Evolutionary Foundations of Human Motivation

—

Some Insights

Ingela Alger (CNRS, TSE, IAST)

TARK Toulouse, July 19, 2019

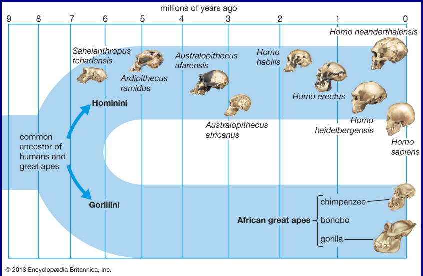
This talk is based on a series of papers with Jörgen Weibull (Stockholm School of Economics & Institute for Advanced Study in Toulouse).

Photo: Ingela Alger

$$\max_{\mathbf{x} \in F(\Omega)} \quad u\left(\mathbf{x}\right)$$

Introduction

Evolution: competition between individuals for survival and reproduction



- Strategic interactions (public goods games, rent seeking, trust games, common pool resource games) must have been common
- Darwinian logic: those alive today had ancestors who were successful at surviving and reproducing; our preferences should reflect this

Introduction

Evolution: our research question

- If preferences that guide the behavior of individuals in strategic interactions are transmitted from one generation to the next, and if the realized material payoffs determine fitnesses, which preferences can evolution by natural selection be expected to lead to? [indirect evolutionary approach, Güth and Yaari, 1992]

Introduction

Evolution: our research question

- If preferences that guide the behavior of individuals in strategic interactions are transmitted from one generation to the next, and if the realized material payoffs determine fitnesses, which preferences can evolution by natural selection be expected to lead to? [indirect evolutionary approach, Güth and Yaari, 1992]
- Goal: understand how the *environment in which a population evolves* affects the evolutionary viability of preferences

Introduction

Evolution: our research question

- If preferences that guide the behavior of individuals in strategic interactions are transmitted from one generation to the next, and if the realized material payoffs determine fitnesses, which preferences can evolution by natural selection be expected to lead to? [indirect evolutionary approach, Güth and Yaari, 1992]
- Goal: understand how the *environment in which a population evolves* affects the evolutionary viability of preferences
- Alger and Weibull (*AER* 2010, *JTB* 2012, *Econometrica* 2013, *GEB* 2016)

Introduction

Evolution: our research question

- If preferences that guide the behavior of individuals in strategic interactions are transmitted from one generation to the next, and if the realized material payoffs determine fitnesses, which preferences can evolution by natural selection be expected to lead to? [indirect evolutionary approach, Güth and Yaari, 1992]
- Goal: understand how the *environment in which a population evolves* affects the evolutionary viability of preferences
- Alger and Weibull (*AER* 2010, *JTB* 2012, *Econometrica* 2013, *GEB* 2016)
- With Jörgen Weibull and Laurent Lehmann (*Evolution* 2015, WP 2019)

Introduction

Evolution: our research question

- If preferences that guide the behavior of individuals in strategic interactions are transmitted from one generation to the next, and if the realized material payoffs determine fitnesses, which preferences can evolution by natural selection be expected to lead to? [indirect evolutionary approach, Güth and Yaari, 1992]
- Goal: understand how the *environment in which a population evolves* affects the evolutionary viability of preferences
- Alger and Weibull (*AER* 2010, *JTB* 2012, *Econometrica* 2013, *GEB* 2016)
- With Jörgen Weibull and Laurent Lehmann (*Evolution* 2015, WP 2019)
- NB: transmission can be biological or cultural

Introduction

Framework

- A large (continuum) population
- Individuals are randomly matched into pairs
- Each pair has a symmetric interaction, with strategy set X
- $w(x, y)$: fitness from playing $x \in X$ against $y \in X$
- Each individual has a *type* θ, which defines a *utility function* $u_\theta\colon X^2 \to \mathbb{R}$
- Type set: Θ (*homo oeconomicus*: $u = w$)

Introduction

Framework

- Consider a population with some resident type θ
- Inject some individuals with some mutant type τ
- Posit an information structure and evaluate fitnesses at Nash equilibrium strategy profile(s)
- The resident type θ withstands the invasion of the mutant type τ if the average fitness of residents exceeds that of mutants, when the mutants are rare
- Seek types which best withstand the invasion against other preference types

