



CENTRO DE
INVESTIGACIÓN EN
COMPLEJIDAD SOCIAL

GAME THEORY AND MORAL BEHAVIOR

**Summer School 2013
Santa Fe Institute - UDD**

Carlos Rodriguez-Sickert
CICS, Facultad de Gobierno

S T R U C T U R E O F T H E T A L K

- **Evolution of homo economicus**
 - Mill's original abstraction (HE version 1.0)
 - Formal Price Theory + Game Theory (HE version 2.0)
 - Advances in the discipline (HE version 3.0)
- **Moral behavior – characterization**
- **Moral behavior – explanations**
(revisiting HE versions 1.0, 2.0, 3.0)

H O M O E C O N O M I C U S 1 . 0
T H E C R E A T U R E ' S B I R T H

HOMO ECONOMICUS 1.0 – THE CREATURE’S BIRTH

The concept was developed in his *Essay on Some Unsettled Questions of Political Economy* (1848) and full-fledged in his *Principles of Political Economy* (1848). In his *Essays*, Mill wrote:

[Political economy] does not treat of the whole of man’s nature as modified by the social state, nor of the whole conduct of man in society. It is concerned with him solely as a being who desires to possess wealth, and who is capable of judging of the comparative efficacy of means for obtaining that end. It predicts only such of the phenomena of the social state as take place in consequence of the pursuit of wealth. It makes entire abstraction of every other human passion or motive; except those which may be regarded as perpetually antagonizing principles to the desire of wealth, namely, aversion to labour, and desire of the present enjoyment of costly indulgences.

METHODOLOGICAL INDIVIDUALISM

**INDIVIDUAL ACTION IS
GOAL-ORIENTED
& ENVIRONMENTAL
CONSTRAINTS**

(John Stuart Mill 1844, Essay V, Ch. 3)

HOMO ECONOMICUS 1.0 – CHARACTERIZATION

▪ Methodological Individualism

The basic **unit of analysis** is the **individual** and not the social system as a whole (social phenomena ← aggregation of **decisions** by individuals)

▪ Individual action is goal-oriented

Choose the most efficient or cost-effective **means** to achieve a specific **end** (Instrumental Rationality in Weberian Terms).

▪ Material self-interest

Individuals are inclined to **accumulate wealth** (What is all that money for?*) but **do not like to work** (**environmental constraint**: unlimited ends, scarce means).

HOMO ECONOMICUS 1.0 – CHARACTERIZATION

▪ Social Action

Methodological individualism doesn't rule out the existence of social phenomena (atomism being an extreme case of MI)*. Take A. Smith famous quote from The Wealth of Nations:

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 interest (Smith 1776, Book I, Ch. II, Section 1)**.

▪ Reduced Scope

It is not claimed that the model applies to every domain of human behavior with the same success.

H O M O E C O N O M I C U S 2 . 0
M I L L & S M I T H ' S M O D E L G E T S F O R M A L

H O M O E C O N O M I C U S 2 . 0

MILL & SMITH'S MODEL GETS FORMAL

In a process, which starts with the work of the marginalists in the 19th century, the economic discipline has built a formal model of:

▪ **Individual Action**

(Structure of Choice)

▪ **Social Action**


(Structure of Interaction and Equilibrium Concepts)

- Price Theory (Marginalists Walras, Jevons, Menger → Arrow-Debreu)
- Game Theory (Von Neumann, Nash, Selten, Harsanyi, M. Smith)

H O M O E C O N O M I C U S 2 . 0

CHARACTERIZATION INDIVIDUAL ACTION

- The agent's choice problem involves a set of possible actions A and a set of consequences C .
- Choice is assumed to be the outcome of rational deliberation. Namely, the decision-maker has in mind a preference relation P in the choice set faced. Also, given any choice problem, the decision-maker will choose an action which leads to an optimal consequence according to P .
- Rationality is taken to require that those preferences exhibit two characteristics:
 - *Completeness*. All consequences can be ranked in an order of preference, that is, given any two alternative consequence c^0 and $c^1 \in C$, the individual either prefers c^0 to c^1 , c^1 to c^0 , or is indifferent between them.
 - *Transitivity*. The order of preference is consistent; that is to say, given any three alternative experiences c^0 , c^1 and c^2 , if c^0 is preferred to c^1 and c^1 to c^2 then c^0 must be preferred to c^2 .

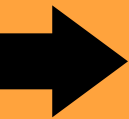


The essence of rational choice is its consistency, but the model is silent regarding the particular content of the agent's preferences.

