Evolution of Preferences in Social Interactions: Some Theoretical Results

Ingela Alger

Toulouse School of Economics, CNRS
Institute for Advanced Study in Toulouse

Pareto Lecture

ASSET

Florence, 10th November 2018

- Presentation based on research agenda developed together with Jörgen Weibull over the past ten years
- Long-term evolution of preferences, when these are transmitted from one generation to the next

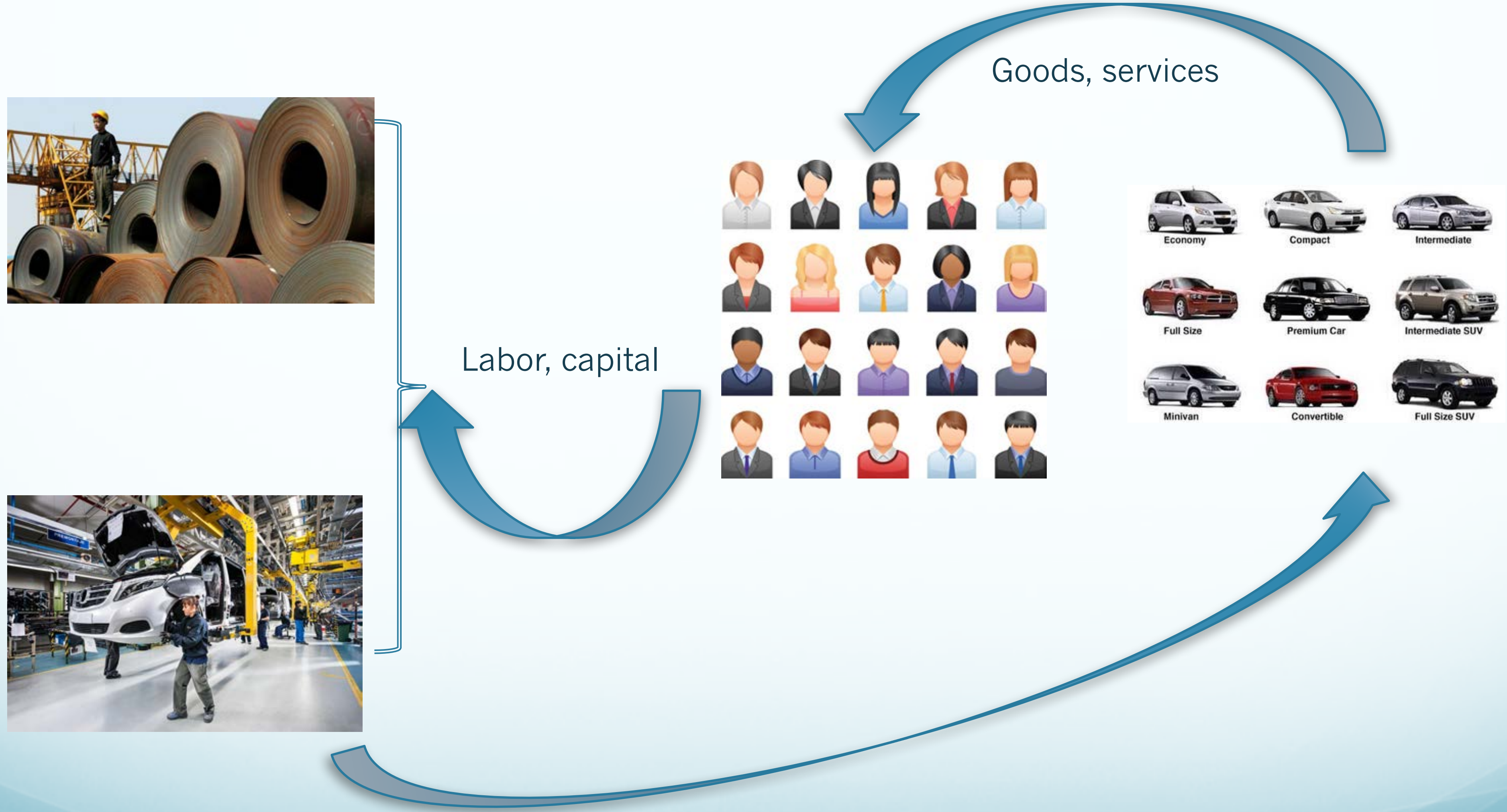
- Presentation based on research agenda developed together with Jörgen Weibull over the past ten years