Introduction

Framework

- Red thread in our work: recognize that the descendants of the initial mutants tend to interact (due to geographical, social, and cultural barriers)
- $\Pr\left[\tau|\tau,\varepsilon\right]$ may be greater than ε
- Write r for $\lim_{\varepsilon\to 0}\Pr\left[\tau|\tau,\varepsilon\right]$ [index of assortativity, Bergstrom, 2003]
  - Uniform random matching $\Rightarrow r = 0$
  - Interactions between siblings who inherited their types from their common parents $\Rightarrow r = 1/2$

Interactions within the family

Insight #1

Evolution by natural selection may favor weaker intra-family altruism in harsh than in generous environments

Interactions within the family

Evolution of altruistic preferences under complete information

- Alger and Weibull (*AER* 2010, *JTB* 2012)
- Interactions under complete information
- Each individual has some *degree of altruism* $\alpha \in (-1, 1) \equiv \Theta$ towards the opponent:

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

- Assume Nash equilibrium uniqueness

Interactions within the family

Evolution of altruistic preferences under complete information

$$\forall \alpha' \neq \alpha, \alpha' \in (-1, 1):$$

$$w\left[x^*\left(\alpha, \alpha\right), x^*\left(\alpha, \alpha\right)\right]$$
$$> (1-r) \cdot w\left[x^*\left(\alpha', \alpha\right), x^*\left(\alpha, \alpha'\right)\right] + r \cdot w\left[x^*\left(\alpha', \alpha'\right), x^*\left(\alpha', \alpha'\right)\right]$$

Interactions within the family

Evolution of altruistic preferences under complete information

$$(r-\alpha)\cdot x_1^*(\alpha,\alpha)+(1-r\alpha)\cdot x_2^*(\alpha,\alpha)=0$$

Interactions within the family

Application: production and sharing within the family

- Time line:

1. A pair of siblings simultaneously choose productive efforts
2. Each sibling's random output is realized, $Y_i \in \left\{Y^L, Y^H\right\}$. It depends probabilistically on own effort.
3. The siblings observe the outputs, and make transfers to each other.

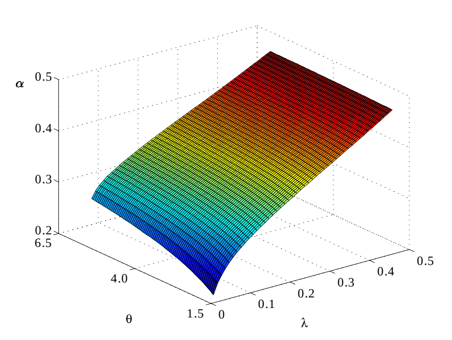
Interactions within the family

Application: production and sharing within the family

- $Y^L = \lambda Y^H$, where $\lambda < 1$ measures *downward risk*
- $p(x) = 1 - e^{-\theta x}$: *return to effort* parameter
- Environment: (λ, θ): an environment (λ', θ') is *harsher* than another environment (λ, θ) if the low output is lower $(\lambda' \leq \lambda)$, and/or the marginal return to effort is smaller $(\theta' \leq \theta)$ with at least one strict inequality

Interactions within the family

Application: production and sharing within the family



- Stable degree of altruism *lower* in harsher environments.
- Intuition: free-rider effect stronger in harsher environments —> more beneficial to mutate towards lower altruism

Interactions beyond the family

- Heifetz, Shannon, and Spiegel (*JET* and *ET* 2007)
- For interactions beyond the family: observability pf preferences is questionable.
- Ok and Vega-Redondo (*JET* 2001), Dekel, Ely, and Yilankaya (*RES* 2007)

Interactions beyond the family

Insight #2

Evolution by natural selection favors Kantian concerns

Interactions beyond the family

Evolution of preferences under incomplete information

- Alger and Weibull (*Econometrica* 2013, *GEB* 2016)
- Each individual's type is his/her *private information*
- Θ: the set of all continuous functions $u : X^2 \to \mathbb{R}$
- Allow for multiple (Bayesian) Nash Equilibria (BNE)

Definition

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

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

Homo moralis is torn between selfishness and a Kantian concern:
- $w (x, y)$: maximizing own fitness
- $w (x, x)$: doing what would be “right for both”, if the other party did the same

Interactions beyond the family

Evolution of preferences under incomplete information

Theorem

(a) Homo moralis with degree of morality $\kappa = r$ is evolutionarily stable against all behaviorally distinguishable types.
(b) Any type θ which is behaviorally distinguishable from homo moralis of degree of morality $\kappa = r$ is evolutionarily unstable.

