

### The Death of the Economic Man

Tittenbrun, Jacek

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

**Empfohlene Zitierung / Suggested Citation:**

Tittenbrun, J. (2013). The Death of the Economic Man. *International Letters of Social and Humanistic Sciences*, 11, 10-34. <https://doi.org/10.18052/www.scipress.com/ILSHS.11.10>

**Nutzungsbedingungen:**

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:  
<https://creativecommons.org/licenses/by/4.0/deed.de>

**Terms of use:**

This document is made available under a CC BY Licence (Attribution). For more information see:  
<https://creativecommons.org/licenses/by/4.0>

# The death of the economic man

**Jacek Tittenbrun**

Chair of History of Sociology, Department of Sociology, Faculty of Social Sciences,  
Adam Mickiewicz University, Wieniawskiego 1, Poznan, Poland  
Tel. 618294000

E-mail address: jacek@amu.edu.pl

## ABSTRACT

The paradigm of rational self-interested actor is still at least implicitly assumed by conventional textbook economics as well as by many other social sciences. This is the case despite its critique on the part of even some economists. However, evidence amassed by other disciplines, such as psychology, including evolutionary psychology, anthropology and sociology unequivocally indicates the model is flawed. The paper, after presenting some historical background to the notion of economic man, reviews assorted topical research leading to the conclusion which is signalled in the title of the paper.

**Keywords:** homo economicus; rationality; self-interest; utility maximisation

## 1. INTRODUCTION

**Homo economicus**, or **Economic human**, is the concept taken for granted in orthodox, neoclassical or textbook economics. It conceives of humans as rational and narrowly self-interested agents who have the ability to make judgments towards their subjectively defined ends. The term "Economic Man" was used for the first time in the late nineteenth century by critics of John Stuart Mill. Below is a passage from Mill's work that those 19th-century critics were referring to:

*"[Political economy] does not treat 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 the comparative efficacy of means for obtaining that end".*

Later in the same work, Mill goes on to write that he is proposing "an arbitrary definition of man, as a being who inevitably does that by which he may obtain the greatest amount of necessaries, conveniences, and luxuries, with the smallest quantity of labour and physical self-denial with which they can be obtained"(1974).

Although the term did not come into use until the 19th century, it is often associated with the ideas of 18th century thinkers like Adam Smith and David Ricardo. In *The Wealth of Nations*, Smith wrote:

*"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". (1996, p. 119).*

This suggests the same sort of rational, self-interested, labor-averse individual that Mill proposes. The use of the Latin form *Homo economicus* is certainly long established; it can be traced back to Vilfredo Pareto (1906) but in actual fact that it may be older. The English term *economic man* can be found even earlier, in John Kels Ingram's "*A History of Political Economy*" (1888). *Homo economicus* bases his choices on a consideration of his own personal "utility function".

Consequently, the "homo economicus" assumptions have been criticized not only by economists on the basis of logical arguments, but also on empirical grounds by cross-cultural comparison. Economic anthropologists have demonstrated that in primitive or traditional societies, choices people make regarding production and exchange of goods follow patterns of reciprocity which differ sharply from what the "homo economicus" model postulates. Such systems have been termed gift economy rather than market economy.

Criticisms of the "homo economicus" model put forward from the standpoint of ethics usually refer to this traditional ethic of kinship-based reciprocity that held together traditional societies.

Economists such as Thorstein Veblen, John Maynard Keynes, Herbert Simon and many proponents of the Austrian school criticised *homo economicus* as an actor with too great of an understanding of macroeconomics and economic forecasting in his decision making. They stress uncertainty and bounded rationality in the making of economic decisions, rather than relying on the rational man who is fully informed of all circumstances impinging on his decisions. They argue that perfect knowledge never exists, which means that all economic activity implies risk.

Other critics of the *Homo economicus* model of humanity, such as Bruno Frey, point to the excessive emphasis on extrinsic motivation (rewards and punishments from the social environment) as opposed to intrinsic motivation. For example, it is difficult if not impossible to understand how *Homo economicus* would be a hero in war or would get inherent pleasure from craftsmanship. Frey and others argue that too much emphasis on rewards and punishments can "crowd out" (discourage) intrinsic motivation: paying a boy for doing household tasks may push him from doing those tasks "to help the family" to doing them simply for the reward.

Another weakness is highlighted by sociologists, who argue that *Homo economicus* ignores an extremely important question, *i.e.*, the origins of tastes and the parameters of the utility function by social influences, training, education, and the like. The exogeneity of tastes (preferences) in this model is the major distinction from *Homo sociologicus*, in which tastes are taken as partially or even totally determined by the societal environment (see below).

Still other critics, drawing on the broadly conceived psychoanalytic approach, questioned the *Homo economicus* model as ignoring the inner conflicts that real-world individuals suffer, as between short-term and long-term goals (*e.g.*, eating chocolate cake and losing weight) or between individual goals and societal values. Such conflicts may lead to "irrational" behavior involving inconsistency, psychological paralysis, neurosis, and/or psychic pain.

The notion of economic man is not only widespread in the social sciences and humanities, but also has readily apparent philosophical consequences, since it obviously implies a determinate view of human nature. So it is certainly worthwhile to take stock of the current state of knowledge in the field in question.

## 2. MORALITY AND TAUTOLOGY

Below there is an interesting attempt to defend the approach under consideration which curiously ends up calling it into question. Namely, the central concept of utility as a goal of an maximising activity of homo economicus is made over-inclusive which robs it of its explanatory value, turning it into an empty tautology; in the light of such views any criticism of proponents of the model of human being concerned for failure to explain human actions motivated by non-rational or non-economic considerations could be easily „refuted” by pointing out that the theory explains a broad variety of actions since humanitarian or even a madman's actions are also rational. This is an excellent example of what Leon Petrazycki termed ‘jumping’, i.e. over-reaching concept.

## 3. BLOCK QUOTE START

There is no logical or formal difficulty in incorporating somewhat moral motivations within the standard framework of utility maximisation (...). In a recent essay Christiano coined a distinction between ‘mainstream’ and ‘revisionist’ rational choice political theory, and cast us in the role of revisionist theorists (...).

One preliminary terminological note requires some attention. While Christiano refers to ‘rational choice political theory’, in this essay we will use the term ‘public choice theory’. However, we do not intend this use of terms to indicate a distinction. Indeed for the purposes of this discussion, we take the two phrases to be interchangeable. We believe that the choice between ‘rational choice political theory’ and ‘public choice theory’ is more a matter of identifying the audience than it is identifying either methodological or substantive distinctions.

A central point of relevance in relation to public choice theory concerns the specification of the content of the idea of rationality (...). Mainstream rational choice theory ‘...adheres strictly to the thesis of homo economicus. In other words, it explains the operation of institutions and justifies the reform of those institutions under the assumption that individuals normally maximise their own utility in every action they undertake (...). The identification of utility maximisation with homo economicus conflates two distinguishable aspects of the homo economicus notion – one referring to the idea of utility maximisation (or the broader structure of rationality) and the other to the precise content of utility functions (or the specific content of rationality). Nevertheless, we broadly accept the charge that the mainstream public choice position is one that emphasises a relatively narrow conception of self-interest as motivating choice in the political as well as the economic domain. By contrast, revisionist public choice theory seeks to move away from the strict conception of homo economicus (...).

While taking the broad idea of ‘utility maximisation’ to be a defining feature of the motivational structure of individuals engaged in either private or public choices, we want to argue for a rather broader concept of ‘utility’ than is typically assumed in the mainstream approach to public choice theory, where utility is often taken to be restricted to a relatively narrow concept of self-interest (...). We introduce the idea of what we there termed ‘somewhat moral motivations’. These motivations are directly responsive to normative considerations; that is, to considerations that are taken to have normative status by the individual concerned. We claim that such motivations can still be analysed in terms of desires and beliefs - the desire to act as morality (or norms) requires, and a set of beliefs about what morality (or norms) does require in particular circumstances. And, clearly, since the desires and beliefs belong to the individual in exactly the same way as do other desires and beliefs (for example, those that relate

to mainstream self-interested preferences) there is no logical or formal difficulty in incorporating these somewhat moral motivations within the standard framework of utility maximisation. We make no strong or specific claims about the precise content of these moral or normative motivations; indeed, we suggest that such motivations will be varied and somewhat unevenly distributed in the population both in terms of their precise content and in terms of their strength (relative, say, to standard preferences).

Motivational heterogeneity in the normative domain is, therefore, a hallmark of our understanding of the revisionist position. On this account, individuals are both rational and somewhat moral, with different individuals taking rather different views both on the specifics of what morality requires, and on the strength of moral claims relative to prudential or other claims. On this view, moral or normative motivations sit alongside other motivations within any individual - they do not pre-empt or otherwise dominate those other motivations but, we argue, their presence will at least sometimes make a difference to action.

To the general idea of somewhat moral or normative motivations, we add the further idea of esteem as a motivational force.

