

Economics : An Experimental Science ?

By Sylvie Thoron

In economics, the study of behavior has been enriched by a productive dialogue between theory and experiments. Challenging the model of rationality associated with *"homo economicus,"* it is shedding new light on decision-making in situations of strategic interaction.

Mainstream economics is best known to the other social sciences for its model: that of the rational and selfish actor. *Homo economicus*, as this famous model is known, makes decisions by perfectly anticipating those of other actors, evaluating risky situations to the best of his abilities and, without a second thought, maximizing his utility, which is defined by his well-being alone. Yet economists themselves have called this model into question by trying to compare it to the behavior of real people.

How did they go about it? They could have visited companies to see how executives make decisions, or homes to see how households buy their groceries. This is, to some degree, what researchers in manager sciences do. Instead, economists bring people to experimental laboratories located in universities to analyze their behavior. An experimental lab is a computer room divided into booths separated by screens. In each booth is a computer, the sole interface between the "subject" and other people. Each participant learns about a situation, described on their screen and explained with instructions, and must make choices. To test the theoretical model of the economic agent, experimental conditions must approximate the theory as much as possible. Every arrangement is intended to create conditions the experimenter can control. Experimenters must be able to work in this setting just as a biologist dissects a frog or raises drosophila populations in different conditions: the economist wants, as experimental literature puts it, to "isolate the effects." This term can be misleading. It does not refer to the observed effects of particular causes but mechanisms that economists assume to exist in order to explain certain behaviors—in other words, causes themselves. The experimental literature is extremely extensive. In this article, we will concentrate on work that focuses on so-called strategic decision-making, in which individuals know that their decisions will influence the decisions of others and vice-versa. We will begin by explaining the method that allows experimenters to shed light on some of the "effects" anticipated in such scenarios. We shall see that this work of observation and interpretation uses the strategic theoretical model of *homo economicus* as its reference point. Experimenters attempt to interpret the differences between the observed behavior and behavior predicted by the theory, leading them to reconsider their model of rationality. To understand their work, we shall see that it is useful to distinguish, within the rationality model, between the way in which economic actors think—that is, the way that they collect and process information—on the one hand, and their motivations, on the other.

Lab Experiments in Economics

From a formal perspective, game theory is used to study *homo economicus*' behavior in situations characterized by strategic interaction. The goal is to describe, as with a board game, who the players are, what rules must be followed, and how gains depend on the decisions players make. All players have access to all available information and use it to the best of their ability to maximize their utility. In order to compare the behavior of real individuals interacting with their peers, experimental literature builds protocols describing how theoretical games are actually played by lab subjects.

Consider the example of the ultimatum game, a classic that was first used as the basis for an experimental protocol by Werner Güth and his coauthors in 1982, and which has been revisited in the experimental literature many times since (see Camerer 2003). It involves a simple interaction between two individuals. Both subjects in their booths familiarize themselves with the protocol's instructions. Each learns that they form a pair with one other participant, though without knowing which one, and that they have at their disposal a small sum of money (around ten euros or so). One of two things can happen to a subject. Either the subject is the ultimatum proposer and has to offer to share the sum with their partner. Or the subject receives the sharing proposal from their partner, in which case they must accept or reject it. If accepted, both pair members receive the proposed amount. If rejected, neither gets anything. This information, which is referred to as the rules of the game, is known to both pair members. It is called "common knowledge," and the experimenter ensures this is so by reading the instructions out loud before the experiment. Another important characteristic of the experimental approach in economics is that at the end of session, subjects are, based on the results, paid in hard cash. The interest of the game lies not in the way it represents a real situation but in the way it proves a tool for dissecting and understanding the subjects' behavior. Game theory predicts that the second subject, if acting as a *homo economicus*, will never veto a proposal involving a sum that is strictly superior to zero, as payment will always be greater than if they reject the split. Consequently, the theory predicts that the first subject will offer their partner the lowest amount above zero—one cent, for example—in order to maximize earnings. What can be observed from these experiments? This game has been used as the basis of experimental protocols so often, and the same characteristics of behavior has been observed so frequently, that they are referred to as "stylized facts": in general, the first subject offers the second nearly 40% of the sum and, generally, the latter accepts, but refuses when the amount proposed is less than 20%. These results are, of course, drawn from processing very diverse statistical data, but they are clearly significant. If one compares the behavior observed to game theory's predictions, the conclusion seems irrefutable. The subject's behavior is not consistent with the theoretical model's predictions; *homo economicus* is too selfish. Is this the whole story?

