

# MORALITY - evolutionary foundations and policy implications

Ingela Alger & Jörgen Weibull

The State of Economics, The State of the World  
Conference 8-9 June 2016 at the World Bank

# 1 Introduction

- The discipline of economics has had a huge impact on policy all over the world
- Probably due to its methodological strength and unity
  - goal-oriented individual behavior
  - equilibrium
- The goal functions (utility, profit) are key ingredients

- Since the days of Adam Smith's *On the Wealth of Nations*, economists usually assume selfishness, *Homo oeconomicus*, an opportunistic creature without morality, who takes any opportunity to enrich him- or herself

“As every individual [...] intends only his own gain [...]”

- Why this premise of pure self-interest?

- Does the assumption have empirical support? In large anonymous markets? In smaller groups?
  - Behavioral and experimental economics say “no” in the second case

- The evolutionary approach: preferences or moral values that lead to behaviors that give their carriers on average better material outcomes will spread (by biological or cultural mechanisms)
- Aim of this talk:
  - discuss some recent theoretical results concerning the evolutionary foundations of human motivation
  - examine implications for economics and policy

## 1.1 Background

- Milton Friedman (1953): “unless the behavior of businessmen in some way or other approximated behavior consistent with the maximization of returns, it seems unlikely that they would remain in business for long”
- More recently, game theorists have shown that *Homo oeconomicus* will prevail in strategic interactions if:
  - the population is large
  - individuals do not know each other’s utility functions
  - players are uniformly randomly matched

- By contrast:
  - in small populations, spitefulness arises [Schaffer (1988)]
  - if individuals know each others' preferences, *Homo oeconomicus* will not prevail [Schelling (1960), Fershtman & Judd (1987), Banerjee & Weibull (1995), Heifetz, Shannon & Spiegel (2007)]
  - likewise, if the random matching is not uniform (think of geographic location, language, culture and religion) *Homo oeconomicus* will not prevail [Alger & Weibull (2016)]

## 1.2 The rest of the talk

- Describe a new class of preferences, which we call *Homo moralis*, and their evolutionary foundation
- Discuss implications of such preferences for canonical economic policy issues
- Compare with other social preferences suggested in behavioral economics
- Conclude and sum up

## 2 Evolution and morality

1. Individuals in a large population are now and then randomly matched into groups to interact with each other
2. The group interaction may involve elements of cooperation and/or conflict, asymmetric information, repetition, reciprocation, etc.
3. Material payoffs from the interaction are *symmetric* and *aggregative* in the sense that a participant's payoff depends only on own action and some aggregate of other group members' actions
4. Each individual has a *utility function*, the expected value of which he or she seeks to maximize
5. Each individual's utility function is his or her *private information*

## 2.1 Definitions

- **Stability:** Generalizing Maynard-Smith's & Price's (1973) notion of an *evolutionarily stable strategy (ESS)*, we analyze which *utility functions*, if any, are *evolutionarily stable* in the sense that, if almost all individuals in the population have such preferences—the *incumbents* or *residents*—these individuals would materially outperform individuals with other behaviorally distinct preferences—the *mutants*—in all equilibria
- **Instability:** A utility function is *evolutionarily unstable* if there exists another utility function such that, no matter how small its population share, there is *some* equilibrium in which the latter utility function materially outperforms the former
- **Assortativity:** The *assortativity profile* of the matching process is the probability vector,  $a$ , for the events that none, some, or all the individuals in a (vanishingly rare) mutant's group also are mutants

## 2.2 Results

- Evolution favors a particular class of utility functions that we call *Homo moralis*
  - Such individuals attach some weight to their own material payoff but also to what can be interpreted as a generalized version of Kantian morality
  - A *Homo moralis* maximizes a weighted average of many terms; each term being the material payoff that she would obtain if—hypothetically—the strategies of none, some, or all the other individuals in her group were replaced by her strategy
  - We call the vector of these probability weights the individual's *morality profile* and denote it  $\mu$

- *Homo oeconomicus* is one extreme member of this class, who pays attention only to his own material payoff
- *Homo kantiensis* is at the opposite extreme, who pays attention only to her hypothetical material payoff if everybody were to use her strategy
- A whole range of *Homo moralis* preferences lie between these extremes

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

[Immanuel Kant, *Groundwork of the Metaphysics of Morals*, 1785]

**Theorem 2.1 (Alger & Weibull, 2016)** *Homo moralis with morality profile  $\mu = a$  is evolutionarily stable. Any preferences that are behaviorally distinct from those of Homo moralis with morality profile  $\mu = a$  are evolutionarily unstable.*

- The *Homo moralis* utility function is simple for pairwise interactions, and, in larger groups, under conditional independence of types; the morality profile then collapses to a single number, the *degree of morality*,  $0 \leq \kappa \leq 1$

## 2.3 Discussion

- Morality has been discussed by many economists, including Adam Smith, Francis Edgeworth, John Harsanyi, John Rawls, Ken Arrow, Amartya Sen,...
- However, to the best of our knowledge, *Homo moralis* preferences have not been discussed, or even known, before
- So how do individuals of this “new species” behave? What are the policy implications if economists’ models are populated not by *Homo oeconomicus* but by *Homo moralis*?

