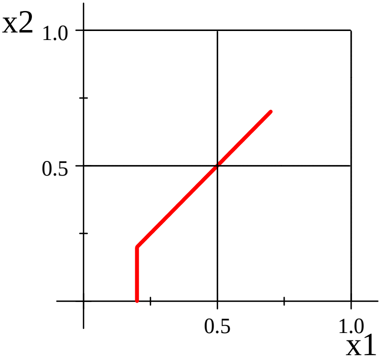
$$\pi(x, y) = \frac{1}{2} v(1 - x_1) \cdot \mathbf{1}_{\{x_1 \geq y_2\}} + \frac{1}{2} v(y_1) \cdot \mathbf{1}_{\{y_1 \geq x_2\}}$$

Homo moralis in action: ultimatum game



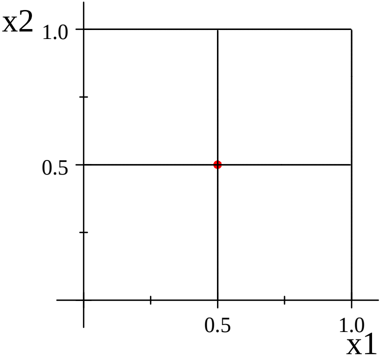
Equilibrium strategies when $\sigma = 0$

Homo moralis in action: ultimatum game



Equilibrium strategies when $\sigma = 1/4$

Homo moralis in action: ultimatum game



Equilibrium strategies when $\sigma = 1$

Morality vs. altruism

- Altruist:

$$u_\alpha(x,y) = \pi(x,y) + \alpha \cdot \pi(y,x),$$

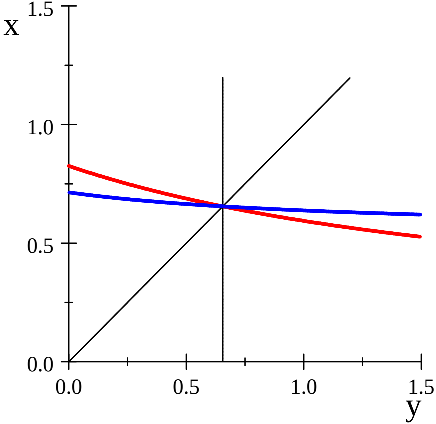
  for some *degree of altruism* $\alpha \in [0,1]$

- *Homo moralis*:

$$u_\kappa(x,y) = (1-\kappa) \cdot \pi(x,y) + \kappa \cdot \pi(x,x)$$

  for some *degree of of morality* $\kappa \in [0,1]$

Morality vs. altruism



Best-reply curves in a public-goods game

Conclusion

- *Homo oeconomicus* thrives in:
  - decision problems
  - under uniform random matching
- In all other situations:
  - natural selection wipes out *homo oeconomicus* and instead favors *homo moralis*
  - the resulting degree of morality is determined by the assortativity in the matching process

Conclusion

- Avenues for further research:
  - interactions between $n > 2$ individuals
  - heterogeneity
  - partial information
  - population processes and stochastic stability
  - implications for political economy & public finance