Not entirely, for there are at least two categories of "effects," in the economic sense of the term, that make it possible to explain, first, why the subject proposing the ultimatum is relatively generous and, second, why their partner refuses sums that are strictly positive. In one effect category, subjects are inherently generous, care for their partners, and take into consideration social norms, values, and so on. To a sociologist or a psychologist, of course, these reasons may seem very different from one another, but not to an economist, who remains faithful to the standard model, for none of these factors is taken into consideration in conceptions of *homo economicus*.

The economic model can, however, account for a second category of effects. They make it possible to explain, at the very least, the behavior of the subject proposing the ultimatum: its effects are strategic. In this situation, the first subject was relatively generous, perhaps, because they anticipated that their partner would otherwise refuse. To make the non-strategic character of the first category clear, one must imagine a different way of "processing" the protocol, comparable in every way to the preceding one except that, in this instance, the subject cannot respond. This is known as the dictator game. The first subject no longer proposes an ultimatum, but dictates a split. This decision determines what the two pair members will receive. By comparing the ultimatum treatment to the dictator treatment, experimenters have observed that first subjects are more generous when proposing ultimatums than when behaving like a dictator, and have concluded that strategic factors nevertheless play a role. Real subjects are, of course, less selfish than *homo economicus*, but they share, to a degree, the latter's taste for strategizing.

Thus by building a protocol that describes the circumstances of the experiment and by varying how it is processed, the experimenter compares behavior and isolates different "effects." Yet the method remains essentially hypothetical and deductive. The construction of protocols, as well as the interpretation of observations, refer to a model of *homo economicus*' strategizing drawn from game theory.

A Reference Point: Strategizing Homo Economicus

For a detailed understanding of how comparisons to the theoretical model work, let us take the example of another game, which is as simple and well-known as the ultimatum game discussed above: the prisoner's dilemma, represented in the matrix below. This game consists of just two players and two strategies, namely: cooperate (C) and defect (D). Cooperation generates collective goods, but is costly for individuals. Payoff is symmetrical. Cooperation between two players results in the greatest total payoff and equal payoff for individual players. The decision on the part of both players to defect generates a situation that is inefficient, in which each layer receives an amount that is less than if both had cooperated. Two symmetrical circumstances remain in which one player cooperates and the other does not. The noncooperator benefits from the others' efforts without assuming their costs and thus receives a payment greater than what could have been earned from full cooperation. The cooperator's fate is, however, worse than when neither cooperates:

		Player 2	
		С	D
	С	8€; 8€	0€; 10€
Player 1	D	10€; 0€	2€; 2€

Prisoner's Dilemma¹

The situation in which both players defect is what game theorists call a Nash equilibrium, that is, a situation in which neither player has an interest in changing strategy, given the other's equilibrium strategy.² The Nash equilibrium in a game consists in what is known as "theoretical prediction." The other outcome of particular interest is when both players cooperate: this is the social optimum. It maximizes total gains, and is also said to be efficient in

¹ Each cell represents an outcome of the game. It is defined by the intersection of a row, the strategy of player 1, and a column, the strategy of player 2. The numbers inside the cells represents the payments the two players receive when these strategies are chosen. The first number refers to player 1's payment, the second to player 2's.

 $^{^2}$ In the prisoner's dilemma, this prediction is particularly robust, as it consists, moreover, in an equilibrium of so-called "dominant strategies": whatever the other player does, the best thing each player can do to maximize their payment is to choose defection.

Pareto's sense, as none other of the three outcomes is preferable for the two players taken together. The Nash equilibrium and the social optimum are two reference points.

It is important to understand what is surprising about this result. The theory predicts that both players will decide not to cooperate despite having an interest in doing so. These two rational and strategic *homines economici*, whose sole objective is to maximize the payments they receive, end up with a gain of $2 \in$ in the equilibrium, when they could have agreed to four times as much! This result is particularly disturbing for economists, as rationality should, by definition, be *homo economicus*' best tool for maximizing profit. Economists have thus tried to understand what prevented the players from cooperating under these circumstances.