# 3 Morality and economics

## 3.1 Trust

- Trust varies across countries and over time, and trust is positively correlated with economic prosperity and growth [Algan and Cahuc (2010)]
- A tool for experimental study: the *trust game* [Berg et al. (1995), Cesarini et al. (2008)]
  - One individual, *the investor*, can send some money to another individual, *the trustee*
  - The sent amount is multiplied by some factor, usually three
  - The trustee decides how much of the gross return to send back to the investor

- If both are *Homo oeconomicus*, the trustee sends back no money and the investor therefore sends nothing
- Yet, in experiments the average investor sends a significant share of her endowment, and most trustees send back part of the gross return

- What will *Homo moralis* do in such an interaction?
- If both are *Homo kantiensis*, the investor will invest all his endowment and the trustee will return *half* the gross return from investment
- For *Homo moralis* of intermediate degree of morality,  $0 < \kappa < 1$ :
  - Full investment obtains for sufficiently high degrees of morality  $\kappa$
  - As  $\kappa$  falls, investment begins to fall
  - Less is paid back, and eventually, at some low  $\kappa$ , investment ceases

## 3.2 Environmental economics

- Arguably, the effectiveness of environmental policy depends to a large extent on human motivation
- Let's analyze the behavior of *Homo moralis* in an otherwise standard model of consumption with external effects [Musgrave (1959), Arrow (1970)]
  - A *continuum* of consumers, two consumption goods, one environmentally neutral and one harmful
  - The quality of the environment depends on *average* consumption of the second good, and hence each individual's impact on the environment is nil

- Each consumer derives utility from own consumption and from the quality of the environment:  $u(x_1, x_2, \bar{x}_2)$
- What happens if all individuals are identical *Homo moralis* with  $u$  as their material payoff function?

- Necessary first-order condition on each individual's consumption:

$$\frac{u_2 \left( x_1^\kappa, x_2^\kappa, x_2^\kappa \right)}{u_1 \left( x_1^\kappa, x_2^\kappa, x_2^\kappa \right)} = p - \kappa \cdot \frac{u_3 \left( x_1^\kappa, x_2^\kappa, x_2^\kappa \right)}{u_1 \left( x_1^\kappa, x_2^\kappa, x_2^\kappa \right)}.$$

$\kappa = 0$  : *Homo oeconomicus* (no concern for the environment)

$\kappa = 1$  : *Homo kantiensis* (socially optimal concern for the environment)

$0 < \kappa < 1$ : *Homo moralis* refrains somewhat from consuming the environmentally harmful good ( $u_3 < 0$ )

- Policy if  $\kappa > 0$ : combine environmental taxes with public information about effects on the environment

### 3.3 Other examples

#### 1. Democracy

- (a) *Homo oeconomicus* participates in general elections only if his expected benefit from participation exceeds his cost, and since the probability for being pivotal is small, the participation rate of *Homo oeconomicus* is low [Krishna & Morgan (2009)], and yet the actual participation rate in general elections is quite high in many countries
- (b) In committees where members have private information, *Homo oeconomicus* will typically not vote informatively even when all members have the same preferences, and thus judgment aggregation may fail [Condorcet (1785), Austen-Smith & Banks (1996)]

2. **Public goods** provision: *free-riding* and *group size* [Nosenzo, Quercia & Sefton (2015)]
  
3. **Tax evasion**: In some countries less evasion than would be rational for *Homo oeconomicus*, given the low conviction probability and relatively mild penalties [Sandmo (2005)]

## 4 Behavioral economics

- Many forms of human motivation have been proposed, and some tested:
  - *altruism* [Becker (1974)]
  - *warm glow* [Andreoni (1990)]
  - *conformity* [Bernheim (1994)]
  - *conditional altruism* [Levine (1998)]
  - *inequity aversion* [Fehr and Schmidt (1999)]
  - *identity* [Akerlof and Kranton (2000)]
  - *honesty* [Alger & Ma (2003), Alger & Renault (2007), Demichelis & Weibull (2008)]
  - *norm adherence* [Lindbeck, Nyberg & Weibull (1999), Brekke, Kverndokk & Nyborg (2003), Huck, Kübler & Weibull (2012)]
  - *image concern* [Bénabou & Tirole (2006), Ellingsen & Johannesson (2008), Falk & Tirole (2016)]

- All these models *presume* that individuals have a concern for something else than their pure material self-interest. By contrast, evolutionary theory may explain *if* and *why*, and may bring these many varieties to a common root. An avenue for future research!

## 5 Conclusion

1. Behavioral and experimental economics, other social and behavioral sciences, everyday observation, and introspection suggest that human motivation is more complex than narrow self-interest
2. The powerful analytical machinery of economics should not be abandoned because of this, but on the contrary be brought to use for agents with more complex motivations
3. Evolutionary stability theory suggests that evolution favors humans motivated by a combination of self-interest and a form of morality

4. By applying the powerful analytical tools of economics, new predictions and policy recommendations will follow, arguably leading to less emphasis on pecuniary incentives and more on information and education about the consequences of our actions for ourselves and others
  
5. Our results being purely theoretical, empirical and experimental work will be necessary to determine the empirical validity of *Homo moralis*

**Final point:** the *status of economics* in the eyes of the general public and in other social and behavioral sciences

- Economics textbooks may give the false impression that selfishness is part of economic rationality [Rubinstein (2006)]
- This misreading of economics probably hurts the reputation of economics
- If economists would instead use partly morally motivated agents, such as *Homo moralis*, then such misunderstandings could be avoided and the critique would fall flat to the ground
- Economic analysis would then not be prejudiced in favor of neither selfishness nor morality, but would allow for the whole spectrum of intermediate degrees

## 6 Post-presentation section

Q: Are different social preferences distinguishable from each other and from morality?

- A stark contrast between *altruism* and *morality* is seen in the following simple coordination game:

	<i>A</i>	<i>B</i>
<i>A</i>	2, 2	0, 0
<i>B</i>	0, 0	1, 1

- Suppose a society is trapped in equilibrium  $(B, B)$ . Consider a small group of mutants who are *altruists*, then a small group of mutants who are *Homo kantianis*