We place this idea as providing one possible link between self-interest and more normative concerns. On the one hand, my quest for esteem is clearly self-interested to the extent that it relates to my quest for others' good opinion of me as a sort of positional consumption good. On the other hand, the means by which I may earn the esteem of others will be closely related to the normative positions that the relevant others take.

I earn the esteem of relevant others to the extent that I am seen to conform to certain prevailing norms, and these norms themselves need have little to do with the direct pursuit of self-interest narrowly conceived. For example, certain acts of public engagement, or charity, are likely to be consistent with social norms that are generally endorsed and these acts are also likely to be relevant to the esteem in which I am held by others.

In this way, my self-interested desire for esteem would reinforce any more directly moral motivation I might have to conform to the relevant normative standard. Taken by itself, the idea of esteem as a motivator might be construed as part of a strategy of reducing apparently normative motivations to self-interested motivations (so as to be more parsimonious' in the account of motivation). Behaviour that may appear on the surface to be moral' or 'normative' in nature would now be explained in terms of the quest for 'esteem' as a private, self-interested good. But we resist this interpretation and favour an account in which esteem may act to reinforce or amplify normative concerns that also have direct motivational force of their own. This richer picture is preferred, in part, because it allows us to understand cases where esteem and morality come apart, as well as cases where they work in concert (...). The expressive argument recognises that there may be benefits (generalised utility) from speech acts (or equivalent acts) that merely express an opinion or view or preference, even if that speech act has no further consequences and, in particular, even if that act plays no causal or instrumental role in bringing about the preference expressed, or realising the underlying view, or persuading others of the underlying opinion, etc. In many circumstances we might expect such expressive benefits to be small relative to the more standard instrumental benefits, and we might also expect the (speech) acts that realise the expressive benefits to be entirely consistent with the acts that bring about the end for which a preference is expressed. But in at least some cases [...] these connections come apart [...]. Expressive behaviour is particularly relevant [...] where the connection between individual action and outcomes is attenuated, and these are also the circumstances in which motivational factors such as morality and esteem may well also come to the fore, simply because the force of mainstream ideas of self-interest is at its weakest in such circumstances (...). Is it still public choice theory at all? (...) We answer with an emphatic

yes. And we do so both because the strong family resemblance outweighs the novelty associated with the revisionist position, and because the seeds of (...) revisionism lie deep in the traditions of public choice theory itself.

The strong family resemblance is in part the result of a shared commitment to analytic modelling that takes the structure of individual rationality as one of its foundations. Our concern is to extend the range of explanation that the rationality approach admits. It would be absurd to construe this aim as abandoning our commitment to that approach.

In respect of the seeds of the revisionist ideas, this is not the place for detailed exegesis, but it should be clear that the idea of motivations that extend beyond the narrowly self-interested are present in the early public choice texts. Indeed, the chief arguments for restricting interest to the narrowly self-interested might be seen to be methodological and strategic rather than substantive [...].

But times and strategic priorities change. Mainstream economic analysis has now shifted to adopt the public choice position to a close approximation, and now builds increasingly sophisticated models to explain [given] phenomena by reference to the narrow self-interest of the hyper-rational individuals involved [...].

The normative criterion underlying public choice theory might be crudely categorised as a form of normative individualism. By this we mean only that whether or not some particular institutional reform is considered, on balance, to be worthwhile will depend crucially upon that reform's impact on individuals and how well their lives go. We do not mean to place any specific limits on what is included in the set of things that make an individual's life go well (or better) and, in particular, we do not necessarily restrict this set to any particular notion of 'welfare'. But we do suggest that a hallmark of normative rational choice theory is that it adopts an essentially individualistic approach to value, so that social or collective value supervenes on individual value.

#### 4. BLOCK QUOTE END (Brennan)

This is by no means the only case of what one can call an insider critique of rational choice. Linda McQuaeg writes in *All You Can Eat: The absurdity of public-choice theory is captured by Nobel Prize-winning economist Amartya Sen in the following little scenario: "Can you direct me to the railway station?" asks the stranger. "Certainly," says the local, pointing in the opposite direction, towards the post office, "and would you post this letter for me on your way?" "Certainly," says the stranger, resolving to open it to see if it contains anything worth stealing.*

It should be noted that scenarios of this type contain an implicit assumption of zero probability of the two strangers ever interacting again. With a positive probability of future interactions and the propensity of humans to use tit-for-tat type strategies (someone harms or helps you, you return the favor in kind), the optimal totally self-interested decision may be to point to the train station. Furthermore, as David S. Friedman observes, the benefit of cheating the stranger on one occasion may not be worth the mental effort of conceiving a way to do so and weighing the odds of suffering the consequences. (Sen, in fact, has participated in the development of public choice theory, in such works as *Collective Choice and Social Welfare*.)

The following case, however, refers to two, no doubt, top representatives of the current being explored. James Buchanan and Gordon Tullock outline the limitations of their methodology: "even if the model (with its rational self-interest assumptions) proves to be useful in explaining an important element of politics, it does not imply that all individuals act in

accordance with the behavioral assumption made or that any one individual acts in this way at all times... the theory of collective choice can explain only some undetermined fraction of collective action. However, so long as some part of all individual behavior... is, in fact, motivated by utility maximization, and so long as the identification of the individual with the group does not extend to the point of making all individual utility functions identical, an economic-individualist model of political activity should be of some positive worth.” (Calculus of Consent, page 29).

Similarly, it is claimed that “McChesney and the public choice theorists in general have a view of human nature (their "economic man") which is grotesquely narrow. Ask yourself why some political actors -judges, for example-are generally viewed as making decisions that are not driven by maximization of personal gain. There is no single, agreed upon answer, but many observers are impressed by the role of specialized legal training, a highly developed institutional structure, ethical and legal strictures against conflicts of interest and corruption, and the culturally instilled drive to please peers and specialized critics. In varying degrees, similar pressures apply to politicians and bureaucrats, and this makes for a far more complex model than we are offered” (Foer 1998).

It is worth seeing what is the viewpoint of a sociologist, who in his book explores the limits of public choice. In the reviewer’s opinion.

## 5. BLOCK QUOTE START

Practitioners may find that Lars Udehn's is not the right book for them. As its subtitle suggests, it is written from a decidedly sociological viewpoint (...). Critics of public choice looking for an erudite discussion of the limitations of economic methodological orthodoxy argued from a sociological perspective that scorns economic imperialism will no doubt find this book much to their liking (...). Udehn begins by calling attention to traditional political science as a field devoted primarily to description and much in need of theoretical guidance: "a discipline with a topic, but no particular approach" (p. 1). Clearly, for Udehn, economics plays a crucial role in illuminating some political phenomena, but so do the sister disciplines of psychology and sociology:

My quarrel is with those who make exaggerated claims concerning the universality and explanatory power of economic theory. The argument of this book is not that public choice is bad, but that there are definite limits to the economic approach to politics; that there are certain phenomena it is not well equipped to deal with, and some phenomena it cannot handle at all. (p. 9) (...). He does not intend to undertake an ideological critique; instead, he wishes to focus on empirical and theoretical issues. The field known as public choice is considered broadly so as to include more than simply the Virginia School originators of the term (...). Udehn invokes Buchanan's characterization of the methodology of public choice, which consists of three elements, namely, self-interest motivation, politics as exchange, and methodological individualism, discussed in turn (...). Udehn takes issue with the assumption of self-interest in politics, which lies at the core of orthodox public choice methodology. His thesis holds that political man acts with regard to both group interests and the public interest in ways that deny solely selfish motivation. Udehn does note the more recent tendency to treat the self-interest axiom broadly enough to include even altruism, but he dismisses the broader interpretation as vacuous. His charge, then, is primarily lodged against an orthodox public choice in which, without apology, self-interest is equated with selfishness.

Udehn alerts readers to empirical shortcomings stemming from the selfishness assumption. He considers a range of evidence concerning the behavior of politicians, bureaucrats, and voters (...). An economist writing on the methodology of public choice might have argued that one requires only a sufficient tendency toward rational, self-interested behavior among some political actors at the margin in order for one to ascribe economic causes to politics (...). He writes: "The point I wish to make is that some use of economic policy to increase the probability of re-election does not prove that politicians are all selfish. It only proves that they are human, perhaps all too human" (p. 69). An orthodox economist writing on methodology could acknowledge difficulties resulting from the axiom of self-interest, but would maintain the axiom nevertheless (either unflinchingly or with a wink) because to do otherwise would be tantamount to abandoning the economic approach. Preserving orthodoxy, however, is not Udehn's mission. He has come instead to highlight the need for a sociology of politics alongside public choice.

Udehn is correct that the task of explaining manifestations of altruism poses a special challenge to economists. Perhaps economists can develop better ways of dealing with other-regarding behavior (...) political man can be so much more than merely selfish; he can be socialized to care about his group interest and the public interest, too (p. 60) (...). He is concerned that the public choice movement has been about more market and less democracy (p. 188), especially as the movement has accompanied the rise of the New Right in politics, a deplorable development in Udehn's opinion (p. 8).