- Long-term evolution of preferences, when these are transmitted from one generation to the next

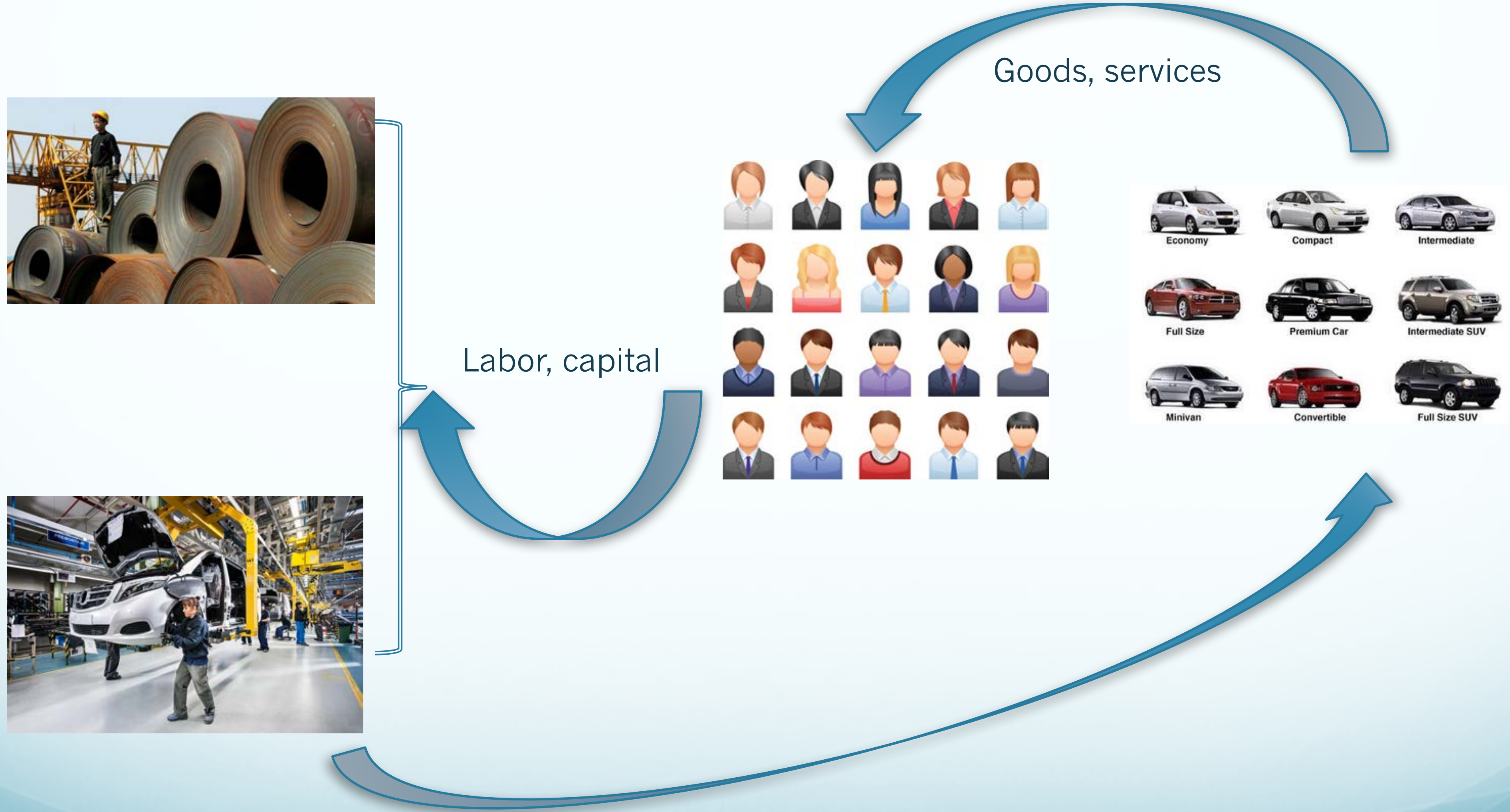
  - "Kinship, Incentives, and Evolution" [*American Economic Review* 2010]

  - "Homo Moralis – Preference Evolution under Incomplete Information and Assortative Matching" [*Econometrica* 2013]

Economics is about resource allocation



The Question: is resource allocation efficient?



The Question: is resource allocation efficient?

1st Theorem of Welfare Economics

A set of sufficient conditions for equilibrium resource allocations to be Pareto efficient.

The Question: is resource allocation efficient?

1st Theorem of Welfare Economics

A set of sufficient conditions for equilibrium resource allocations to be Pareto efficient.

However, this set of conditions is rarely – if ever – met.

The reign of *Homo oeconomicus*

- For about half a century, economists have devoted much attention to externalities and information asymmetries.
- Typically, models are inhabited by the selfish *Homo oeconomicus*:
  - an opportunistic creature without concern for others or for moral values
  - a tradition that dates back to Adam Smith [1776]:

“It is not from the benevolence of the butcher,
the brewer, or the baker, that we expect our dinner,
but from their regard to their own self interest.”

The reign of *Homo oeconomicus*

Do these models deliver effective policy recommendations?

Do they deliver desirable recommendations?

The reign of *Homo oeconomicus*

Do these models deliver effective policy recommendations?

Do they deliver desirable recommendations?

Preferences are central

Many alternative preferences

Altruism [G. Becker]

Warm glow [J. Andreoni]

Fairness concerns/inequity aversion [M. Rabin, E. Fehr and K. Schmidt]

Conditional altruism [D. Levine]

Conformity [D. Bernheim]

Desire to avoid social stigma [A. Lindbeck, S. Nyberg, and J. Weibull]

Identity concerns [G. Akerlof and R. Kranton]

Efficiency concerns [G. Charness and M. Rabin]

Moral motivation [K.A. Brekke, S. Kverndokk, and K. Nyborg]

Image concerns [R. Bénabou and J. Tirole, T. Ellingsen and M. Johannesson]

Honesty concerns [I. Alger and R. Renault]

One step beyond: evolutionary foundations of human motivation

- Which preferences should we *expect* humans to have, from first principles ?
  - *Homo oeconomicus* ?
  - If not, then what ?
  - And why ?
- If preferences are transmitted from one generation to the next, then evolutionary logic can be used to examine these questions

One step beyond:
evolutionary foundations of human motivation

- Evolutionary logic:
  - Human populations have evolved under scarcity of resources
  - Not all who are born survive and not all who survive reproduce
  - Darwinian logic: if we have inherited our ancestors' traits, these should reflect the ability to survive and reproduce

One step beyond:
evolutionary foundations of human motivation

- Some early contributions in economics

  Frank [1987]

  Güth and Yaari [1992]

  Bergstrom [1995]

  Robson [1996]

  Bester and Güth [1998]

  Ok and Vega-Redondo [2001]

  Sethi and Somanathan [2001]

  Dekel, Ely and Yilankaya [2007]

  Heifetz, Shannon and Spiegel [2007]

Common theoretical framework

- Analysis at the population level
- A sequence of generations
- Each individual inherits his/her preference trait from someone in the preceding generation
- In each generation individuals are matched to interact
- Preferences, information, and beliefs determine equilibrium behaviors
- Equilibrium behaviors determine individual reproductive successes
- Preferences that lead to a higher individual reproductive success are represented in a larger population share in the next generation

Common theoretical framework

- While evolution is dynamic, a static stability concept is often adopted
- Initially, some preference trait – the *incumbent* trait – is present in all individuals
- Suddenly a *mutant* preference trait appears
- The incumbent preference trait is *evolutionarily stable* against the mutant trait if incumbents achieve a higher reproductive success than rare mutants
  [builds on the ESS concept, due to Maynard Smith and Price (1973)]

Today

- Focus on two theoretical insights:

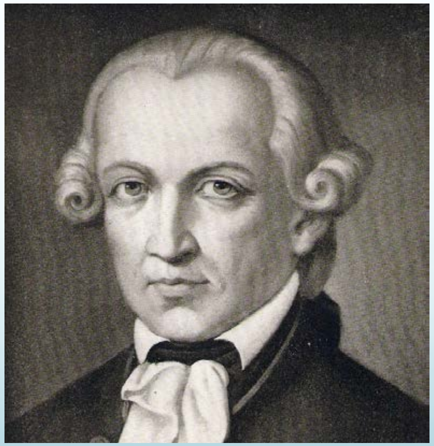
  A. Preferences that give an edge in the evolutionary race may depend on the environment in which the population evolves

Today

- Focus on two theoretical insights:

  A. Preferences that give an edge in the evolutionary race may depend on the environment in which the population evolves

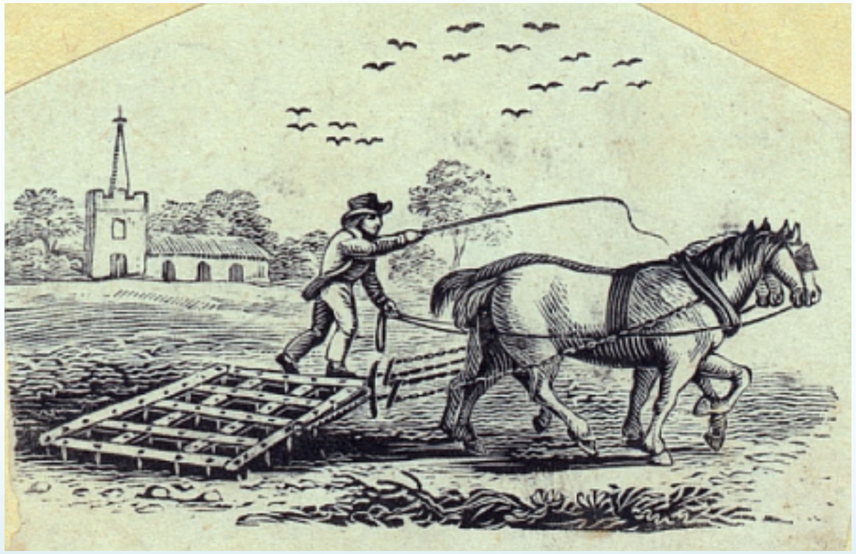
  B. In an analysis with minimal restrictions on the set of potential preferences, the math led to a novel class of (moral) preferences



A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

Imagine… a pre-industrial population



A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

- In each generation, individuals who survive until sexual maturity mate
- Each couple gets two children
- As young adults, children interact in sibling pairs:

  - Each sibling chooses a production effort
  - Output is uncertain (high or low)
  - Siblings observe each other's output
  - They may transfer output to each other

- Efforts and transfers determine each sibling's expected reproductive success (probability of surviving until sexual maturity)

A. Evolutionarily stable intra-family altruism
[Alger and Weibull, *AER* 2010]

- Effort and transfer choices are guided by preferences
- Preference trait in focus: the degree of altruism towards one's sibling
- Specifically:

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

where $\pi(x,y)$ is own reproductive success,

$\pi(y,x)$ is that of the sibling,

and $\alpha \in [-1,1]$

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

- Siblings observe each other's preferences
- Each sibling pair thus plays a game of complete information
- Focus on games such that in any sibling pair there is a unique Nash equilibrium
- For any given pair of altruism degrees, α, and β, let:

$x^*(\alpha, \beta)$ and $x^*(\beta, \alpha)$ denote the equilibrium strategies,

and $\Pi(\alpha, \beta)$ and $\Pi(\beta, \alpha)$ the equilibrium reproductive successes:

$$\Pi(\alpha, \beta) = \pi(x^*(\alpha, \beta), x^*(\beta, \alpha)) \text{ and } \Pi(\beta, \alpha) = \pi(x^*(\beta, \alpha), x^*(\alpha, \beta))$$

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

- Suppose that there is some incumbent degree of altruism α
- Suddenly a mutant degree of altruism α' appears
- Recall:

"The incumbent preference trait is *evolutionarily stable against* against the mutant trait if incumbents achieve a higher reproductive success (RS) than rare mutants"

- Formally: let ε be the share of mutants in the population
- α is evolutionarily stable against α' if, for all small ε, incumbents achieve a strictly higher average RS than mutants
- Calculating these average RS's requires knowing the probabilities that carriers of the incumbent and the mutant preference traits get to interact with each other

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

- Assume that each child inherits the preference trait from one parent, with equal probability for each parent (genetic or cultural transmission)

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

- Assume that each child inherits the preference trait from one parent, with equal probability for each parent (genetic or cultural transmission)
- Then, α is ***evolutionarily stable against α'*** if:

$$\Pi(\alpha, \alpha) > \frac{1}{2} \cdot \Pi\left(\alpha', \alpha\right) + \frac{1}{2} \cdot \Pi\left(\alpha', \alpha'\right)$$

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

- Assume that each child inherits the preference trait from one parent, with equal probability for each parent (genetic or cultural transmission)

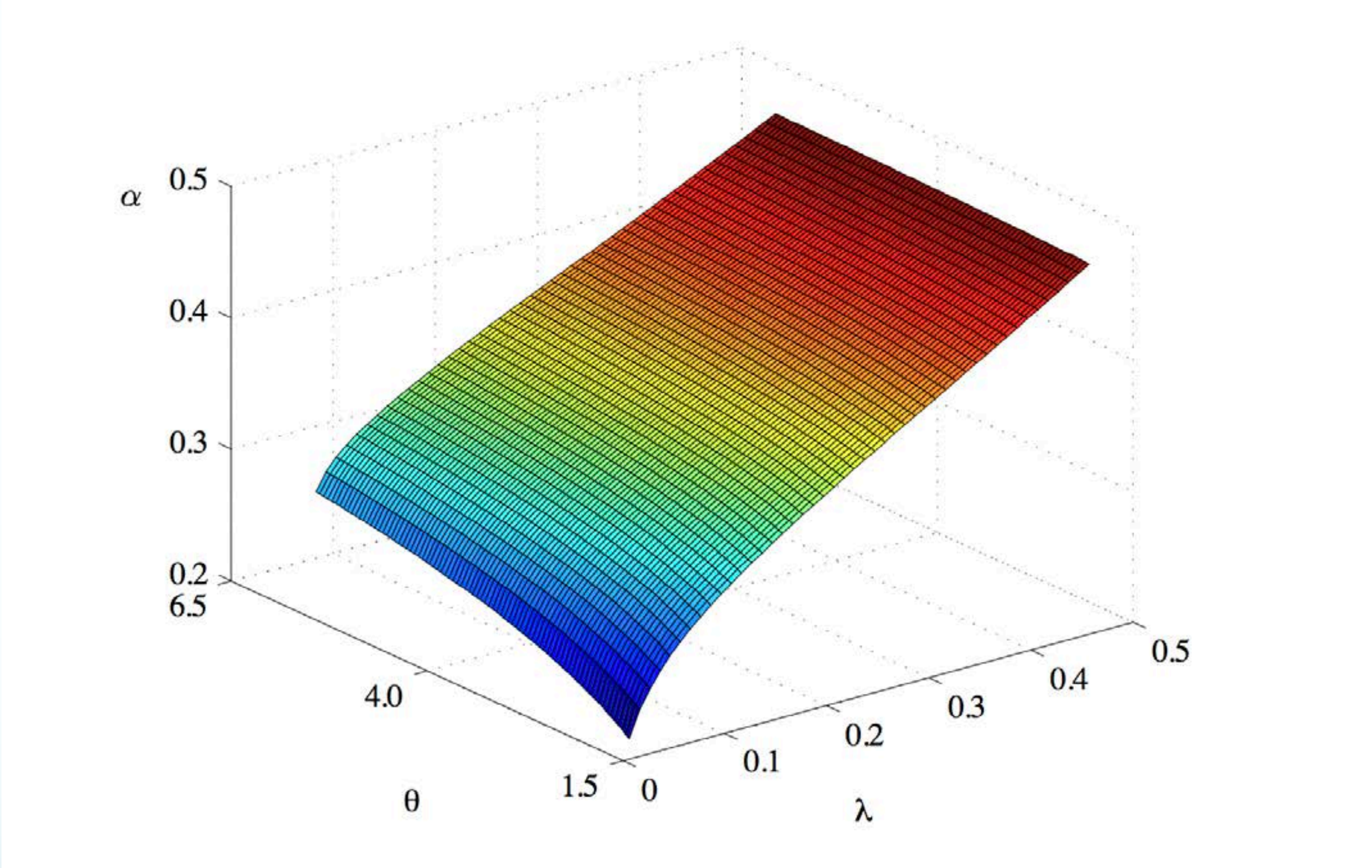
- Then, α is ***evolutionarily stable against α'*** if:

$$\Pi(\alpha, \alpha) > \frac{1}{2} \cdot \Pi\left(\alpha', \alpha\right) + \frac{1}{2} \cdot \Pi\left(\alpha', \alpha'\right)$$

- And α is ***evolutionarily stable*** if it is evolutionarily stable against all α'

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]



λ: the ratio of low to high output

θ: the marginal return to effort

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

Key result:

the evolutionarily stable degree of intra-family altruism
may depend on exogenous features of the environment
in which the population evolves, such as the climate

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

- Intra-family altruism may affect many decisions:
  - productive efforts [Azam and Gubert 2005, Baland et al. 2016, Alger et al. 2018]
  - investment and savings decisions [Di Falco and Bulte 2011 & 2013, Jakiela and Ozier 2016; Alby, Auriol, and Nguimkeu 2018]
- A population in which intra-family altruism is weak may develop differently from one in which it is strong

A. Evolutionarily stable intra-family altruism

[Alger and Weibull, *AER* 2010]

Broader implication (1):

a society's development path may depend on

exogenous features of the environment in which the

population evolves, such as the climate,

through the impact that these features has on preferences

A. Evolutionarily stable intra-family altruism
[Alger and Weibull, *AER* 2010]

Broader implication (2):

can this kind of analysis help explain differences in
preference distributions across cultures [see work by Armin Falk] ?

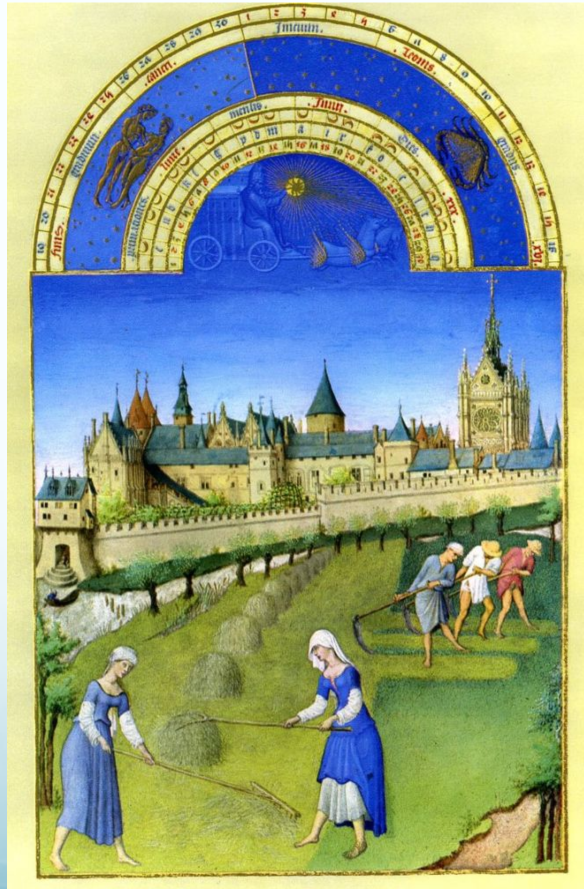
B. A novel class of preferences emerges
[Alger and Weibull, *Econometrica* 2013]

- Objective:
- Formulate and analyze a general model of preference evolution
- Make minimalistic assumptions on the set of potential preferences
- The math leads to a novel class of preferences

B. A novel class of preferences emerges

[Alger and Weibull, *Econometrica* 2013]

Imagine… a pre-industrial population



B. A novel class of preferences emerges
[Alger and Weibull, *Econometrica* 2013]

- In each generation, adults produce offspring (asexually) and die
- Each child inherits preference trait from the unique parent
- Young individuals are randomly matched into pairs to interact
- Set of strategies: X (compact and convex)
- Expected reproductive success from playing x against y : $\pi(x,y)$
- The function π: $X^2 \rightarrow \mathbb{R}$ is continuous

B. A novel class of preferences emerges
[Alger and Weibull, *Econometrica* 2013]

- Strategy choices are guided by preferences
- Set of potential preferences: the set of continuous functions $u: X^2 \rightarrow \mathbb{R}$
- Interacting individuals do not observe each other's preferences
- Each matched pair thus plays a game of incomplete information
- We do not require equilibrium uniqueness

B. A novel class of preferences emerges

[Alger and Weibull, *Econometrica* 2013]

- Although matching is random, it may be assortative:

For a given population state $s = (\theta, \tau, \varepsilon)$, let:

- $\Pr[\tau|\tau, \varepsilon]$ be the probability for a mutant to be matched with a mutant
- $\sigma = \lim_{\varepsilon \to 0} \Pr[\tau|\tau, \varepsilon]$
- $\sigma \in [0, 1]$ be the *index of assortativity* (Bergstrom, 2003)

B. A novel class of preferences emerges

[Alger and Weibull, *Econometrica* 2013]

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

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

Average material payoffs in population state $s = (\theta, \tau, \varepsilon)$, given some BNE (x^*, y^*)

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

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

B. A novel class of preferences emerges

[Alger and Weibull, *Econometrica* 2013]

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 BNE* (x^*, y^*) *in all states* $s = (\theta, \tau, \varepsilon)$ *with* $\varepsilon \in (0, \bar{\varepsilon})$.

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 BNE* (x^*, y^*) *in state* $s = (\theta, \tau, \varepsilon)$.

B. A novel class of preferences emerges

[Alger and Weibull, *Econometrica* 2013]

Theorem *(a) Homo moralis with degree of morality* $\kappa = \sigma$ *is evolutionarily stable against all behaviorally distinguishable types.*

(b) Any type θ *which is behaviorally distinguishable from homo moralis of degree of morality* $\kappa = \sigma$ *is evolutionarily unstable.*

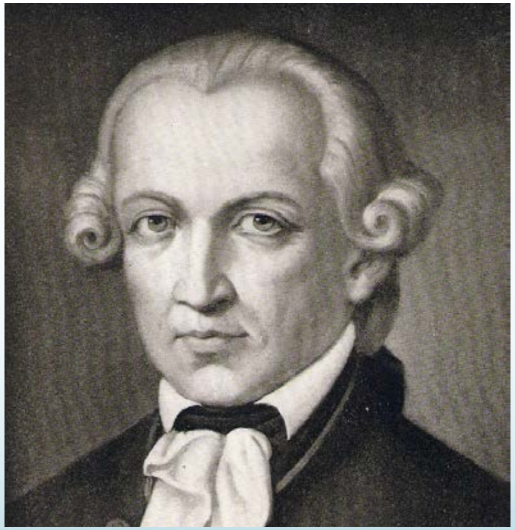
B. A novel class of preferences emerges

[Alger and Weibull, *Econometrica* 2013]

Definition *A homo moralis is an individual with utility function*

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

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



Immanuel Kant (1724-1804)

B. A novel class of preferences emerges

[Alger and Weibull, *Econometrica* 2013]

- Intuition:

(i) *HM* with $\kappa = \sigma$ preempts mutants.

For instance, for $n = 2$: a resident population of *HM* with $\kappa = \sigma$ play some

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

while a vanishingly rare mutant type, who plays some $z \in X$, obtains expected material payoff

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

(ii) Any resident type θ that does not behave like *HM* with $\kappa = \sigma$ can be invaded by a mutant that is "committed" to a strategy that would maximize the mutant's material payoff in the limit as the mutant population share $\varepsilon \to 0$

B. A novel class of preferences emerges
[Alger and Weibull, *Econometrica* 2013]

Key results:

- a hitherto unknown preference class emerges from the analysis

- evolution by natural selection favors preferences
reminiscent of Kantian morality
whenever a rare mutant is more likely than an incumbent
to interact with another rare mutant

B. A novel class of preferences emerges
[Alger and Weibull, *Econometrica* 2013]

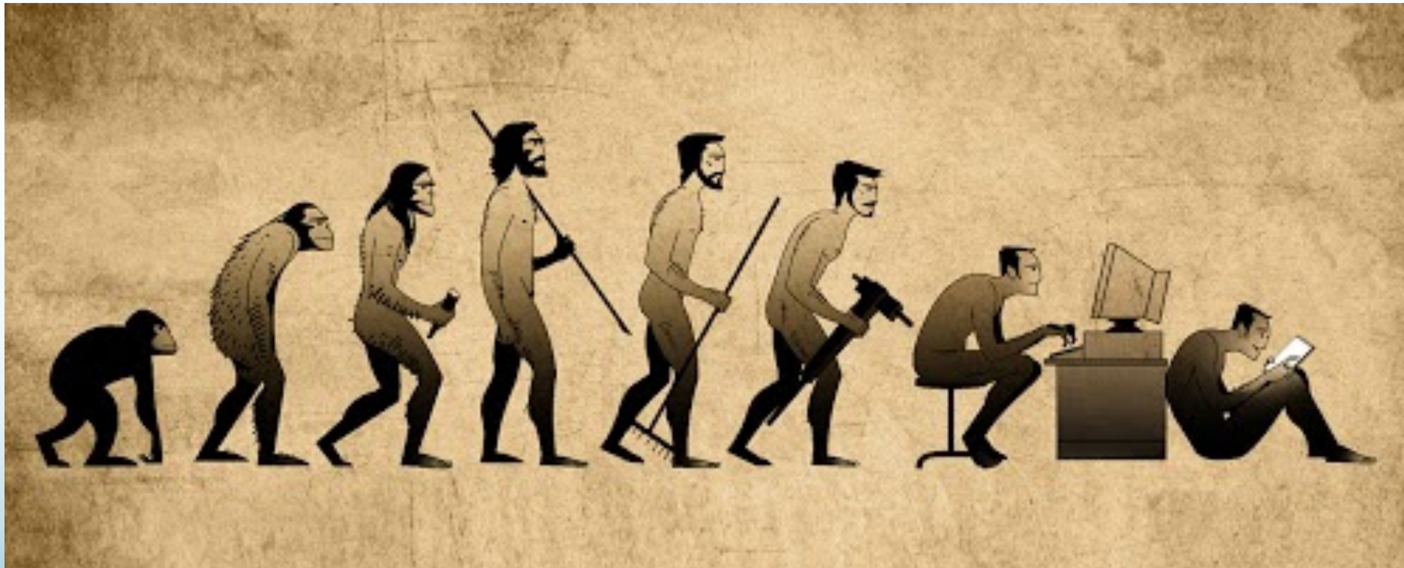
- Importantly, *homo moralis* preferences sometimes lead to behaviors which differ drastically from those predicted for other preferences:
  - Laffont [1975]: general equilibrium analysis with Kantian agents
  - Bergstrom [1995]: analysis of altruists and semi-kantians in some games
  - Alger and Weibull [2017] “Strategic Behavior of Moralists and Altruists”
  - Alger and Weibull [2019] “Morality: Evolutionary Foundations and Policy Implications” [*The State of Economics, the State of the World*, Kaushik Basu and Claudia Sepulveda, eds., MIT Press]

Concluding remarks

- Empirically relevant assumptions on preference distributions are necessary for:
  - effective policy recommendations
  - relevant welfare analyses
- Evolutionary theory helps economists achieve a deeper understanding of the ultimate causes behind preferences
- Today I have summarized research which shows that:
  - preferences may differ between populations because of exogenous factors affecting selective pressure
  - theoretical analyses can unveil novel preference classes

Concluding remarks

- Evolutionary logic may also help predict how changes in the economic environment may change our preferences in the long run!



Now and the future

- Theoretical work on preference evolution in group-structured populations:
  - Lehmann, Alger, and Weibull "Does Evolution Lead to Maximizing Behavior?" [*Evolution* 2015]
  - Alger, Weibull, and Lehmann "Evolution of Preferences in Group-Structured Populations: Genes, Guns, and Culture" [WP TSE 18-888]
- Experimental work: Alger, van Leeuwen, and Weibull: a horse race between preferences [WIP]

- ERC Advanced Grant: *EvolvingEconomics*

- A survey: Alger and Weibull "Evolutionary Models of Preference Formation" [WP TSE 18-955]

Merci !