EVOLUTION, MAXIMIZING BEHAVIOR, AND PRO- (OR ANTI-) SOCIALITY

Ingela Alger (CNRS, TSE, IAST)

Ecology and evolutionary biology, deterministic and stochastic models *IMT October 10 2017*

1 Question

- A population in which individuals are randomly matched into pairs
- Each pair plays a symmetric game
- Common strategy set: $X \subseteq \mathbb{R}^k$ (X is compact and convex)
- Payoff (fitness) from using strategy $x \in X$ against $y \in X$: $\pi(x, y)$
- $\pi : X^2 \to \mathbb{R}$ continuous

- Example 1 [two farmers working in a team]:

$$\pi(x, y) = (x + y)^{0.1} - x^2/2$$

- Example 2 [two hunters working in a team]:

$$\pi(x, y) = (xy)^{0.25} - x^2/2$$

- *Question: which strategy are individuals expected to play?*
- Answer: it depends on the level at which selection occurs
- Economics offers some useful tools to study this

2 Strategy evolution

From now on: continuum population

Definition [Maynard Smith and Price (1973)]: x is ESS under uniformly random matching if for each $y \neq x$, there exists $\bar{\varepsilon} > 0$ such that for all $\varepsilon \in (0, \bar{\varepsilon})$:

$$(1 - \varepsilon) \cdot \pi(x, x) + \varepsilon \cdot \pi(x, y) > (1 - \varepsilon) \cdot \pi(y, x) + \varepsilon \cdot \pi(y, y).$$

ESS is not used in economics as a solution concept. Why?

- In ESS theory: each individual is equipped with a *strategy* to play

- In economic theory: each individual adapts the strategy choice to the situation. This feature is fundamental for the questions asked by economists. For instance: how do people respond to a price change? To a tax increase?

- In economic theory: each individual is equipped with a *goal function* (or utility function), which guides the strategy choice

- In the two-player interaction described above, it would be standard to assume that each individual has a *goal function* $u : X^2 \rightarrow \mathbb{R}$; maximization of this goal function guides the individual's strategy choice

- For economists, the natural question to ask is thus: *which goal function are individuals expected to have?*

Goal functions: examples

$$u(x, y) = \pi(x, y)$$

Goal functions: examples

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

[Becker, G. 1976. "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology," *Journal of Economic Literature*, 14:817–826]

Goal functions: examples

$$\begin{aligned} u(x,y) &= \pi(x,y) - \alpha \cdot \mathsf{max}\{0, \pi(y,x) - \pi(x,y)\} \\ &\quad - \beta \cdot \mathsf{max}\{0, \pi(x,y) - \pi(y,x)\} \end{aligned}$$

[Fehr, E., and K. Schmidt. 1999. “A theory of Fairness, Competition, and Cooperation,” *Quarterly Journal of Economics*, 114:817-868]

Recall: for x to be ESS, it must be that

$$\lim_{\varepsilon \to 0} \left[(1-\varepsilon) \cdot \pi(x,x) + \varepsilon \cdot \pi(x,y) \right] \geq$$
$$\lim_{\varepsilon \to 0} \left[(1-\varepsilon) \cdot \pi(y,x) + \varepsilon \cdot \pi(y,y) \right] \text{ for all } y \in X, y \neq x,$$

or

$$\pi(x,x) \geq \pi(y,x) \text{ for all } y \in X, y \neq x,$$

or

$$x \in \arg\max_{y \in X} \pi(y,x)$$

Any ESS is *as if* each individual sought to maximize the goal function $u(x,y) = \pi(x,y)$

But what if selection were to operate at the level of the goal function? Would $u(x,y) = \pi(x,y)$ then be an evolutionarily stable goal function?

The literature on preference evolution in economics shows that the answer depends on a certain number of factors... I will here focus on the role of *information*.

3 Preference evolution: general framework

- Indirect evolutionary approach [Frank, 1987, Fershtman and Judd, 1987, Güth and Yaari, 1992, Bester and Güth, 1998]
  - individuals with given goal functions are matched to play a game
  - each individual best-responds to the other's strategy, given his goal function [i.e., they play some Nash equilibrium of this game]
  - each gets the material payoff (fitness) associated with the equilibrium strategies

- Need to make assumption about the set Θ of potential goal functions

- Need to make assumption about the information that an individual has about his opponent's type

- Apply and extended ESS concept to three cases:

1. Complete information + a specific parametric class of goal functions

2. Incomplete information + the set of all continuous goal functions

3. Incomplete information + the set of all continuous goal functions + random but assortative matching

4 Preference evolution under complete information: an illustration

- Suppose that each individual is equipped with preferences of the form

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

- Call α the *degree of altruism*: this is the trait that evolution selects for or against

- The individuals observe each other's degree of altruism (i.e., the game is played under complete information)

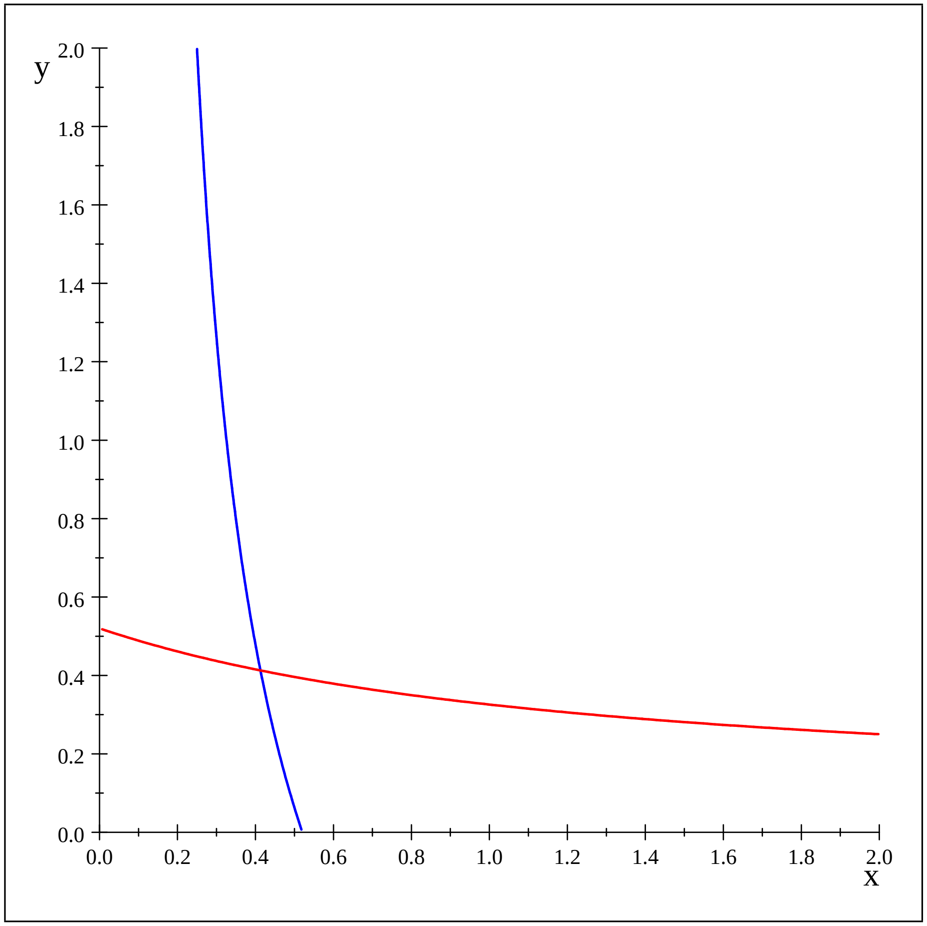
4.1 An example: a public goods game

$$\pi(x, y) = (x + y)^{1/2} - x^2$$

A Nash equilibrium solves:

$$\begin{cases} x^* \in \arg\max_{x \in X} & \pi(x, y^*) + \alpha \cdot \pi(y^*, x) \\ y^* \in \arg\max_{y \in X} & \pi(y, x^*) + \beta \cdot \pi(x^*, y) \end{cases}$$

Best response curves, with $\alpha = 0.5$ for both individuals:



4.2 Related to the biology literature on the evolution of "response rules":

McNamara, Gasson, and Houston (Nature, 1999) (negotiation rule)

Agrawal (Science, 2001) (reaction norm)

André and Day (JTB, 2004), Taylor and Day (JTB, 2004) (response rule)

Akçay, Van Cleve, Feldman, and Roughgarden (AmNat 2009), Akçay and Van Cleve (PNAS 2012) (behavioral response)

4.3 Analysis

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

- Suppose the set of potential degrees of altruism is $\Theta = (-1, 1)$
- Objective: identify *evolutionarily stable degrees of altruism*
- Consider a population with some incumbent or resident degree of altruism α, and inject a share ε of individuals with some mutant degree of altruism β
- We focus on games with a unique Nash equilibrium, that, moreover, is pure, interior and regular