## 6. BLOCK QUOTE END (Schap 1997)

It is with this issue of economics vs. sociology that the following comments are concerned: "Homo economicus and homo sociologicus are the two straw men of social science.

1. For homo economicus, each social act is a considered choice, an exercise in naked self-interest. For homo sociologicus there is no choice. Man is a boat without rudder, drifting at the mercy of the powerful tides of social forces. The substantive question underlying these caricatures is the scope of instrumental behavior. Each of us at different times pursue material gain, consciously follow social custom, act reflexively from habit or tradition and respond to emotions. Weber (1947) developed a typology of action which distinguished affectual choice, seated entirely in emotional responses; traditional actions, rooted in social custom; value-rational or axiological choice, which flows from assertions of belief, regardless of consequences; and instrumental choice, behavior whose purpose is to produce particular outcomes in order to achieve specific goals. The domains of these types are fluid. What was once rational becomes customary; and what was customary or emotionally determined can be reexamined rationally.

The answer to the question of when and why individuals feel and act upon social obligations, or act in altruistic ways, is contested by these two explanatory traditions. Some rational choice theorists argue that obedience to norms is a direct expression of self-interest. Norm obedience directly generates utility. For example, Riker and Ordeshook (1968) and Tullock (1967)

Eliminate the paradox of voting by having the act of voting itself generate utility beyond its (negligible) instrumental effect on the outcome. Others, such as Ellison (1994) and Kandori (1992), view social norms as implicit long-term contracts. When social interaction is recurrent,



reciprocity is rational. These explanations are not without problems. Making norm obedience an argument of a utility function simply pushes back the explanatory problem. How did norms get there? Why do some social prescriptions become normative while others do not?

2. The long-term contracting explanations are based on what some might regard as a weakness of dynamic game theory. When social interaction is recurrent, reciprocity is rational if individuals are sufficiently patient. But there are many distinct self-consistent social formations. Nearly every outcome is rational if individuals are sufficiently patient. This is the content of the so-called 'folk theorems' for repeated games. Now the problem is pushed back to equilibrium selection. Why (or when) do we observe those equilibria which are norm-respecting and not some others? Homo sociologicus offers a different set of answers. Much effort has been expended demonstrating that commitment to norms is complementary to, rather than consequent of, instrumental behavior.

3. As put succinctly by Durkheim, 'It is therefore in the nature of society itself that we must seek the explanation of social life.'

4. I will not attempt to summarize this vast and varied literature here, but I claim that any such survey would conclude that what is being rejected by this tradition is not rational action per se, but methodological individualism.

5. Thus this tradition is not only at odds with rational choice theory; it is also in opposition to exchange theorists such as Homans, whose individualist views were developed from behaviorist rather than utilitarian principles. The methodological holist position does not preclude understanding the functionality of social norms, and thus, to some degree, their persistence, for any truly dysfunctional norm would adversely affect the future of any society that took it up. But it is less helpful in understanding the origins of norms, and functionality alone cannot explain persistence.

6. Having given up on explanations of adherence to norms that are based solely on individual incentives, some social scientists who hold to the holist tradition nonetheless recognise that the active agents in society are individuals, and that norms emerge as an aggregate of their activities. Durkheim writes: society is not the mere sum of individuals, but the system formed by their association represents a specific reality which has its own characteristics. Undoubtedly no collective entity can be produced if there are no individual consciousnesses: this is a necessary but not a sufficient condition. In addition, these consciousnesses must be associated and combined, but combined in a certain way. It is from this combination that social life arises and consequently it is this combination which explains it. By aggregating together, by interpenetrating, by fusing together, individuals give birth to a being, psychical if you will, but one which constitutes a psychical individuality of a new kind. For Durkheim, social life, and in particular, the 'collective consciousness', is an emergent property of a system of interacting individuals.

7. Evolutionary game theory offers one approach to the study of emergent properties in systems of interacting agents. The evolutionary game theory program is to replace or supplement highly detailed explorations of individual rationality in a social environment with an examination of social interactions and rough descriptions of adaptive behavior. The outcome of an evolutionary game theory analysis is a description of equilibrium as a population state, that is to say, as an emergent property of the system whose interactions are captured in the evolutionary model. As a method for modelling emergence, evolutionary game theory is one bridge across the individualism/holism gap. B. Emergence and Equilibrium Selection In the context of

evolutionary game theory, social norms are often understood as solutions to coordination problems which arise in a population of individuals large enough that most members never directly interact with one another, thus precluding the possibility of effective collective action. The process of norm emergence has been modeled in evolutionary game theory as a population of individuals engaging in multiple strategic interactions (...). Inertia is the force of habit and tradition. Only occasionally are choices rationally examined for their instrumental value, and payoff perturbations or simple noise in choice are a first pass at accounting for a non-instrumental component in choice.

The cost of focusing on interactions in evolutionary game theory models has been the oversimplification of the individual. Indeed utility maximization is a component of choice, but the optimization problem is so sparse that it is hard to say that the decision model is in any sense rational "(Blume, Easley 2008).

It is also, without doubt, interesting to see how our economic man fares in the context of field ethnographic studies: "research in experimental economics has emphatically falsified the textbook representation of *Homo economicus*, with hundreds of experiments that have suggested that people care not only about their own material payoffs but also about such things as fairness, equity, and reciprocity<sup>1</sup>.<sup>2</sup><sup>3</sup> The following thinker writes in the same vein: "In the

---

<sup>1</sup> Consider just one example of such studies: „The ultimatum game could not be simpler. Two players are allotted a sum of money. The first player, now often called the Proposer, offers some portion of the money to the second player, called the Responder. If the Responder accepts, she gets what was offered, and the Proposer gets the rest. If the Responder rejects the offer, both players get nothing. This game first attracted attention because the empirical results differed so dramatically from the predictions of game theory, which assumes self-interest. If both players are income maximizers, and Proposers know this, then the Proposer should offer a penny (or the smallest unit of currency available), and the Responder should accept. Instead, offers typically average about 30-40 percent of the total, with a 50-50 split often the mode. Offers of less than 20 percent are frequently rejected. These facts are not now in question" Cammerer 2004.

<sup>2</sup> Regarding other similar experimental results, we learn that they „all posit some form of nonselfish behaviour, since contributing to a public good is an action against self-interest"(Bardsley 2000).

<sup>3</sup> Yet another proof of the fallacy pertinent to the central building block of public choice and other theories based on neoclassical economics is provided by other experimenters' data by virtue of which

BLOCK QUOTE START the behaviors displayed by our subjects provide considerable evidence, consistent with many others in the literature,

in support of the proposition that most people are not simple maximizers of their own monetary payoffs. In particular, subjects in the A role show a clear tendency to try to do better than the self-interested equilibrium outcome by attempting to exploit the natural human reciprocity they presume to characterize a substantial fraction of people, and subjects in the B role often live up to the A players' expectations. The introduction of pre-play communication possibilities generates evidence that reciprocity is considerably strengthened by cues suggesting that it is expected. That is, the notion that in sending money A is banking on B's trustworthiness may or may not occur to B herself in a trust game without communication, but this is more likely to be evident to B when A begins with a proposed plan of action for both parties, and clearer still when A can use words to articulate her hopes. When it's harder for B to escape the knowledge that A is counting on a certain action from him, it becomes harder for B to 'stiff' A and retain a good self-image (compare Charness and Dufwenberg, 2006). It is also likely that pre-play exchange of text messages strengthens B's inclination to return money to A because it helps to transform A into a real person in B's eyes in a way that can't occur under the trading of row and column proposals only, or in the complete absence of communication. When communication was ruled out, B may even have doubted the experimenter's assertion that he is matched with a real player (who is after all said to be seated in a different room, and who B will never meet). The mere possibility of such doubt can weaken the social obligation of reciprocity. But even if B takes as given that A exists, the communication probably imparts a concreteness and individuality to A that is better at activating the impulse toward reciprocity. The implications for the possibility of 'self-committing' to agreements are at least as interesting. In traditional economics, a promise is of no account unless backed by external penalties, as for instance may happen through reputation. Yet the large majority of our subjects appeared capable of self-commitment, at least in circumstances in which modest amounts of money were on the

20th century, rational choice theory replaced what was once a substantive concept of utility with purely formal requirements: the axiomatic framework of rationality. Action was henceforth conceived of as the product of the agent's beliefs and desires directed to securing what he or she most wants, and the issue of the motives for action was ruled out of the core of the theory. Statements about agents' preferences and beliefs, which were required to account for actual behavior, were given the status of auxiliary assumptions.