They first observed that situations in which the strategic stakes are the same as the prisoners' dilemma are not common in economics since, most of the time, economic actors do not find themselves in such ephemeral relationships, in which decisions are taken once and for all. Specifically, the question of cooperation is raised when two actors have a relationship that persists over time, which compels them to renew their commitment, but which also allows them to back out of their decisions. Thus game theorists have proposed a new game that simply repeats the matrix of the prisoners' dilemma several times. They then show that when the game is repeated indefinitely, cooperation becomes perfectly rational. Indeed, a long-term relationship makes it possible to promote cooperation with a "threat," possibly consisting in nothing more than a return to the state of non-cooperation (the so-called "trigger" strategy).

Unfortunately, if players know that a relationship will end, everything becomes impossible. According to the rationality that is assumed in a game repeated in this way, *homo economicus* anticipates the playing out of the entire game, which they "resolve" through backward induction, starting with the end. Yet in the final stage it is no longer possible to brandish a threat because the relationship has no future. *Homo economicus* decides, in this instance, all at once, just as in the simple version of the game, and chooses not to cooperate. Once the relationship between actors is seen as being finite, there is no longer a credible threat and thus no longer any cooperation. Of course, when there is a significant number of repetitions, backward induction becomes particularly artificial. The fact remains that *homo economicus*' perfect rationality leads once again to an inefficient outcome.

Is Homo Economicus too Intelligent? Limited Rationality

It is possible to generalize the prisoner's dilemma to a game in which more than two players are involved. It becomes the contribution to public goods game. This game tries to understand the strategic stakes lurking behind real and complex problems, such as different countries' contributions to reducing greenhouse gases in the struggle against climate change, or the implementation of any system of voluntary contributions for providing a public good.

The contribution to public goods game has served as the basis for numerous experimental protocols. In this instance, it involves setting up in a lab a group of more than two subjects who are still physically isolated in their booths, but who can communicate through screens according to precise rules described by the experimenter. Each group member receives a sum at the outset of the experiment and then decides whether or not to put all or some of this sum into the group pot. Whatever does not go into the group pot is held onto privately. The instructions then explain how the group pot makes it possible to generate a surplus or public good that will later be distributed equally among its members. In formal terms, the sum of the group's contributions is multiplied by a coefficient, which varies depending on protocol and treatment, but which is always less than the number of group members. Consequently, subjects are encouraged to contribute to the group pot to produce a surplus, but since this surplus is distributed equally, irrespective of individual contributions, each subject will profit most if only other players contribute. The greatest surplus occurs when each subject puts into the pot the entire quantity they were allocated, which corresponds to the social optimum. Let us take one example with a coefficient of two and four subjects who are each given 5€. If each puts their total sum into the group pot, the public good to be distributed will be 40€ and each one will receive 10€ have only put in 5€. But if only three subjects put in 5€ and one contributes nothing, the latter, who conserved their allotment, will get 12.50€, and the others, only 7.50€.

The theory predicts that *homo economicus* will not contribute at the Nash equilibrium and that the sum of contributions is nil. Indeed, each player does not even have to think about what the others will do. Whatever they do, the best answer is to contribute nothing to the group pot. This situation is, however, very inefficient, for while each player holds onto this allotment, the group loses an opportunity to do something for the common good. We find ourselves before the same problem as the prisoner's dilemma: inefficient equilibrium.

Before turning to a few general thoughts, a few of the protocol's features should be pointed out. In general, in experiments, the game is repeated several times. In other words, at each stage, the same amount is made available to each subject, who must decide how much of it to contribute to the shared pot. The purpose of this repetition is for subjects to familiarize themselves with and understand the rules. In theory, however, as with the prisoner's dilemma, this repetition has no consequence on the behavior observed on the part of *homo economicus*, who, from the first to the last stage, does not contribute. What, on the other hand, about the experiment subjects? In this instance, too, extensive observations amount to genuine stylized facts, in contrast to theoretical predictions. Initial contributions are, generally speaking, around 40 to 60% of the allotment, before declining in a way that brings them closer to the theoretical result.

According to the first interpretation, the subject's rationality is limited vis-a-vis *homo economicus*'. They are not capable of backward induction as described above and thus they reason in the repeated game as before an infinite horizon. This interpretation could explain why their

behavior comes closer to theoretical behavior as the experiment's end approaches and the experiment's horizon becomes clearly finite. The subjects would thus seem to be the victims of shortsightedness that does not allow them to reason rationally until they are close enough to the experiment's conclusion, but this limitation of their rationality brings them closer to the social optimum.

Some economists have sought to eliminate the strategic effects of repeated games while giving subjects the possibility of practicing the protocol's rule a few times in order to master them. Then how can one avoid distorting the nature of the game by repeating it several times? They proposed comparing two treatments. In the first, known as the partners' treatment, the group of subjects remains the same for the experiment's entire duration; in the second, known as the strangers' treatment, the group is reconstituted at each stage. The underlying principle is that game strategies can only be implemented amidst a group of partners. Threats cannot be a deterrent without a long-term relationship. It is thus to be expected that the subjects of the strangers' treatment contribute less.

Yet the results leave *homo economicus* perplexed. In both treatments, the previous pattern persists. Contributions are at first relatively generous, then they drop, reaching low levels by the experiment's end. It would clearly seem that the decline is less significant in the case of strangers' treatment than in the partners' treatment, but the subjects in the strangers' treatment ultimately contribute a good deal more than subjects in the partners' treatment (Ledyard 1995). None of this seems to make any sense. Interpretations based on the strategic cooperation of *homo economicus*, even when they assume limited rationality, do not work, since subjects cooperate even better when they are not involved in ongoing relationships.

Reconsidering How Rationality is Defined

Intrigued by these results, experimenters began examining individual data more closely. Rather than attempting to detect average and representative behavior, they discovered a great diversity of types of behavior. Thus the literature elaborated a different kind of interpretation, which maintained that subjects' rationality is not limited compared to that of *homo economicus*, but that they subscribe to different kinds of rationality. In other words, there is not just one kind of rationality, which is assumed to allow an optimal handling of the information individuals require to maximize their utility, but several. Rather than interpret this diversity as evidence of behavior that is "deviant" vis-à-vis an unsurpassable rational norm, experimenters have begun to study heterogeneous behavior as endowed with its own logic and to establish typologies on this basis.

In experiments based on the contribution to public goods game that interest us, three broad categories of subjects emerge. Opportunists, who behave like *homo economicus*, are the

minority. Subjects who, to the contrary, put almost all of their allotment into the group pot, regardless of the actions of others, are also a minority. There are, finally, conditional contributors, who initially make large contributions but whose generosity then becomes contingent on that displayed by other participants at earlier stages. This typology suggests an entirely different interpretation of the stylized facts discussed above. According to this interpretation, contributions made at the sequence's outset are the average of the greatest contributions made by the unconditional and the conditional contributors and the lesser or noncontributions made by the opportunists. Then, as the conditional contributors observe that total contributions are falling below their expectations, they begin to cut their own contributions, triggering a dangerous domino effect. Because they do not resort to threats but look to the past, conditional contributors will be no more optimistic with a group of partners than with a group of strangers. Their declining contributions are not a punishment inflicted on insufficiently cooperative partners but evidence of their efforts to adapt to circumstances. Thus in the strangers' treatment, some conditional contributors manage to become optimistic in light of their knowledge that the group's composition will change and contribute more than at the previous stage, slowing down the shared pot's decline.

This interpretation based on different kinds of rationality recalls a very different kind of experiment conducted in the 1980s by the American political scientist Robert Axelrod (1984). He launched a competition among researchers in different fields-political scientists, economists, game theorists, mathematicians, and others-asking them to propose strategies for playing the prisoner's dilemma. If the theory holds that the equilibrium of dominant strategies should be robust, as players do not need to know the strategy of their competitors to make their own choice, it is because it rests on a strong hypothesis: that everyone reasons in the same way. Axelrod initiated his competition because he had doubts about this robustness. Each participant submitted their strategy as a mini program. Axelrod paired each strategy with another proposed strategy, randomly selected from all the proposals, with which it would "play" the prisoner's dilemma a hundred times. For each strategy, he repeated this operation dozens of time, always connecting it to a randomly selected competitor. In this way, he calculated a score based on a strategy's average gains. The winning strategy was a very simple one, proposed by the political scientist Anatol Rapoport and dubbed "Tit for Tat." Tit for Tat always begins by cooperating then, at each stage, does what the competitor did at the previous one. Homo economicus wins against Tit for Tat (barely) when they play alone, but Tit for Tat does far better when it confronts an array of different strategies.

Axelrod's experiment made a case for the development of a new approach to situations of economic interaction, an alternative to standard game theory, known as evolutionary game theory. According to this approach, actors need not be as rational as *homo economicus*. They can be seen, rather, as automatons with very simple yet different strategies. Evolutionary games are based on methods constructed by analogy with Darwinian natural selection. This approach was, incidentally, first elaborated in the 1970s in biology by John Maynard Smith (Maynard Smith 1982), before being imported into economics.

The goal of this approach is to know how, among a diverse population of agents, the proportions of these various strategies will evolve. According to one method, a diverse population of individuals is considered. Each one has a fixed strategy. Individuals in this population meet randomly on many occasions. They are "reproduced" in proportion to the utility they draw from these interactions. This process, which clearly references the idea of natural selection, is known as the "replicator dynamic." This dynamic can, over the long term, converge with a situation of fixed proportions between various strategies.

Another method consists in imagining the existence of a prevailing strategy that everyone has adopted. The initial population is, in this case, homogeneous. The question is whether such a situation is stable. "Mutants," who follow alternative strategies, are introduced randomly into this population at each period. A strategy is said to be evolutionarily stable if a mutant strategy is unable to invade the initial population. In other words, individuals following the mutant strategy who are introduced into the population cannot achieve greater utility than those who continue to employ the old strategy. With this approach, rationality is thus transferred from the individual to society. In contrast to the methodological individualism favored by economists and despite significant technical limitations, this theoretical literature continues to develop in parallel to studies devoted to individual rationality. The latter have, moreover, undergone additional upheavals, as we shall see.

Is Homo Economicus too Selfish?

Alongside the development of these considerations of rationality in the sense of reasoning and anticipation, other economists have dwelt on subjects' motivations. They have proposed to modify the standard model of *homo economicus* by modifying nothing more than the motivations attributed to him, leaving the standard game theory model's hypotheses about his reasoning habits intact. Thus the theory of social preferences acknowledges the fact that individuals behave as if they evaluated other participants' payoffs positively or negatively. There are several ways to modify the standard model to incorporate this interpretation.

The simplest and best-known model is the so-called inequality aversion model proposed in 1999 by Ernst Fehr and Klaus Schmidt. The authors proposed to modify *homo economicus*' utility function, which is entirely based on material payoff, by integrating a single element, the difference between the individual's payoff and that of another individual or of an average of individuals. The utility function is thus the weighted sum of a player's material payoff and of the "social" component that is inequality aversion. Thus an individual's utility does not necessarily rise with payoff if the latter becomes far greater than that of their partner or partners.

Diversity in this model resides in different levels of sensitiveness to inequality between individuals. Actors are, however, always more sensitive to inequalities that disadvantage them than those that advantage them; they value equal payment but are more or less slightly biased in their own favor. Fehr and Schmidt have shown that their model makes it possible to faithfully reproduce the stylized facts derived from the ultimatum protocol. The model was then used to explain behavior observed in other games that guarantee victory. It is now safe to say that this model has proved phenomenally successful in economic literature. Of the articles published in the *Quarterly Journal of Economics* since its creation, Fehr and Schmidt's is the fourth most cited. Yet the productive dialogue between experiments and theory does not stop here.

Only three years after the publication of Ferh and Schmidt's article, a group of experimenters, Cherry et al. (2002), published new results that called into question the famous stylized facts that, it was believed, had been explained. To the traditional protocol based on the dictator's game, the authors proposed to add a first stage, during which the dictator must accomplish one task in order to earn the sum that they can then share in the second stage. It would appear that dictators become very selfish at this stage and their behavior is once against consistent with the dominant theory's predictions. Yet this article does not endorse this theory, but demonstrates that the explanation in terms of inequality aversion is probably too simple. The experiment makes it possible to understand that the protocols that form the basis of the vast experimental literature on ultimatum games does not see the origin of the sum to be distributed as significant. In other words, in protocols as well as in utility functions, money has no smell. This hypothesis, the authors conclude, should have been called into question along with that of actors' selfish behavior.

A new experimental literature began to emerge, based on a model protocol comprising two stages: a first stage in which, unlike the Cherry et al framework, all subjects participate in earning the sum that will then be distributed in the second stage. It becomes clear why inequality aversion was considered relevant, since the failure to include a production phase meant that both subjects were effectively given equal rights. Equal sharing is thus the principle best suited to this context. Consequently, a different and more general interpretation has been proposed, which holds that subjects refer to principles of justice (Cappelen et al. 2007, Rodriguez and Moreno 2012). But how do principles of justice influence behavior? The literature offers different answers, some of which are contradictory. Thus according to Cappelen et al (2007), individuals are defined by ideals of justice that guide them in their choices. For Rodriguez and Moreno (2012), on the other hand, individuals do not really have ideals and, like a good moral *homo economicus*, they choose, based on circumstance the principle of justice that will maximize their utility. Is it still possible, in light of these experiments, to claim that principles of justice modify motivations alone?

Misperceptions about Economics

It would be a mistake to believe that economics rests on a single and uncontested model of rationality. The homo economicus model undoubtedly still has a bright future ahead of it, notably because it is the foundation of macroeconomics, the branch of economics that is most often used to advise public decision-makers. Yet behavioral studies is a dynamic research field that explores highly varied trails and thrives on a productive dialogue between theory and experiments, yet without renouncing the hypothetical-deductive approach that economists are fond of. The well-known game theorist Ariel Rubinstein wrote in 2001 that "experimental economics has become economic orthodoxy." This literature has indeed gained stature in economics by establishing itself as a homogeneous sub-discipline. It has its rules, such as paying subjects or preventing them to communicate in order to ensure better control. It has its method, based on comparing observations and theoretical predictions. Yet new research is already occurring beyond the beaten paths, which occasionally diverges from the elegant, newly established disciplinary edifice. This new research seeks to enrich experimentation by organizing it in the field, beyond the lab, accepting the risk that not every control requirement will be satisfied. They test more inductive methods and no longer consider the homo economicus model as their reference point. They seek points of comparison in other fields, such as psychology and neurosciences. It is undeniable that these developments and the ongoing renewal of experimental and behavioral economics have, at present, profoundly modified economic theory and finance.

Further reading :

- Axelrod, Robert, (1984), *The Evolution of Cooperation*, New York, Basic Books.
- Camerer, C., (2003), Behavioral Game Theory, Princeton, Princeton University Press.
- Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen and Bertil Tungodden, (2007), "The Pluralism of Fairness Ideals: An Experimental Approach", *The American Economic Review*, Vol. 97, No. 3, pp. 818-827.
- Cherry, Todd L., Peter Frykblom and Jason F. Shogren, (2002), "Hardnose the Dictator", *The American Economic Review*, Vol. 92, No. 4, pp. 1218-1221
- Fehr, Ernst, and Klaus M. Schmidt, (1999), "A Theory of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics*, 114(3): 817-68.
- Güth, W., Schittberger R. et Schwarze B., (1982), "An Experimental Analysis of Ultimatum bargaining", *Journal of Economic Behavior and Organization*, vol. 3, p. 367-388.
- Konow, James, (2003), "Which Is the Fairest One of All? A Positive Analysis of Justice Theories", *Journal of Economic Literature*, Vol. 41, No. 4, pp. 1188-1239.

- Ledyard J. (1995), "Public Goods: A Survey of Experimental Research", in Kagel, J. and Roth, A. (eds.), *The Handbook of Experimental Economics*, Princeton, Princeton University Press.
- Maynard Smith, J. (1982), *Evolution and the Theory of Games*, Cambridge, Cambridge University Press.
- Nowak, M.A., (2006), *Evolutionary Dynamics*, Harvard, The Belknap press of Harvard University Press.
- Rodriguez-Lara, Ismael and Luis Moreno-Garrido, (2012), "Self-interest and fairness: self-serving choices of justice principles", *Experimental Economy*, 15, p. 158–175.
- Rubinstein, Ariel, (2001), "A Theorist's view of Experiments", *European Economic Review*, vol. 45, p. 615-628.

First published in laviedesidees.fr, October 21, 2016. Translated from the French by Michael Behrent with the support of the Institut Français

Published in Books & Ideas, March 27, 2017.