- Intuition: *HM* preempts mutants
- A resident population of *HM* play some x_r such that

$$x_r \in \arg\max_{x \in X} (1 - r) \cdot w(x, x_r) + r \cdot w(x, x)$$

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

$$(1 - r) \cdot w(z, x_r) + r \cdot w(z, z)$$

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

- The tendency for individuals sharing a common ancestor to interact arises in populations structured into groups, with limited migration between them
- Our ancestors (last 2 million years) lived in small groups (5-150 grown-ups), extending beyond the nuclear family
- Impact of such group structure on preferences?

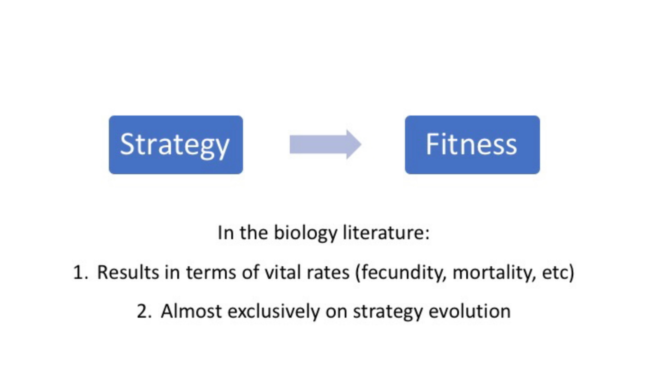
Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

- Population dynamics in populations structured in groups: a long-standing tradition in biology (Wright, 1931)
- Combine the island model with game theory
- Lehmann, Alger, and Weibull (*Evolution* 2015), Alger, Weibull, and Lehmann, WP 2018)

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population



Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

Our contributions:

1. Results in terms of fitness and of material payoffs
2. Allow for preferences to guide the strategy choice

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

Insight #3

Evolution by natural selection favors Kantian concerns at the fitness level...

and Kantian concerns mixed with spite or altruism at the material payoff level

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

- An infinite number of *islands* of size n
- Evolution takes place perpetually over discrete time; each *demographic time period* consists of two phases:

1. *Phase 1*: the n adults in each island interact (X, π)
2. *Phase 2*: the realized material payoffs determine each adult's survival and fecundity; following reproduction, offspring may migrate from their native island to other islands (probability $m > 0$). After migration, individuals compete for available spots; at the end there are exactly n adults in each group.

- This determines each adult's *individual fitness*: the expected number of her *immediate descendants* who have secured a "breeding spot" in the next demographic time period
- Fitness of i: $w\left(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*\right)$

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

- Assume that, initially, everybody has the same utility function u_θ; suddenly exactly one individual with another utility function, u_τ, appears.
  - Does the resident type withstand the invasion of the mutant type?
- Each individual's type is his/her *private information*
- Θ: the set of all continuous functions $u : X^2 \to \mathbb{R}$
- Allow for multiple (Bayesian) Nash Equilibria (BNE)

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

- Any BNE defines a Markov chain that induces a probability distribution over possible mutant *local lineage* realizations
- u_θ is *uninvadable against* u_τ if u_τ is bound to disappear from the population in finite time
- u_θ is *uninvadable in* Θ if it is uninvadable against all $u_\tau \in \Theta$

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

Theorem

Uninvadability requires residents to play some strategy satisfying:

$$x^* \in \arg\max_{x \in X} \quad [1 - r(x_i, x^*)] \cdot \tilde{w}(x_i, x_j, x^*) + r(x_i, x^*) \cdot \tilde{w}(x_i, x_i, x^*),$$

where $r(x_i, x^*)$ is the probability for a randomly drawn mutant playing x_i that his neighbor is also a mutant, when residents play x^*.