There is now a wide set of experimental results showing that, contrary to the predictions of standard rational choice theory, most individuals do not solely try to maximize their self-interest but instead behave as if they were guided by social or moral motives (...). The explanation of the "puzzling facts" that experimental findings have brought into light call for theoretical elements that are deeply alien to standard game and rational choice theory. If morally driven behavior is to be seriously taken into account in economics, the ontology of the individual is to be revised to encompass the social nature of the constitution of human identity and the communicative dimension of action. The other related challenges are to theoretically integrate the deontological moral elements and the specificity of moral sentiments that undeniably influence real agents' behavior and that are not reducible to utility maximization. Putting forward utility maximization, pervasively interpreted as self-interest maximization, as normative behavior may have the regrettable effect of making it come true. As Hannah Arendt (1983, p. 401) put it: "What is particularly awkward in modern theories of behavior is not that they may be false, it is that they might become true." (Lopes 2008).

However, "this research has left fundamental questions unanswered: are such social preferences stable components of human nature; or, are they modulated by economic, social, and cultural environments? Until now, experimental research could not address this question because virtually all subjects had been university students, and while there are cultural differences among student populations throughout the world, these differences are small compared with the full range of human social and cultural environments. A vast amount of ethnographic and historical research suggests that people's motives are influenced by economic, social, and cultural environments, yet such methods can only yield circumstantial evidence about human motives. In combining ethnographic and experimental approaches to fill this gap, in reporting the results of a large cross-cultural study aimed at determining the sources of social (non-selfish) preferences that underlie the diversity of human sociality. The same experiments that provided evidence for social preferences among university students were performed in fifteen small-scale societies exhibiting a wide variety of social, economic,

---

table (a player B who received 30 and had agreed to return 20 is giving about \$2.80 of real money to an unknown individual each time he follows through, given the conversion rate of our experiment currency). Most people seem to feel at least somewhat bound by their promises, an observation that offers considerable reassurance for the possibility of CO-operation within organizations, fulfillment of obligations by trusted agents, etc. Human nature includes a complex amalgam of motivations. While self-interest may indeed be the predominant consideration to keep in mind when analyzing most economic interactions, it is not the sole operative force and considering it alone gives misleading predictions about behavior, particularly in organizations. Which economic arrangements are most viable and most able to serve the goals of human beings acting as individuals and as societies can be studied more insightfully when we are open, in our research, to including aspects of the human social persona alongside the stylized 'economic man'

BLOCK QUOTE END (Putterman 2009)

The findings of the studies reviewed in this paper are reassuring because government, non-profit organizations, and other forms of collective action including that in firms, are necessary complements to markets and profit-seeking, and their effective functioning depends on the traditional economic model being wrong in its depiction of how people behave in their dealings with one another. People must at least sometimes be trustworthy and honest, and be able to count upon this from one another, if arrangements predicated on a modicum of pro-social action are to be viable.

and cultural conditions by experienced field researchers who had also done long-term ethnographic field work in these societies. The results (...) demonstrated no society in which experimental behaviour is consistent with the canonical model of self-interest, and showed that variation in behaviour is far greater than previously thought, and that the differences between societies in market integration and the importance of CO-operation explain a substantial portion of the variation found (which individual-level economic and demographic variables could not). The results also trace the extent to which experimental play mirrors the patterns of interaction found in everyday life” (Henrich et al. 2004).

It is, too, this still relatively new branch of economics dubbed behavioural economics that in an important way contributed to weakening of the fundament of its senior discipline:

**BLOCK QUOTE START** Economics traditionally conceptualizes a world populated by calculating, unemotional maximizers that have been dubbed Homo Economicus. In a sense, neo-classical economics has defined itself as explicitly “anti-behavioral”. Indeed, virtually all the behavior studied by cognitive and social psychologists is either ignored or ruled out in a standard economic framework. This unbehavioral economic agent has been defended on numerous grounds: some claimed that the model was “right”; most others simply argued that the standard model was easier to formalize and practically more relevant. Behavioral economics blossomed with the realization that neither point of view was correct.

Empirical and experimental evidence mounted against the stark predictions of unbounded rationality. Further work made clear that one could formalize psychological ideas and translate them into testable predictions. The behavioral economics research program has consisted of two components:

1. Identifying the ways in which behavior differs from the standard model.
2. Showing how this behavior matters in economic contexts (...). Market forces and learning do not eliminate the importance of human actions.

Is Homo Economicus the Only One Who Survives? Many economists have argued that a combination of market forces (competition and arbitrage) plus evolution should produce a world similar to that described in an economics textbook: do only the rational agents survive? Or, do the workings of markets at least render the actions of the quasi-rational irrelevant? These are questions that have been much studied in the past two decades, and the early impressions of many economists that markets would wipe out irrationality were, well, optimistic.

Consider a specific example: human capital formation. Suppose that a young economist, call him Adam, decides to become a behavioral economist, perhaps because Adam mistakenly thinks this will lead to riches, or because he thinks it is going to be the next fad, or because he finds it interesting and lacks the willpower to study “real” economics.

Whatever the reason for the choice, let’s assume for the sake of argument that this decision was a mistake for Adam by any rational calculation. So, what will market forces do? Well, Adam may be poorer because of this choice than if he had sensibly chosen to study corporate finance, but he will not be destitute. Adam might even realize he could switch to corporate finance and make tons more money but is simply unable to resist the temptation to continue wasting his time on behavioral economics. Thus, markets per se do not necessarily solve the problem: they provide an incentive to switch, but they cannot force Adam’s hand.

What about arbitrage? In this case, like most we study in economics outside the realm of financial markets, there is simply no arbitrage opportunity available. Suppose a wise arbitrageur is watching Adam’s choices, what bet can she place? None. The same can be said if Adam saves too little for retirement, picks the wrong wife, or buys the wrong car. None of these irrational

acts generates an arbitrage opportunity for anyone else. Indeed, economists now realize that even in financial markets there are important limits to the workings of arbitrage. First, in the face of irrational traders, the arbitrageur may privately benefit more from trading that helps push prices in the wrong direction than from trading that pushes prices in the right direction. Put another way, it may often pay “smart money” to follow “dumb money” rather than to lean against it (Haltiwanger and Waldman, 1985; Russell and Thaler 1985). For example, an extremely smart arbitrageur near the beginning of the tulip mania would have profited more from buying tulips and further destabilizing prices than by shorting them. Second, and slightly related, arbitrage is inherently risky activity and consequently the supply of arbitrage will be inherently limited (De Long, Shleifer, Summers and Waldman, 1990). Arbitrageurs who did decide to short tulips early would probably have been wiped out by the time their bets were proven to be “right”. Add to this the fact that in practice most arbitrageurs are managing other people’s money and, therefore judged periodically, and one sees the short horizons that an arbitrageur will be forced to take on. This point was made forcefully by Shleifer and Vishny (1997) who essentially foresaw the scenario that ended up closing Long Term Capital Management.

So, markets per se cannot be relied upon to make economic agents rational. What about evolution? An old argument that individuals who failed to maximize should have been weeded out by evolutionary forces, which presumably operated during ancient times. Overconfident hunters, for example, presumably caught less prey, ate less and died younger. Such reasoning, however, has turned out to be faulty. Evolutionary arguments can just as readily explain overconfidence as they can explain appropriate levels of confidence. For example, consider individuals playing a war of attrition (perhaps in deciding when to back down during combat). Here overconfidence will actually help.

Seeing the overconfidence, a rational opponent will actually choose to back down sooner. As can be seen from this example, depending on the initial environment (especially when these environments have a game theoretic component to them), evolution may just as readily weed out rational behavior as it does weed out quasi-rational behavior. The troubling flexibility of evolutionary models means that they can just as readily argue for bounds on rationality.

The final argument is that individuals who systematically and consistently make the same mistake will eventually learn the error of their ways. This kind of argument has also not stood up well under theoretical scrutiny. First, the optimal experimentation literature has shown that there can be a complete lack of learning even in infinite horizons). The intuition here is simple: as long as there are some opportunity costs to learning or to experimenting with a new strategy, even a completely “rational” learner will choose not to experiment. This player will get stuck in a non-optimal equilibrium, simply because the cost of trying something else is too high. Second, work on learning in games has formally demonstrated Keynes’ morbid observation on the “long run”. The time required to converge to an equilibrium strategy can be extremely long. Add to this a changing environment and one can easily be in a situation of perpetual non-convergence. In practice, for many of the important decisions we make, both arguments apply with full force. The number of times we get to learn from our retirement decisions is low (and possibly zero). The opportunity cost of experimenting with different ways of choosing a career can be very high.

The upshot of all these theoretical innovations has been clear. One cannot defend unbounded rationality on purely theoretical grounds. Neither arbitrage, competition, evolution, nor learning necessarily guarantees that unbounded rationality must be an effective model. In the end, as some might have expected, it must ultimately be an empirical issue (...). Standard economic model of human behavior includes (at least) three unrealistic traits: unbounded

rationality, unbounded willpower, and unbounded selfishness. These three traits call for modification.

Herbert Simon (1955) was an early critic of modeling economic agents as having unlimited information processing capabilities. He suggested the term “bounded rationality” to describe a more realistic conception of human problem solving capabilities. As stressed by Conlisk (1996), the failure to incorporate bounded rationality into economic models is just bad economics the equivalent to presuming the existence of a free lunch. Since we have only so much brainpower, and only so much time, we cannot be expected to solve difficult problems optimally. It is eminently “rational” for people to adopt rules of thumb as a way to economize on cognitive faculties. Yet the standard model ignores these bounds and hence the heuristics commonly used. As shown by Kahneman and Tversky (1974), this oversight can be important since sensible heuristics can lead to systematic errors.