H O M O E C O N O M I C U S 2 . 0

CHARACTERIZATION SOCIAL ACTION (PRICE THEORY)

- Price Theory: In a private ownership economy, outputs and inputs are voluntarily exchanged in markets.
- Social Action Outcome
 - GE Theory focus on existence: there is a price vector \mathbf{p} , which is said to be
 - Efficient (1st Theorem WE). However, this only holds iff
 - No Information asymmetries
 - No externalities
 - Agents are price takers ~ many agents.
- Mechanism (from individual rational choice to equilibrium):
 - Excess demand/supply model (intuitive, but not fully specified, attempt: tâtonnement stability models a la Walras, e.g., Samuelson, 1947).
 - Thermodynamic interpretation (Smith & Foley, JEDC 2008)




Expectations about other people's choice are fully condensed in the price vector, i.e., they do not involve a rationalization process of their choices.

H O M O E C O N O M I C U S 2 . 0

CHARACTERIZATION SOCIAL ACTION (GAME THEORY)

- A game is an analytical artefact which captures the essential features of a particular interaction structure, involving: Players, Strategy Space (Actions, Sequence), Preferences over Consequences (combination of strategies), Information about other player's actions and types.

Equilibrium Concepts	Mechanisms
Iterated Elimination of Dominated Strategies, Backward Induction (sequential)	Theory of Mind (Second order rationality)
Nash Equilibrium (every player plays a BR against current play), BNE (uncertainty)	In a NE, there are no deviation incentives (so mechanism is not fully specified)
Perfect Bayesian Equilibrium	Bayes-updated beliefs (when actions might convey info about types)

 Expectations about other people's choices can be the result of a rationalization process of the agent and thus involve additional rationality requirements.

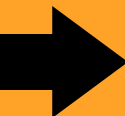
H O M O E C O N O M I C U S 3 . 0
S T A R T I N G T O L E A R N

H O M O E C O N O M I C U S 3 . 0 S T A R T I N G T O L E A R N

- **The structure of individual action coincides with that of HE 2.0.**
- **The concept of rational learning (or rational adaptation)**
 - As a justification of Nash Equilibrium.
 - As an alternative to the hyper-rationality research program on refinements of NE (~ Bounded Rationality)

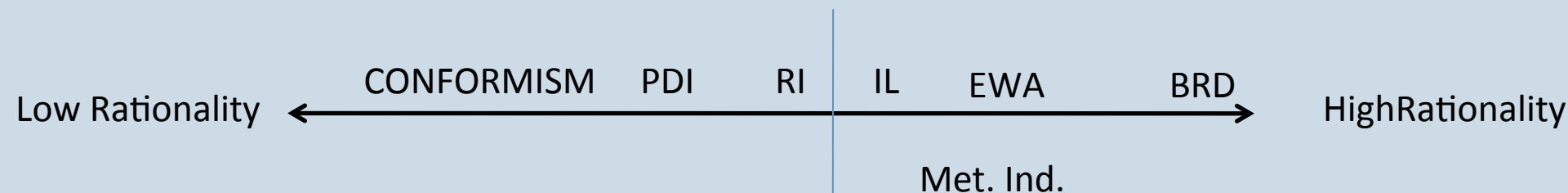
H O M O E C O N O M I C U S 3 . 0 C H A R A C T E R I Z A T I O N

- A rational learner, in its standard version, i.e., Best-response dynamics (as in Fudenberg and Levine, 1998), will start with a prior of other people's choices and as the game is recurrently played, update her beliefs and play a best-response accordingly (noise is also incorporated).
- Social Action corresponds to the social dynamics generated by the concurrent learning process of the community of agents and the notion of equilibrium relates to the convergence of this dynamical system.

 Individual action shares its structure with HE 2.0.
The mechanics which link individual action and social action are fully specified.

HOMO ECONOMICUS KEEPS EVOLVING PROLIFERATION OF LEARNING STRATEGIES

- Alternative forms of learning (in social contexts) are:
 - Conformism (Boyd & Richerson, 1984)
 - Payoff-Dependant imitation (EGT, Taylor and Jonkers 1978)
 - Reinforcement Learning (Roth and Erev, GEB 1995)
 - Inductive Learning (El Farol Problem, Arthur, AER 1994)
 - EWA Learning (Ho and Camerer, ECONOMETRICA 1999)
 - Best-response dynamics (Fudenberg and Levine, JEDC 1995)
- We can order this alternative learning strategies according to the rationality requirements each learning rule involves:



**C H A R A T E R I Z A T I O N
O F M O R A L B E H A V I O R**

TAXONOMY OF (OBSERVABLE) MORAL BEHAVIOR

	<p>Norm (Negative Social Response against deviation)</p> <p>[3]</p>	<p>Superogatory virtues (positive social response)</p> <p>[4]</p>
<p>Altruism (Individually costly pro-social behavior)</p> <p>[1]</p>	<ul style="list-style-type: none"> • Share with fairness (inequity aversion) • Honour trust (positive reciprocity) 	<p>Superogatory behavior (unconditional altruism)</p>
<p>Individually costly, no direct pro-social effect</p> <p>[2]</p>	<p>Membership norms</p>	<p>Enforcing behavior: Costly sanctions of norm deviators (negative reciprocity)</p>

**MORAL BEHAVIOR – EXPLANATIONS
REVISITING THE EVOLUTION OF HE**

H O M O E C O N O M I C U S 1 . 0

- Material self-interest → No space for moral behavior.
- Cooperation in this scheme can be achieved only via incentive-compatible schemes
 - Repeated games (long-run benefits of cooperation are larger than short-term benefits of opportunism)
 - Third party-enforcement (a la Hobbes)
- This is not to say that Mill or Smith would deny the existence of moral behavior [reduced scope in Mill's Quote, e.g., auctions or investment banking].
- In fact, Adam Smith wrote another book (TMS) which opens as follows:

How selfish so ever man may be supposed, there are evidently some principles in his nature which interest him in the fortune of others, and render their happiness necessary to him though he derives nothing from it except the pleasure of feeling it (Smith 1759, Part I, Section I, Chapter 1).

H O M O E C O N O M I C U S 2 . 0 AND MORAL CONSISTENCY OF INDIVIDUAL CHOICE

- Individual action structure in the HE version 2.0 is based on the idea of rationality as consistency.
- Andreoni and Miller (ECONOMETRICA, 2002) consider a dictator game such that the allocator s faces the following budget constraint

$$\pi_s + p \pi_o = m,$$

- varied p across treatments and rationalize their choices using a function of the form

$$U_s = u(\pi_s, \pi_o)$$

- They show that most agents satisfy the Generalized Axiom of Revealed Preferences (GARP).
- They also reported a heterogeneous structure (e.g., a significant proportion of the agents choices could be rationalized as purely selfish $U_s = \pi_s$; and among those who didn't one could find different preference structures. For instance

Rawlsian preferences: $U_s = \min\{\pi_s, \pi_o\}$

Utilitarian preferences (assuming $U_s = \pi_s$): $U_s = \pi_s + \pi_o$

 Rationality ~ Consistency → There is room for moral behavior to the extent that moral choices are consistent.

ARQUETYPICAL STRUCTURES (Sequential Games and Social Preferences)

	Anomalies	Explanations	Refinements
Dictator game (Forsythe et al, GEB 1994)	Positive allocations	Unconditional altruism (Edgeworth onwards)	Altruism and rationality (Andreoni & Miller, 2002)
Ultimatum game (Guth et al 1982)	Rejection unfair offers	Inequity aversion (Fehr & Schmidt, AER 2000)	Negative reciprocity, intentions matter (Falk & Fischbacher, EI 2001, Fowler et al 2005)
Trust game (Berg et al 1996)	Positive amount sent back	Positive reciprocity (Dufwenberg & Kirchsteiger, 2004)	Intentions matter (Cox, 2004)

ARQUETIPICAL STRUCTURES (Collective Action Problems; recurrent, $N > 2$)

	Anomalies	Explanations
Public good (Fehr & Gächter, AER 2002)	Positive contributions	Altruism, inequity aversion, positive reciprocity.
	Costly sanctioning (2 nd order public good)	Negative reciprocity (Dufwenberg & Kirchsteiger, 2004)
Common pool resources (Ostrom & Walker, 1992)	No enforcement => Erosion of cooperation	Conditional cooperation <= Self-biased inequity aversion (Fischbacher et al, 2003) Heterogeneous reciprocity (Rodríguez-Sickert, Guzmán, Cardenas, JEBO 2008)

UTILITY IS NOT FUTILITY

1. Utility function used by economists a la Mill:

$$U(c_{EGO}, c_{ALTER}) = f(c_{EGO}) \sim c_{EGO}$$

2. Experimental economists foster the emergence of a new breed of behavioral economists who are trying to fix the the utility function:

$$U(c_{EGO}, c_{ALTER}) = c_{EGO} + f(c_{EGO}, c_{OTRO}, \text{morals}_{EGO})$$

3. New problem for mechanism design: preferences are not independent from the institutional environment

H O M O E C O N O M I C U S 2 . 0 AND INTERDEPENDANCE OF MORAL CHOICES

- In Rodriguez-Sickert et al (JEBO, 2008) and Gelchich et al (2013) , we consider a common pool resource game with payoffs increasing in the agent's own extraction and decreasing in other players average extraction such that the NE among selfish agents is full-extraction and the social optimum is minimum extraction.
- We show that three types of preferences fit the experimental results:
 - Selfish agents who care only about their own material payoffs
 - Unconditional cooperators who experience guilt from deviating from a norm that precludes over-extraction.
 - Conditional cooperators whose guilt is alleviated by the aggregate level of deviation.
- In the absence of external enforcement, conditional cooperators are the modal type (creating two NE: a high cooperation NE and a low cooperation NE). Agents, in fact do coordinate in these two NE: High cooperation (after a public signal of cooperation) and in a the low cooperation NE 10 rounds later.

H O M O E C O N O M I C U S 2 . 0 AND INTERDEPENDANCE OF MORAL CHOICES

- When non-deterrent enforcement is in place, unconditional cooperators become the modal type and there is a unique NE which agents play for all 10 rounds.

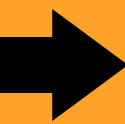
➔ Institutions influence Preferences, i.e., they are not like the Rocky Mountains (Bowles, SCIENCE 2008)

➔ The formation of expectations plays an important role in the emergence of cooperation → There is no one to one map between preferences and behavior.

H O M O E C O N O M I C U S 3 . 0

LEARNING AND THE FORMATION OF EXPECTATIONS

- Rational Learning configures the formation of expectations within a social dilemma.
- In Rodriguez-Sickert, Guzman and Cardenas (JEBO, 2008) we show that an EWA Learning model can explain the transition from the high cooperation equilibrium to the low cooperation one:
 - Violations (accidental or not), coupled with reciprocal preferences, account for the erosion.

 Rational Learning provides a mechanism to explicitly link individual moral dispositions and social outcomes.

 Morality at the individual level is still defined by the agent's preferences

H O M O E C O N O M I C U S 3 . 0

LOW RATIONALITY LEARNING RULES

- Low rationality learning rules such as conformism or payoff-dependent imitation do not involve preferences, they map directly on behavior.
- Boyd, Gintis, Bowles, and Richerson (2003, PNAS), for instance, report simulations that show that when social learning takes the form of payoff-dependent imitation, cooperation together with enforcing behavior can evolve for large groups.
- Thus, the two levels of analysis that characterizes H.E 2.0 and H.E. 3.0 are collapsed into one level of analysis: behavior.

➔ Low Rationality learning rules provide an explicit mechanism which links individual action and social action.

➔ However, we cannot define a particular agent's moral inclination anymore. The agent is immersed in the social structure.

H O M O E C O N O M I C U S 3 . 0

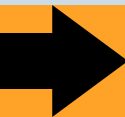
THE COEVOLUTION OF LEARNING RULES AND BEHAVIOR

- If a diversity of learning rules are available, one can assume that learning rules and behavior can coevolve.
- In Guzman, Rodriguez-Sickert and Rowthorn (2007, EHB) we consider a PGG game (as in BGBR, 2003).
 - There are two learning rules in our model: conformism and payoff-dependent imitation, which evolve by natural selection,
 - and three behavioral strategies: cooperate, defect, and cooperate, plus punish defectors, which evolve under the influence of the prevailing learning rules.
 - Group and individual level selective pressures drive evolution.
- The presence of conformists dramatically increases the group size for which cooperation can be sustained (wrt BGBR). The results are robust: they hold even when migration rates are high, and when conflict among groups is infrequent.
- In Rodriguez-Sickert, Rowthorn and Guzman (2013, mimeo) we consider an additional learning rule: Rational Learners
 - Results hold for the baseline parameters: In the long-run distribution, defectors disappear. Best responders do not invade.
 - However, for low migration rates a new long-run distribution emerges: rational learners who behave as defectors invade all groups.

H O M O E C O N O M I C U S 3 . 0

LEARNING RULES AS THE FUNDAMENTALS OF MORAL

- In a context in which learning rules and behavioral rules coevolve, again two level of analysis emerge.
- One way to think about the choice of would be in terms of learning rules as being vertically transmitted:
 - Simple Conformism: “my son, our people is wise, follow them in moral issues.”
 - Payoff dependent imitation or selective conformism: “my son, imitate those who do good, they are the respectable ones.”
 - Rational learning : “forget about the community, learn from past behavior and do what it is convenient for you my son).”
- However, in this context one should consider yet another learning rule, one in which behavior is not updated, e.g.,
 - Kantians : “my son, just do what is right and cooperate in PG games”

 In the context of diverse learning they can be interpreted as the real fundamentals of morality (replacing preferences).