- $V(\alpha, \beta)$: equilibrium material payoff to α-altruist playing against a β-altruist

- Definition: α is an evolutionarily stable degree of altruism if for each $\beta \neq \alpha$, there exists $\bar{\varepsilon} > 0$ such that for all $\varepsilon \in (0, \bar{\varepsilon})$,

$$(1 - \varepsilon) \cdot V(\alpha, \alpha) + \varepsilon \cdot V(\alpha, \beta) > (1 - \varepsilon) \cdot V(\beta, \alpha) + \varepsilon \cdot V(\beta, \beta)$$

4.4 Main result

Proposition 4.1 *[Alger and Weibull, JTB 2012] For any locally evolutionarily stable degree of altruism α^*:*

(i) $\alpha^* = 0$ *if* $\frac{\partial^2 \pi(x,y)}{\partial x \partial y} = 0$ *(selfishness, hedonism)*

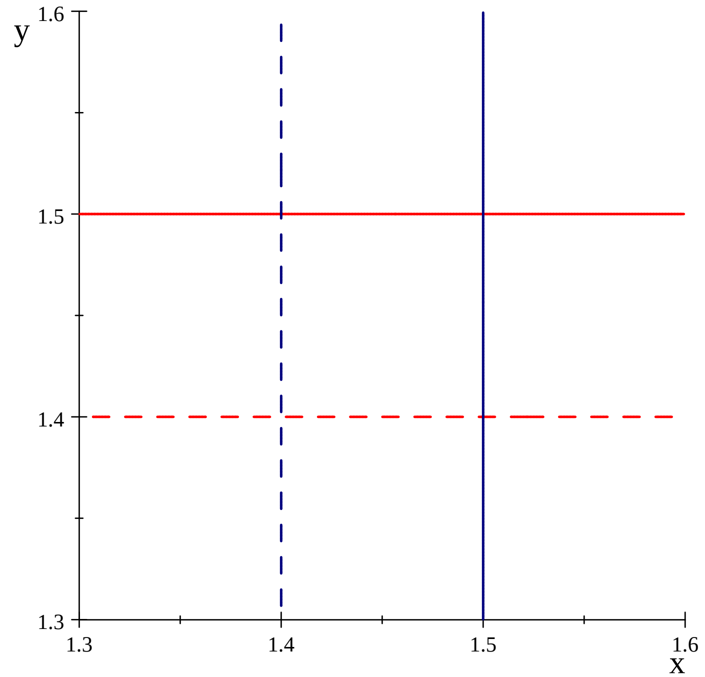
(ii) $\alpha^* < 0$ *if* $\frac{\partial^2 \pi(x,y)}{\partial x \partial y} < 0$ *(hedonism and some Schadenfreude)*

(iii) $\alpha^* > 0$ *if* $\frac{\partial^2 \pi(x,y)}{\partial x \partial y} > 0$ *(hedonism and some empathy)*

- Intuition? Reminiscent of the idea that commitment may have a strategic value [Schelling (1960)]

- Public goods game with $\frac{\partial^2 \pi(x,y)}{\partial x \partial y} = 0$: $\quad \pi(x,y) = x + y - x^2/2$

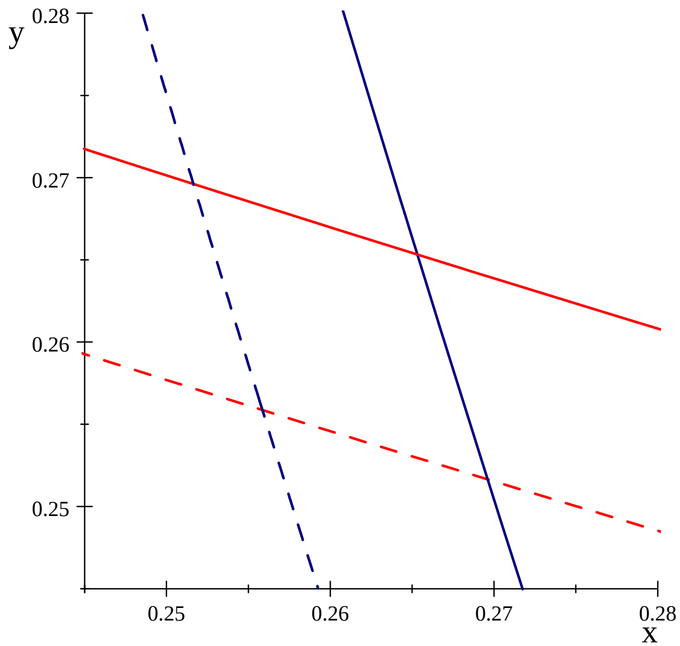
- Incumbents make the same effort, whether playing against an incumbent or a mutant.