Departures from rationality emerge both in judgments (beliefs) and in choice. The ways in which judgment diverges from rationality is long and extensive (see Kahneman, Slovic and Tversky, 1982). Some illustrative examples include overconfidence, optimism, anchoring, extrapolation, and making judgments of frequency or likelihood based on salience (the availability heuristic) or similarity (the representativeness heuristic).

Many of the departures from rational choice are captured by prospect theory (Kahneman and Tversky 1979), a purely descriptive theory of how people make choices under uncertainty (see Starmer 2000 for a review of literature on non-EU theories of choice).

Prospect theory is an excellent example of a behavioral economic theory insofar as its key theoretical components incorporate important features of psychology. Consider three features of the prospect theory value function.

1. It is defined over changes to wealth rather than levels of wealth (as in EU) to incorporate the concept of adaptation.
2. The loss function is steeper than the gain function to incorporate the notion of “loss aversion”; the notion that people are more sensitive to decreases in their well being than to increases. [...].
3. Both the gain and loss function display diminishing sensitivity (the gain function is concave, the loss function convex) to reflect experimental findings. To fully describe choices prospect theory often needs to be combined with an understanding of “mental accounting” (Thaler, 1985). One needs to understand when individuals faced with separate gambles treat them as separate gains and losses and when they treat them as one, pooling them to produce one gain or loss. A couple of examples can illustrate how these concepts are used in real economics contexts. Consider overconfidence. If investors are overconfident in their abilities, they will be willing to make trades even in the absence of true information. This insight helps explain a major anomaly of financial markets. In an efficient market when rationality is common knowledge, there is virtually no trading, but in actual markets there are hundreds of millions of shares traded daily and most professionally managed portfolios are turned over once a year or more. Individual investors also trade a lot: they incur transaction costs and yet the stocks they buy subsequently do worse than the stocks they sell. An example involving loss aversion and mental accounting is Camerer et al’s (1997).

Study of New York City taxi cab drivers. These cab drivers pay a fixed fee to rent their cabs for twelve hours and then keep all their revenues. They must decide how long to drive each day. A maximizing strategy is to work longer hours on good days (days with high earnings per hour such as rainy days or days with a big convention in town) and to quit early on bad

days. However, suppose cabbies set a target earnings level for each day, and treat shortfalls relative to that target as a loss. Then, they will end up quitting early on good days and working longer on bad days, precisely the opposite of the rational strategy. This is exactly what Camerer et al find in their empirical work.

Having solved for the optimum, Homo Economicus is next assumed to choose the optimum. Meanwhile, real humans, even when they know what is best, sometimes fail to choose it for self-control reasons. Most of us at some point have eaten, drank, or spent too much, and exercised, saved, or worked too little, haven't we?

Finally, people are boundedly selfish. Although economic theory does not rule out altruism, as a practical matter economists stress self interest as the primary motive. For example, the free rider problems widely discussed in economics are predicted to occur because individuals cannot be expected to contribute to the public good unless their private welfare is thus improved. In contrast, people often take selfless actions. For instance, in 1993, 73.4 % of all American households gave some money to charity, the average dollar amount being 2.1 % of household income. Also, 47.7 % of the population does volunteer work with 4.2 hours per week being the average hours volunteered. Similar selfless behavior is observed in controlled laboratory experiments. Subjects systematically often CO-operate in public goods and prisoners dilemma games, and turn down unfair offers in "ultimatum" games (...). Twenty years ago (...) one leading economist had called the efficient markets hypothesis the best established fact in economics. Times change. Now, as we begin the 21 st century finance is perhaps the branch of economics where behavioral economics has made the greatest contributions. How has this happened?

Two factors contributed to the surprising success of behavioral finance. First, financial economics in general, and the efficient market hypothesis in particular, generated sharp, testable predictions about observable phenomena. Second, there are great data readily available to test these sharp predictions. We briefly summarize here a few examples. The rational efficient markets hypothesis makes two classes of predictions about stock price behavior. The first is that stock prices are "correct" in the sense that asset prices reflect the true or rational value of the security. In many cases this tenet of the efficient market hypothesis is untestable because intrinsic values are not observable. However, in some special cases the hypothesis can be tested by comparing two assets whose relative intrinsic values are known. One class of these is called "Siamese Twins": two versions of the same stock that trade in different places.

A specific well-known example is the case of Royal Dutch Shell as documented in Froot and Dabora (1999). The facts are that Royal Dutch Petroleum and Shell Transport are independently incorporated in the Netherlands and England respectively. The current firm emerged from a 1907 alliance between Royal Dutch and Shell Transport in which the two companies agreed to merge their interests on a 60:40 basis. Royal Dutch trades primarily in the US and the Netherlands and Shell trades primarily in London. According to any rational model, the shares of these two components (after adjusting for foreign exchange) should trade in a ratio of 60:40. They do not; the actual price ratio has deviated from the expected one by more than 35 %. Simple explanations such as taxes and transactions costs cannot explain the disparity (see Froot and Dabora). This example illustrates that prices can diverge from intrinsic value because of limits of arbitrage. Some investors do try to exploit this mispricing, buying the cheaper stock and shorting the more expensive one, but this is not a sure thing, as many hedge funds learned in the Summer of 1998 (when, at the time hedge funds were trying to get more liquidity, the pricing disparity widened).

The Royal Dutch Shell anomaly is a violation of one of the most basic principles of economics: the law of one price. Another similar example is the case of closed-end mutual funds (Lee, Shleifer, and Thaler 1991). These funds are much like typical (open-end) mutual funds except that to cash out of the fund, investors must sell their shares on the open market. This means that closed-end funds have market prices that are determined by supply and demand, rather than set equal to the value of their assets by the fund managers as in an open-end fund. Since the holdings of closed-end funds are public, market efficiency would lead one to expect that the price of the fund should match the price of the underlying securities they hold (the net asset value or NAV). Instead, closed-end funds typically trade at substantial discounts relative to their NAV, and occasionally at substantial premia. Most interesting from a behavioral perspective is that closed-end fund discounts are correlated with one another and appear to reflect individual investor sentiment. (Closed-end funds are primarily owned by individual investors rather than institutions.) Lee, Shleifer and Thaler find that discounts shrink in months when shares of small companies (also owned primarily by individuals) do well, and in months when there is lots of IPO activity, indicating a “hot” market. Since these findings were predicted by their theory, they move the research beyond the demonstration of an embarrassing fact (price not equal to NAV) toward a constructive understanding of how markets work.

The second principle of the efficient market hypothesis is “unpredictability”. In an efficient market it is not possible to predict future stock price movements based on publicly available information. Many early violations of this had no explicit link to behavior. Thus it was reported that small firms, firms with low price earnings ratios earned higher returns than other stocks with the same risk. Also, stocks in general, but especially stocks of small companies have done well in January and on Fridays (but poorly on Mondays).

An early study by De Bondt and Thaler (1985) was explicitly motivated by the psychological finding that individuals tend to over-react to new information. For example, experimental evidence suggest that people tended to underweight base rate data (or prior information) in incorporating new data. De Bondt and Thaler hypothesized that if investors displayed this behavior, then stocks that had performed quite well over a period of years will eventually have prices that are too high. Individuals overreacting to the good news will drive the prices of these stocks too high. Similarly, poor performers will eventually have prices that are too low. This yields a prediction about future returns: past “winners” ought to underperform while past “losers” ought to outperform the market.

Using data for stocks traded on the New York Stock Exchange, De Bondt and Thaler found that the 35 stocks that had performed the worst over the past five years (the losers) outperformed the market over the next five years, while the 35 biggest winners over the past five years subsequently underperformed. Follow-up studies have shown that these early results cannot be attributed to risk (by some measures the portfolio of losers is actually less risky than the portfolio of winners), and can be extended to other measures of overreaction such as the ratio of market price to the book value of equity.

More recent studies have found other violations of unpredictability that have the opposite pattern from that found by DeBondt and Thaler, namely underreaction rather than overreaction. Over short periods of time, e.g., six months to one year, stocks display momentum the stocks that go up the fastest for the first six months of the year tend to keep going up. Also, after many corporate announcements such as large earnings changes, dividend initiations and omissions, share repurchases, splits, and seasoned equity offerings, there is an initial price jump on the day of the announcement followed by a slow drift in the same direction for as long as a year or more (see Shleifer 2000).



These findings of underreaction are a further challenge to the efficient markets hypothesis, but also to behavioral finance. Do markets sometimes overreact, and sometimes underreact? (...). Underreaction appears at short horizons while overreaction appears at longer horizons (...). The prediction that investors will be less willing to sell a loser than a winner is striking since the tax law encourages just the opposite behavior. Nevertheless, Odean (1998) finds evidence of just this behavior. In his sample of the customers of a discount brokerage firm, investors were more likely to sell a stock that had increased in value than one that had decreased in value. While around 15 % of all gains were realized, only 10 % of all losses are realized.