- Kantian concern at the fitness level

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

- Weak selection (material payoffs affect fitness marginally)

Theorem

Under weak selection, v is uninvadable:

$$\begin{aligned} v\left(x_i, x_j\right) &= (1-r) \cdot \left[\pi\left(x_i, x_j\right) - \lambda \cdot \pi\left(x_j, x_i\right)\right] \\ &\quad + r \cdot \left[\pi\left(x_i, x_i\right) - \lambda \cdot \pi\left(x_i, x_i\right)\right] \end{aligned}$$

where λ is the coefficient of fitness interdependence:

$$\lambda = \left(-\frac{\partial w\left(\bar{\pi}_i, \bar{\pi}_j, \bar{\pi}^*\right)}{\partial \bar{\pi}_j}\right) / \left(\frac{\partial w\left(\bar{\pi}_i, \bar{\pi}_j, \bar{\pi}^*\right)}{\partial \bar{\pi}_i}\right).$$

- A mix of self-interest, a Kantian concern, and a comparison with other's material payoff: *other-regarding Kantians*

Interactions beyond the family (2)

Evolution of preferences under incomplete information in a structured population

[D] *(i)* $X = \mathbb{R}$, *(ii)* $\pi : \mathbb{R}^n \to \mathbb{R}$ *is continuously differentiable, and* *(iii)* $p_k : X^2 \to [0,1]$ is differentiable for all $k \in \{0, 1, ..., n-1\}$.

Theorem

Under weak selection, uninvadability requires residents to play some strategy satisfying:

$$\hat{x} \in \arg\max_{x \in X} \ (1-\kappa) \cdot \pi\left(x, \hat{\mathbf{x}}^{(n-1)}\right) + \kappa \cdot \pi\left(x, \mathbf{x}^{(n-1)}\right),$$

where $\kappa(x)$ *is the coefficient of scaled relatedness:*

$$\kappa = \frac{r - \frac{1}{n-1}\lambda\left[1 + (n-2)\,r\right]}{1 - \lambda r}$$

Three canonical scenarios

Genes

$$w\left(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*\right) = s\left(\pi_i\right) + m \cdot \left[1 - s\left(\bar{\pi}^*\right)\right] n \cdot \frac{f\left(\pi_i\right)}{nf\left(\bar{\pi}^*\right)}$$

$$+ (1-m) \cdot \left(n - \sum_{j=1}^{n} s\left(\pi_j\right)\right) \cdot \frac{f\left(\pi_i\right)}{(1-m)\sum_{j=1}^{n} f\left(\pi_j\right) + nmf\left(\bar{\pi}^*\right)}$$

$s\left(\pi_i\right)$: probability that i survives to the next demographic time period

$f\left(\pi_i\right) > 0$: i's expected number of offspring

Three canonical scenarios

Genes

Suppose that $s(\pi_i) = s_0$ and $f(\pi_i) = f_0 \cdot \exp(\delta \cdot \pi_i)$. Then:

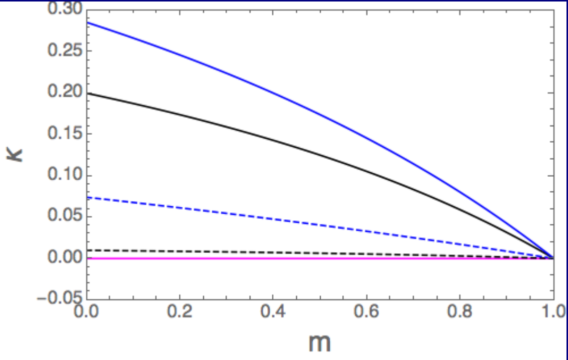
$$r_0^A = \frac{(1-m)^2 + \left(1+m^2\right) s_0}{n - (n-1)(1-m)^2 + \left(1-(n-1)m^2\right) s_0}$$

$$\lambda_0^A = \frac{(n-1)(1-m)^2}{n-(1-m)^2}$$

$$\kappa_0^A = \frac{2(1-m)s_0}{2(1-m)s_0 + n\left[2 - m(1-s_0)\right]}$$

Three canonical scenarios

Genes



Black solid: $s_0 = 1/n$ and $n = 2$ Black dashed: $s_0 = 1/n$ and $n = 10$ Blue solid: $s_0 = 0.8$ and $n = 2$ Blue dashed: $s_0 = 0.8$ and $n = 10$ Pink: $s_0 = 0$