Best-reply curves for $\alpha = 0.5$ and $\alpha = 0.4$

- Public goods game with $\frac{\partial^2 \pi(x,y)}{\partial x \partial y} < 0$: $\quad \pi(x,y) = (x+y)^{0.1} - x^2/2$

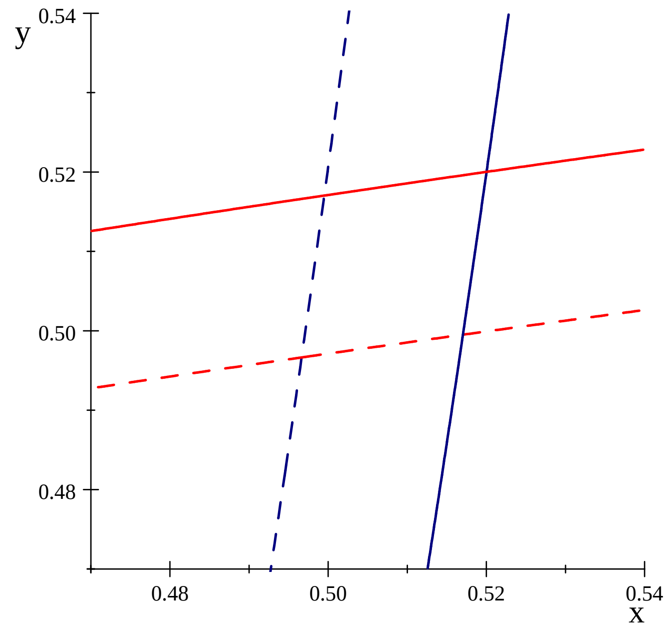
- Incumbents make a *higher* effort towards slightly less altruistic mutants than towards incumbents: compared to the linear case, there is an additional *benefit* of mutating towards lower altruism.



Best-reply curves for $\alpha = 0.5$ and $\alpha = 0.4$

- Public goods game with $\frac{\partial^2 \pi(x,y)}{\partial x \partial y} > 0$: $\quad \pi(x, y) = (xy)^{0.25} - x^2/2$

- Incumbents make a *lower* effort towards slightly less altruistic mutants: compared to linear case, there is an additional *cost* of mutating towards lower altruism.



Best-reply curves for $\alpha = 0.5$ and $\alpha = 0.4$

5 Preference evolution under incomplete information

- Suppose now that matched individuals do not observe each other's type

- Let Θ be the set of all continuous functions $u : X^2 \to \mathbb{R}$

- Consider a population with some incumbent goal function $u_\theta \in \Theta$ and inject a share ε of individuals with some mutant goal function $u_\tau \in \Theta$

- Each individual best-responds to the population state $s = (u_\theta, u_\tau, \varepsilon) \in \Theta^2 \times (0,1)$, given his goal function:

Definition 5.1 *In any population state* $s = (u_\theta, u_\tau, \varepsilon) \in \Theta^2 \times (0,1)$*, a (type-homogenous)* **Bayesian Nash equilibrium** *is a strategy pair* $(\hat{x}, \hat{y}) \in X^2$ *such that*

$$\begin{cases} \hat{x} \in \arg\max_{x \in X} & (1-\varepsilon) \cdot u_\theta(x, \hat{x}) + \varepsilon \cdot u_\theta(x, \hat{y}) \\ \hat{y} \in \arg\max_{y \in X} & (1-\varepsilon) \cdot u_\tau(y, \hat{x}) + \varepsilon \cdot u_\tau(y, \hat{y}) \end{cases}$$

Definition 5.2 (Alger and Weibull, Econometrica 2013) *A goal function* $u_\theta \in \Theta$ *is* **evolutionarily stable against** $u_\tau \in \Theta$ *if* $\exists\ \bar{\varepsilon} > 0$ *such that individuals with* u_θ *earn a higher average material payoff than individuals with* u_τ *in all Nash equilibria in all population states* $s = (u_\theta, u_\tau, \varepsilon)$ *with* $\varepsilon \in (0, \bar{\varepsilon})$.