In labor economics, experimental and empirical work has underlined the importance of fairness considerations in setting wages. For example, wages between industries differ dramatically, even for identical workers and most interestingly, even homogeneous workers (such as janitors) earn higher wages when they work in industries where other occupations earn more. In Law and Economics, we have seen the importance of “irrelevant” factors in a jury’s decision to sentence or in the magnitude of awards they give. BLOCK QUOTE END (Mullainathan, Taylor).

## 7. CO-OPERATION

The connection of the heading of this subchapter with the topic of the previous one is readily apparent: „economic analysis and virtually all game theory starts with the assumption that people are both rational and selfish. For example, predictions that players will defect in the prisoner's dilemma game and free ride in public goods environments are based on both assumptions. People are assumed to be clever enough to figure out that defection or free riding is the dominant strategy, and are assumed to care nothing for outcomes to other players; moreover, people are assumed to have no qualms about their failure to do "the right thing." (Dawes, Taylor 1988).

The predictions derived from the assumption of rational selfishness are violated in many familiar contexts. Public television successfully raises enough money from viewers to continue to broadcast. The United Way and other charities receive contributions from many if not most citizens. Even when dining at a restaurant away from home in a place never likely to be visited again, most patrons tip the server. And people vote in presidential elections where the chance that a single vote will alter the outcome is vanishingly small. As summarized by Jack Hirshleifer (1985, p. 55), " ... the analytically uncomfortable (though humanly gratifying) fact remains: from the most primitive to the most advanced societies, a higher degree of CO-operation takes place than can be explained as a merely pragmatic strategy for egoistic man" (Dawes, Taylor 1988)."

„A curious thing about of the human animal is that he is prone to sacrifice himself for non relatives who are unlikely to return the favour. This phenomenon, which can be termed gratuitous CO-operation, eludes traditional evolutionary explanations. A recent literature attempts to explain gratuitous CO-operation applying the tools of several behavioural sciences: biology, cultural anthropology, evolutionary psychology, and economics. The adjective ‘gratuitous’ derives from the Latin word *gratuitus*, meaning free, freely given, spontaneous.

The authors of a paper overviewing the existing literature and attempting to integrate the various findings in a fairly coherent framework state:

The main conclusions of our work are as follows. First: conformism (the tendency to imitate the most frequent behaviour), altruistic punishment (the tendency to punish free riders

at a cost to oneself) and gratuitous CO-OPERATION coevolve. Group selection, the natural selection of competing groups, is a necessary ingredient in this evolutionary soup.

Second: gratuitous CO-operation is not the result of a cognitive impairment. That is, people don't CO-OPERATE by mistake. The cognitive-impairment assumption, which pervades the literature on gratuitous CO-operation, is not only at odds with modern theories of the evolution of intelligence: it is an unnecessary assumption.

Third: conformism in CO-OPERATIVE dilemmas should be understood as a learned preference, instilled by parents in their children. In simple words, people are smart; if they conform, they do because they wish to follow the crowd.

Fourth: within groups, CO-operation is a cycle of sudden rise, long lasting prosperity, and abrupt decadence. Occasional "group mutations" are needed to stabilise CO-operation in the system as a whole; for example, the emergence of a charismatic leader who brings order to a group and leads it to war against other groups. Our simulations show that even if group mutations occur very infrequently, every 500 years or so, a culture of conformism, altruistic punishment, and gratuitous

CO-operation can persist in the long-run. (...). In 1986, six hundred thousand liquidators firemen, soldiers, workers, medics, and many volunteers entered the exploded Chernobyl Nuclear Plant. Exposing themselves to deadly radiation, they extinguished the fire, cleaned the area of radioactive debris, and built the concrete sarcophagus that now seals the reactor. Of the 40 firemen who were the first to deal with the disaster, the majority died within the first three months. The remaining ones are all dead today. About 60 thousand of the liquidators have died since 1986. Amongst the survivors, some 165 thousand are now chronically ill or disabled. (...). If everyone else stays and helps with the exploded plant, the family of the liquidator is saved: he obtains a benefit of  $B > 0$ .

The cost for a liquidator of staying and helping is  $C > 0$ . (...) If everyone else stays and helps with the exploded plant, the liquidator's family will be saved: better for him to run with his family and enjoy life than to expose himself to the radiation ( $B > B - C$ ). And if everyone else flees the site of the disaster, the liquidator's family is doomed no matter what; there is nothing the liquidator can do about it. Run, and he and his family will have a better chance of surviving ( $0 > -C$ ). Thus, a selfish liquidator will conclude it is best for him to flee regardless of what his fellow liquidators do. And if all liquidators are selfish, then all liquidators will flee. And yet they stayed. And by staying, they averted a catastrophe of undreamed consequences. The heroism of the liquidators is hard to explain from a biological point of view. Why should anyone kill himself (i.e., remove himself from the gene pool) for the sake of mankind? A gene that predisposed us to such an act should have gone extinct millennia ago. Kin selection, the evolution of selfish genes that move you to sacrifice yourself for blood relatives, is clearly not the answer to the conundrum.

The liquidators would have done their families a much better service by fleeing with them. Reciprocal altruism, the exchange of favours in the course of time, neither is the 2 Kin selection and the related principle of inclusive fitness were proposed by Hamilton (1964). Not really altruism, but long-term self interest. The notion is due to Trivers (1971). answer. When you are due to die within three months, there is not much time to get repaid for any favour. The explanation must lie elsewhere.

The Chernobyl dilemma is an example of what is known as a CO-OPERATIVE dilemma, a situation in which a person must decide whether or not to subordinate his own interests to those of the group. And the behaviour of the liquidators represents an extreme example of gratuitous (...) CO-operation: the sacrifice of oneself for others to whom one is not genetically related, and who will have little or no chance of returning the favour.

Three themes recur in the literature that deals with the evolution of gratuitous CO-operation: group selection, altruistic punishment, and conformism. (...). The hypothesis of group selection can be traced back to Charles Darwin himself. In *The Descent of Man*, he speculates about the development of moral faculties in humans: "It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an increase in the number of well-endowed men and advancement in the standard of morality will certainly give an immense advantage to one tribe over another. A tribe including many members who from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection." (Darwin 1871, p. 166). According to Darwin, tribes compete amongst themselves for survival, and those formed by CO-operative people drive to extinction those formed by egoists.

Modern humans appeared on earth some 200 thousand years ago. And during the first 190 thousand years they lived on the basis of hunting and gathering. Every morning, the men separated into small parties and went hunting. On return to the camp, those men who had succeeded in catching an animal shared the meat amongst all the members of the band. And, since success in the hunt depended in large measure on luck, sharing was a form of insurance that guaranteed the availability of food (almost) every day (Gurven 2004).

The behaviour of our ancestors can only be described as gratuitous CO-operation. Hunting is costly in reproductive terms: it consumes thousands of calories which are precious when one lives on the edge of starvation. A selfish hunter may plausibly blame bad luck if he returns to the campsite with empty hands. Fighting is also costly since there is a risk of death or serious injury. Nevertheless, our ancestors did their share of hunting and risked their lives fighting for the band.

The theory of group selection identifies the intense competition between bands as the explanation of our ancestors' behaviour. This account makes good sense except for one thing. It is not clear which must predominate: individual selection in favour of free riders (that is, those who do not CO-operate with the band), or the selection of groups in favour of CO-operative bands. The direction in which this tension is resolved depends, amongst other things, on the frequency of conflicts and migration rates between groups. It has been argued that migration rates among early humans were too high and conflict rates too low for group selection on its own to be effective (Williams 1966). Assuming this to be true, we are forced to discard group selection as a sufficient explanation for the existence of gratuitous CO-operation.

Altruistic punishers may enforce CO-operation, but only for a brief period of time. A second explanation for gratuitous CO-operation is based on what is called altruistic punishment (Fehr and Gächter 2002). In most human societies, to be a free rider is not really free.

The free rider may be exposed to social punishment: from disapproval to ostracism to physical aggression. If the proportion of punishers is sufficiently high, free riding may end up as more costly than CO-operating. And if that is the case, free riders will eventually disappear from the population or will become a very small fraction of it.

Punishing free riders is also costly: the main cost being the risk that free riders will defend themselves or retaliate if punished. For this reason, punishment of free riders is called altruistic: punishers sacrifice themselves to protect society from the spread of free riding [...]. Conformists prevent the spread of free riding (...) much of the time people limit themselves to repeating what the majority does. This form of learning is called conformism (...). If the number of conformists is large, the Law of Large Numbers makes it exceedingly unlikely that a substantial fraction of these conformists will simultaneously and independently innovate or

accidentally free ride. An increase on the group size, which is bad for CO-operation when people imitate profitable behaviour, helps to stabilise CO-OPERATION when people conform to the most common behaviour.

This story, just like the others, is unsatisfactory. It eliminates the question, “where does CO-operation come from?” only to raise another: “where does conformism come from?” (...) experiments in which people play some stylised version of the CO-OPERATIVE dilemma, such as the prisoner dilemma or the public good game, suggest people do not CO-operate by mistake.

Experimental subjects understand very well the logic of this dilemma; they know that free riding is the optimal behaviour from a selfish perspective. Nevertheless, most of the subjects choose to conform to the majority: they CO-OPERATE if others CO-OPERATE, and they don't CO-OPERATE if others don't. This pattern of behaviour has been termed conditional CO-operation (Fischbacher et al. 2004).

Arguably, the CO-OPERATIVE dilemma is just as easy to understand in real life as it is in the lab.

Once again, consider the case of the Chernobyl Dilemma. Sarah Wallace (2007) interviewed many men who volunteered to be liquidators. Judging from their answers, their behaviour can hardly be attributed to misunderstanding. Here are two examples.

One of the liquidators interviewed by Wallace was a physics professor in Zhytomyr Oblast, who helped with the evacuation close to the nuclear plant. Wallace reports: “He would not change his decision to be a liquidator. He thinks that people need to help each other. He does not consider what he did to be very special.” Ever since the disaster, the professor has suffered cardiovascular and neurological diseases. He attributes his condition to being exposed for so long to the radiation.

Another liquidator was a construction worker who in 1986 resided in Kyiv. Says Wallace: “He and his colleagues were asked to go to Chernobyl soon after the disaster. Their job was to seal the floor of the reactor with metal insulation and concrete so that irradiated matter would not seep into the groundwater. He agreed to go to Chernobyl out of patriotism. In his own words, if he did not do it, who would?” This man has had health problems since 1986, most likely due to the dose of 28.2 roentgens that he received while working at Chernobyl. During the interview, he declared that when he volunteered he was fully aware of the risks and that, if he could go back to 1986, he would volunteer again. (...). People understand that CO-operation is a form of self-sacrifice (the mere existence of the term “self-sacrifice” reveals how well we understand the costly nature of CO-OPERATION). Even so we choose to conform to altruistic norms. Cognitive impairment cannot account for this. The only way to make sense of the facts is to conclude that conformism in CO-operative dilemmas is an expression of preference. (...). Conformism, CO-operation, and altruistic punishment coevolve; the mixture can explain gratuitous CO-operation in large groups we simulated our model for conditions that approximate those in which early humans lived, and found that the mixture formed by altruistic punishment, conformism, and group selection was capable of sustaining gratuitous CO-operation in groups of thousands. This result obtains even when there is a high rate of migration between groups, or when conflict amongst groups is infrequent. In our model, altruistic punishment, conformism and CO-operation coevolve: none of them can develop fully without the presence of the others. ” (Rowthorn et al. 2009, pp. 1-16).

„When they are given the opportunity to punish free riders, stable CO-operation is maintained, although punishment is costly for those who punish. (...) if some people care about equity the puzzles can be resolved. It turns out that the economic environment determines whether the fair types or the selfish types dominate equilibrium behavior (...). Almost all economic models assume that all people are exclusively pursuing their material self-interest and do not care about “social” goals per se. This may be true for some (maybe many) people (...). There is, however, also substantial evidence suggesting that fairness motives affect the behavior of many people. The empirical results of Kahneman, Knetsch, and Thaler (1986), for example, indicate that customers have strong feelings about the fairness of firms’ short-run pricing decisions which may explain why some firms don’t fully exploit their monopoly power.

There is also a lot of evidence suggesting that firms’ wage setting is constrained by workers’ views about what constitutes a fair wage (Blinder and Choi 1990; Agell and Lundborg 1995; Bewley 1995; Campbell and Kamlani 1997).

According to these studies, (...) firms’ REFUSE a lot to cut wages since the fear that workers will perceive pay cuts as unfair which in turn is expected to affect work morale adversely (...). There are also many well-controlled bilateral bargaining experiments which indicate that a nonnegligible fraction of the subjects do not care solely about material payoffs (...). There is similarly conflicting evidence with regard to CO-operation. Reality provides many examples indicating that people are more CO-operative than is assumed in the standard self-interest model. Well-known examples are that many people vote, pay their taxes honestly, participate in unions and protest movements, or work hard in teams even when the pecuniary incentives go in the opposite direction. This is also shown in laboratory experiments” (Fehr, Schmidt 1999).

## 8. COMMON-POOL RESOURCES

Goods that are generally considered to be "public goods" yield non-subtractive benefits that can be enjoyed jointly and simultaneously by many people who are hard to exclude from obtaining these benefits. Common-pool resources yield benefits where beneficiaries are hard to exclude but each person's use of a resource system subtracts units of that resource from a finite total available for harvesting.

To put it differently,

**BLOCK QUOTE START** Public goods are indivisible (not in discrete units) goods and services distributed under non-competitive market conditions to consumers who, as individuals, cannot easily be excluded from enjoyment of the benefits (ex., roads). Private goods, in contrast, are goods and services which can be allotted in discrete units and distributed under competitive market conditions to consumers based on ability to pay (ex., automobiles).

Externalities. Public and private goods form a continuum, not a dichotomy. In between are goods and services which spill over into the other sector in positive or negative directions. Air pollution produced by a paper mill is a public good illustrative of a negative externality. Improved quality of life provided by a public park is a public good which illustrates a positive externality for the nearby real estate market.

Public goods such as national defense benefit all consumers within the jurisdiction of the country, regardless of the amount (if any) they pay for it in the form of taxes. A self-interested, rational, benefit-maximizing individual cannot purchase more of the public good than already received and has no incentive to try to do so. Rational individuals also have no incentive to conserve public goods held in common (hence the "tragedy of the commons" described by

Hardin (1968), in which benefit-maximizing rational individuals intensively graze their own sheep in the common pasture, harming it and leaving everyone worse off). Below we shall see whether the grim implications for voluntary action stemming from this position are indeed true. The logic of public goods and free ridership undermines voluntary collective action.

## 9. EXCLUDABILITY AND THE FREE-RIDER PROBLEM

When it is difficult or costly to exclude beneficiaries from a good once it is produced, it is frequently assumed that such a good must be provided publicly, rather than privately. When the benefits of a good are available to a group, whether or not members of the group contribute to the provision of the good, that good is characterized by problems with excludability. Where exclusion is costly, those wishing to provide a good or service face a potential free-rider or collective-action problem (Olson 1965). This is not to say that all individuals will free-ride whenever they can. A strong incentive exists to be a free-rider in all situations where potential beneficiaries cannot easily be excluded for failing to contribute to the provision of a good or service.

Public sector provision of common-pool resources or infrastructure facilities raises additional problems in determining preferences and organizing finances.

When exclusion is low-cost to the supplier, preferences are revealed as a result of many quid pro quo transactions. Producers learn about preferences through the consumers' willingness to pay for various goods offered for sale. Where exclusion is difficult, designing mechanisms that honestly reflect beneficiaries' preferences and their willingness to pay is complex, regardless of whether the providing unit is organized in the public or the private sphere. In very small groups, those affected are usually able to discuss their preferences and constraints on a face-to-face basis and to reach a rough consensus. In larger groups, decisions about infrastructure are apt to be made through mechanisms such as voting or the delegation of authority to public officials where the difficulties of translating preferences into collective choices that adequately reflect individual views are present (Arrow 1951; Shepsle 1979). In the early 1980s, the National Academy of Sciences organized a research panel on the study of common property. Ronald Oakerson (1992) wrote a framework paper for the panel that was used in the organization of a series of case studies of how diverse peoples had devised institutional arrangements related to common-pool resources (see also Thomson, Feeny, and Oakerson 1992; E. Ostrom 1992). Oakerson's presentation of the framework has influenced an untold number of studies of common-property regimes in many diverse sectors in all regions of the world. (...). When allowed to communicate or to use sanctioning mechanisms, the behavior observed in the lab is not consistent with these theoretical models but is similar to what we have observed in field settings. We have consequently developed a theory of how boundedly rational individuals use heuristics such as "measured responses" to stabilize agreements achieved in settings where there are no external enforcers to impose rules on participants (E. Ostrom, Gardner, and Walker 1994). (... The set of researchers) studied approximately 50 in-shore fisheries and irrigation systems, respectively, and were able to isolate key rules that were positively associated with higher performance levels. In *Governing the Commons* (E. Ostrom 1990), I was able to draw on the framework and on an analysis of the extensive case studies we were all reading at that time to elucidate some aspects of a theory of common-pool resources. In particular, I examined the key design principles that characterized robust, self-organized institutions for achieving sustainable resource use of very long periods of time as well as developing an initial theory of institutional change.

The second database focused entirely on irrigation systems and has been used to code more than 175 irrigation systems in Nepal (...).

We have been able to challenge many of the empirical assumptions used by development scholars who have presumed that farmers are unable to self-organize and engage in costly collective action without the imposition of rules from external authorities. We have found that farmer-managed irrigation systems in Nepal are able to outperform agency-managed systems in regard to agricultural productivity when we have controlled for factors such as size of group, length of canal, and type of terrain. The third database is an integral part of the International Forestry Resources (...) the only forests where deforestation is not extensive are where local institutional arrangements are viewed by local residents as legitimate and are monitored extensively.

(The study-note: J. T.) provides an empirical challenge to the presumption of many scholars that collective action becomes progressively more difficult as the size of the group increases from a very small face-to-face group. He shows that moderately sized villages are better able to generate the labor needed to protect local forests than are very small villages (...). Theoretical work in the public choice tradition focuses only on one arena and takes the variables specifying the situation and the motivational and cognitive structure of an actor as givens. The task of analysis is then to predict the behavior of individuals, assuming that some kind of equilibrium is likely in a fixed situation (...). When individuals face a relatively simple decision situation where institutions generate accurate information about the variables relevant to a particular problem, that problem can be adequately represented as a straightforward, constrained maximization problem.

Many of the situations of interest in understanding public choices about common-pool resources, however, are uncertain, complex, and lack the selective pressure and information-generating capabilities of a competitive market. Therefore, one can substitute the assumption of bounded rationality that persons are intendedly rational but only limitedly so for the assumptions of perfect information and utility maximization used in axiomatic choice theory (see Simon 1965, 1972; Williamson 1985; E. Ostrom, Gardner, and Walker 1994: ch. 9). Information search is costly, and the information-processing capabilities of human beings are limited. Individuals, therefore, often must make choices based on incomplete knowledge of all possible alternatives and their likely outcomes. With incomplete information and imperfect information-processing capabilities, all individuals may make mistakes in choosing strategies designed to realize a set of goals (V. Ostrom 1986). Over time, however, they can acquire a greater understanding of their situation and adopt strategies that result in higher returns. Reciprocity may develop, rather than strictly narrow, short-term pursuit of self-interest (Oakerson 1993; E. Ostrom 1998). DEPENDING UPON THE ANALYTICAL STRUCTURE OF A SITUATION and the particular assumptions about the actor used, the analyst makes strong or weak inferences about results. In tightly constrained, one-shot action situations under conditions of complete information, where participants are motivated to select particular strategies or chains of actions that jointly lead to stable equilibria, an analyst can frequently make strong inferences and specific predictions about likely patterns of behavior and outcomes.

When there is no limit on the number of appropriators from a common-pool resource or on the amount of harvesting activities they undertake, for example, one can develop a mathematical model of an open-access, common-pool resource (see, for example, E. Ostrom, Gardner, and Walker 1994). When the net benefits of harvesting to each entrant increase for the initial set of resource units sought and decrease thereafter, each appropriator acting independently tends to make individual decisions that jointly yield a deficient (but stable) equilibrium. A model of an open-access, common-pool resource generates a clear prediction of

a race to use up the resource, leading to high social costs. Both field research and laboratory experimental research strongly support the predictions of overuse and potential destruction of open-access, common-pool resources where appropriators do not share access to collective-choice arenas in which to change the open-access structure they face (E. Ostrom, Gardner, and Walker 1994).

Many arenas, however, do not generate such unambiguous results. Instead of completely independent decision making, individuals may be embedded in communities where initial norms of fairness and conservation may change the structure of the situation dramatically. Within these situations, participants may adopt a broader range of strategies. Further, they may change their strategies over time as they learn about the results of past actions. The analyst examining these more open, less-constrained situations makes weaker inferences and predicts the patterns of outcomes that are relatively more or less likely to result from a particular type of situation. In laboratory experiments, for example, giving subjects in a common-pool resource situation opportunities to communicate generally increases the joint outcomes they achieve (see E. Ostrom, Gardner, and Walker 1994 and cites contained therein). In field settings, multiple studies have shown that individuals have far more capabilities to change rules to reduce the incentives to overproduce and in many cases achieve sustainable use of renewable resources (Tang 1992; Bromley et al. 1992; Lam 1998).

In field settings, it is hard to tell where one action arena starts and another stops. Life continues in what appears to be a seamless web as individuals move from home to market to work (action situations typically characterized by reciprocity, by exchange, or by team problem solving or command). Further, within arenas, choices of actions within a set of rules as contrasted to choices among future rules are frequently made without recognising that the level of action has shifted. So, when a "boss" says to an "employee", "How about changing the way we do X?," and the two discuss options and jointly agree upon a better way, they have shifted from taking actions within previously established rules to making decisions about the rules structuring future actions. In other words, using IAD language, they have shifted to a constitutional choice or a collective-choice arena. BLOCK QUOTE END (Ostrom, OSTROM 2004)

## 10. CONCLUSION OR A SIGH OF RELIEF

This emotion refers to all too human state of mind that the reader of the present article may enjoy. After all, it turns out that this rather nasty picture of human nature as driven by egoistic motives is clearly not supported by research. Values and sentiments or emotions play at least equally important role in behaviour as reason. Even this side of human conduct is a result of its social embeddedness. This is still more apparent in the case of the second main ingredient of homo oeconomicus, i.e. self-interest. While not dismissing the results of laboratory experiments amply reported in the preceding text, it is real-life examples that have the greatest value for anyone studying this issue. And such positive instances of human altruism abound. Again, it stems from the fact that we, human beings are social entities. We are dependent on our peers for not only the material means of subsistence but also for a multitude of the ideal means of a variety of our actions, inclusive of work.

Thus, the fundamental assumption of methodological individualism pertinent to the notion of economic man as well is wrong. It is not the case that the single atoms enter into interactions, because at each moment of one's life one is involved in the web of social relations, some of which are deliberate and realised, but another portion of these relationships is of an



indirect character, i.e. they are mediated by various material and nonmaterial objects. By the same token, one becomes a member of the community far extended over one's life span, one has to do in the course of endless everyday activities with not only one's contemporaries but also with the dead. What is more, such bonds may in many cases affect our life choices, personality etc. to a greater degree than those people with whom we enter into face-to-face interactions. Suffice it to cite an example of long dead writers such as Dostoevsky, Turgenev, or Hesse who are in a position deeply to influence our world outlook, personal convictions etc.

To sum up, the fundamental reason for which any Robinson Crusoe view of members of our species is doomed to failure is the basic existential fact that for any given individual her or his biophysical and psychic nature and her/his social environment are so inextricably intertwined so as to become indistinguishable in practical terms. Whilst by no means denying that each individual is a specific structure, it must at the same time be noted that she or he constitutes an open system, constantly exchanging various inputs and outputs with her or his societal, including biophysical environment.

## References

- [1] Blume, Lawrence E., David Easley (2008). "rationality," *The New Palgrave, Dictionary of Economics*, 2nd Edition. Brennan, Geoffrey.
- [2] PUBLIC CHOICE THEORY.  
<http://www.socialsciences.manchester.ac.uk/disciplines/politics/researchgroups/mancept/workingpapers/documents/revisionistpublicchoicetheorynpe>.
- [3] Fehr Ernst, Klaus Schmidt (1999). A theory of Fairness, Competition and Cooperation *The Quarterly Journal of Economics*. August.
- [4] Foer Albert A. (1988). Review of: Money for nothing: Politicians, Rent, Extraction and Political Extortion, Author: Fred S. McChesney, Publisher: Harvard University Press (Cambridge, MA), 1997.
- [5] Henrich Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis (2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford University Press.
- [5] Hirsch Paul, Stuart Michaels, Ray Friedman (1990). "Clean Models vs. Dirty Hands: Why Economics Is Different from Sociology." In Sharon Zukin and Paul DiMaggio, eds. *Structures of Capital: The Social Organization of the Economy*: 39-56. Cambridge; New York and Melbourne: Cambridge University Press, 1990 (ISBN 0-521-37523-1)
- [6] Lopes Helena (2008). From Self-Interest Motives to Justice Motives: The Challenges of Some Experimental Results. *American Journal of Economics and Sociology*, Volume 67, Issue 2.
- [7] Mill John Stuart (1874). "On the Definition of Political Economy, and on the Method of Investigation Proper to It," *London and Westminster Review*, October 1836. *Essays on Some Unsettled Questions of Political Economy*, 2nd ed. London: Longmans, Green, Reader & Dyer, essay 5, paragraphs 38 and 48.
- [8] Putterman Louis. (2009). *Human nature, communication and trust*. *Annals of Public and Cooperative Economics*. Volume 80, Issue 1.

- [9] Smith Adam (1996). "*On the Division of Labour*," The Wealth of Nations, Books I-III. New York: Penguin Classics.
- [10] Ostrom Elinor, Vincent Ostrom (2004). The Quest for Meaning in Public Choice. *The American Journal of Economics and Sociology*. Volume 63, Issue 1.
- [11] Schap David (1997). Independent review of: *The Limits of Public Choice: A Sociological Critique of the Economic Theory of Politics*. New York: Routledge 1996.

( Received 26 August 2013; accepted 31 August 2013 )