Three canonical scenarios

Guns

$$w\left(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*\right) = \left[(1-\rho) + 2\rho v\left(\boldsymbol{\pi}, \bar{\pi}^*\right)\right] \cdot \left[m \cdot \frac{f\left(\pi_i\right)}{f\left(\bar{\pi}^*\right)} + \right.$$

$$\left. (1-m)\, n \cdot \frac{f\left(\pi_i\right)}{(1-m)\sum_{j=1}^{n} f\left(\pi_j\right) + nmf\left(\bar{\pi}^*\right)}\right]$$

ρ: probability that any given island is drawn into war

$v\left(\boldsymbol{\pi}, \bar{\pi}^*\right)$: probability that an island, in which material payoff profile $\boldsymbol{\pi} \in \mathbb{R}^n$ obtains, wins a war when the average payoff in the rest of the population is $\bar{\pi}^*$

Three canonical scenarios

Guns

Suppose that $f\left(\pi_i\right) = f_0 \cdot \exp\left(\delta \cdot \pi_i\right)$ (as in the preceding example) and $v_n\left(\boldsymbol{\pi}, \bar{\pi}^*\right) = \frac{\exp(\delta \cdot n\bar{\pi})}{\exp(\delta \cdot n\bar{\pi}) + \exp(\delta \cdot n\pi^*)}$. Then:

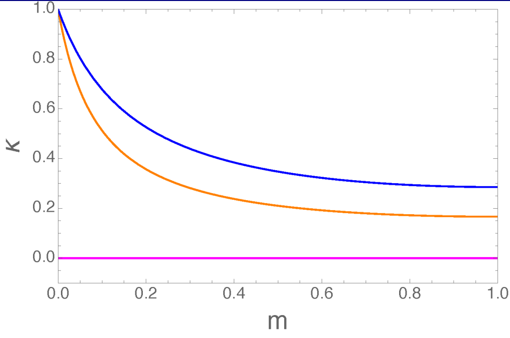
$$r_0^B = \frac{\left(1-m\right)^2}{n - \left(n-1\right)\left(1-m\right)^2}$$

$$\lambda_0^B = \frac{\left(n-1\right)\left(1-m\right)^2 - \rho\left(n-1\right)n/2}{n - \left(1-m\right)^2 + \rho n/2}$$

$$\kappa_0^B = \frac{\rho}{\rho + 2m\left(2-m\right)}$$

Three canonical scenarios

Guns



Pink: $\rho = 0$ Orange: $\rho = 0.4$ Blue: $\rho = 0.8$

Three canonical scenarios

Culture

$$w\left(\pi_i, \boldsymbol{\pi}_{-i}, \bar{\pi}^*\right) = s\left(\pi_i\right) + \; m \cdot \left[1 - s\left(\bar{\pi}^*\right)\right] \cdot \frac{f\left(\pi_i\right)}{f\left(\bar{\pi}^*\right)} + \; (1-m) \cdot \left(n - \sum_{j=1}^{n} s\left(\pi_j\right)\right) \cdot \frac{f\left(\pi_i\right)}{\sum_{j=1}^{n} f\left(\pi_j\right)}$$

$s\left(\pi_i\right)$: probability that i's child emulates i's trait

$(1-m) \cdot \left[n - \sum_{j=1}^{n} s\left(\pi_j\right)\right]$: expected number of children in i's island who did not emulate their parent's trait and who will emulate i's trait

$m \cdot \left[1 - s\left(\bar{\pi}^*\right)\right]$: expected number of children from other islands who did not emulate their parent's trait and who will emulate i's trait

$f\left(\pi_i\right)$: attractiveness of the trait used by i

Three canonical scenarios

Culture

Suppose that $f(\pi_i) = f_0 \cdot \exp(\delta \cdot \pi_i)$ and $s(\pi_i) = s$. Then:

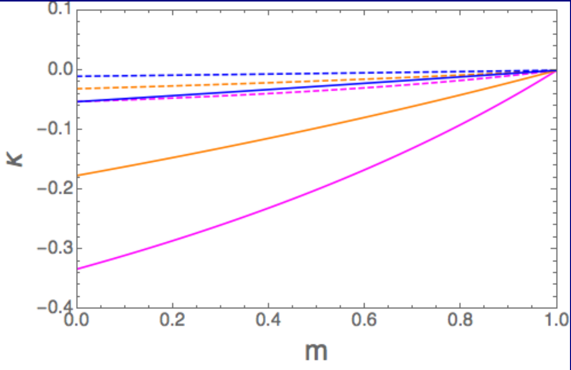
$$r_0^C = \frac{(1-m)\left[2s_0 + (1-m)(1-s_0)\right]}{n(1+s_0) - (1-m)(n-1)\left[2s_0 + (1-m)(1-s_0)\right]}$$

$$\lambda_0^C = \frac{(n-1)(1-m)}{n-(1-m)}$$

$$\kappa_0^C = -\frac{(1-m)(1-s_0)}{2n - \left[m(n-1)+1\right](1-s_0)}$$

Three canonical scenarios

Culture



Pink: $s_0 = 0$, $n = 2$ Orange: $s_0 = 0.4$, $n = 2$ Blue: $s_0 = 0.8$, $n = 2$

Pink dashed: $s_0 = 0$, $n = 10$ Orange dashed: $s_0 = 0.4$, $n = 10$ Blue dashed: $s_0 = 0.8$, $n = 10$

Concluding remarks

- Theory helps us understand how evolutionary forces may have shaped *homo sapiens*

Concluding remarks

- Theory helps us understand how evolutionary forces may have shaped *homo sapiens*
    - impact of environment on preferences

Concluding remarks

- Theory helps us understand how evolutionary forces may have shaped *homo sapiens*
  - impact of environment on preferences
  - discovery of novel preference classes

Concluding remarks

- Theory helps us understand how evolutionary forces may have shaped *homo sapiens*
  - impact of environment on preferences
  - discovery of novel preference classes
  - papers on implications of these novel preferences:
    Alger and Weibull (*The State of Economics, the State of the World* 2019, *Games* 2017), Sarkisian (*Games* 2017), Muñoz (2019)

Concluding remarks

- Theory helps us understand how evolutionary forces may have shaped *homo sapiens*
  - impact of environment on preferences
  - discovery of novel preference classes
  - papers on implications of these novel preferences: Alger and Weibull (*The State of Economics, the State of the World* 2019, *Games* 2017), Sarkisian (*Games* 2017), Muñoz (2019)
- Recent surveys:

Concluding remarks

- Theory helps us understand how evolutionary forces may have shaped *homo sapiens*
  - impact of environment on preferences
  - discovery of novel preference classes
  - papers on implications of these novel preferences: Alger and Weibull (*The State of Economics, the State of the World* 2019, *Games* 2017), Sarkisian (*Games* 2017), Muñoz (2019)
- Recent surveys:
  - Preference evolution in strategic interactions: Alger and Weibull (*Annual Review of Economics* 2019)

Concluding remarks

- Theory helps us understand how evolutionary forces may have shaped *homo sapiens*
    - impact of environment on preferences
    - discovery of novel preference classes
    - papers on implications of these novel preferences: Alger and Weibull (*The State of Economics, the State of the World* 2019, *Games* 2017), Sarkisian (*Games* 2017), Muñoz (2019)
- Recent surveys:
    - Preference evolution in strategic interactions: Alger and Weibull (*Annual Review of Economics* 2019)
    - Evolutionary game theory: Newton (*Games* 2017)

Concluding remarks

- Theory helps us understand how evolutionary forces may have shaped *homo sapiens*
  - impact of environment on preferences
  - discovery of novel preference classes
  - papers on implications of these novel preferences: Alger and Weibull (*The State of Economics, the State of the World* 2019, *Games* 2017), Sarkisian (*Games* 2017), Muñoz (2019)
- Recent surveys:
  - Preference evolution in strategic interactions: Alger and Weibull (*Annual Review of Economics* 2019)
  - Evolutionary game theory: Newton (*Games* 2017)
- For preference evolution in decision problems: see the work of Arthur Robson

Merci !

cnrs
dépasser les frontières

Toulouse School of Economics

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

IAST
INSTITUTE for ADVANCED STUDY in TOULOUSE

erc