Definition 5.3 (Alger and Weibull, Econometrica 2013) *A goal function* $u_\theta \in \Theta$ *is* **evolutionarily unstable against** $u_\tau \in \Theta$ *if* $\exists\ \bar{\varepsilon} > 0$ *such that individuals with* u_τ *earn a higher average material payoff than individuals with* u_θ *in all Nash equilibria in all population states* $s = (u_\theta, u_\tau, \varepsilon)$ *with* $\varepsilon \in (0, \bar{\varepsilon})$.

6 Main result

Theorem 6.1 *(Alger and Weibull, 2013) If for all strategies of the opponent there is a unique best response of an individual of type $u_\theta = \pi$, then $u_\theta = \pi$ is evolutionarily stable against all u_τ which do not give rise to the same behavior as $u_\theta = \pi$. In interactions for which there is a unique Nash equilibrium strategy between two individuals with $u_\theta = \pi$, then all other goal functions than $u_\theta = \pi$ [except those that give rise to the same behavior as $u_\theta = \pi$] are evolutionarily unstable.*

- Proof topological; establishes and uses the *upper hemi-continuity* of the Nash-equilibrium correspondence at $\varepsilon = 0$

  [Berge (1959): *Espaces Topologiques*]

7 Evolution under incomplete information, and random but assortative matching

- Definition [Grafen (1979), Hines and Maynard Smith (1979)]: x is ESS under random but assortative matching if for each $y \neq x$, there exists $\bar{\varepsilon} > 0$ such that for all $\varepsilon \in (0, \bar{\varepsilon})$:

$$\begin{aligned} & \Pr[x|x,\varepsilon] \cdot \pi(x,x) + \Pr[y|x,\varepsilon] \cdot \pi(x,y) \\ > \; & \Pr[x|y,\varepsilon] \cdot \pi(y,x) + \Pr[y|y,\varepsilon] \cdot \pi(y,y) \,. \end{aligned}$$

- Consider a population with some incumbent goal function $u_\theta \in \Theta$ and inject a share ε of individuals with some mutant goal function $u_\tau \in \Theta$

- Each individual best responds to the population state $s = (u_\theta, u_\tau, \varepsilon) \in \Theta^2 \times (0, 1)$, given his goal function:

Definition 7.1 *In any population state* $s = (u_\theta, u_\tau, \varepsilon) \in \Theta^2 \times (0, 1)$, *a (type-homogenous)* **Nash equilibrium** *is a strategy pair* $(\hat{x}, \hat{y}) \in X^2$ *such that*

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

- Assume that the conditional probabilities $\Pr[\theta|\theta,\varepsilon]$ and $\Pr[\theta|\tau,\varepsilon]$ are continuous functions of ε.

- Assume also that

$$\lim_{\varepsilon \longrightarrow 0} \Pr[\tau|\tau,\varepsilon] = \sigma$$

$\sigma \in [0,1]$ measures the assortativity in the matching process [Ted Bergstrom, 2003]

- Then the Nash-equilibrium correspondence is *upper hemi-continuous* at $\varepsilon = 0$, and the same proof idea as before can be applied...

- For $\kappa \in [0,1]$, let

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

- An individual with this goal function is torn between selfishness and Kantian morality:

  - $\pi(x,y)$: maximization of own material payoff

  - $\pi(x,x)$: "doing the right thing" (in terms of material payoffs), "if upheld as a universal law" (Kant)

- *Homo moralis*

8 Main result

Theorem 8.1 *(Alger and Weibull, 2013) If for all strategies of the opponent there is a unique best response of an individual with* homo moralis *goal function with $\kappa = \sigma$, then this goal function is evolutionarily stable against all u_τ which do not give rise to the same behavior as this goal function. In interactions for which there is a unique Nash equilibrium between two individuals with* homo moralis *goal function with $\kappa = \sigma$, all other goal functions [except those that give rise to the same behavior as HM with $\kappa = \sigma$] are evolutionarily unstable.*

- An evolutionary foundation for Kantian morality.

9 Bottomline

- Biologists have powerful tools to model ultimate mechanisms for trait selection

- Economists have powerful tools to model proximate mechanisms for behavior

- Building bridges between the two literatures is arguably a fruitful approach to better understand human behavior

- Lehmann, Alger and Weibull (Evolution 2015): take an uninvadable strategy in a structured population; is this strategy *as if* individuals sought to maximize some goal function?

- Further work in progress with Laurent Lehmann and Jörgen Weibull...

Merci !

cnrs
dépasser les frontières

Toulouse
School
of Economics

IAST
INSTITUTE for
ADVANCED
STUDY in
TOULOUSE

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR