

Evolutionary game theory and the emergence of morality

A joint interview with Jörgen Weibull and Jean-François Laslier

Jörgen, first of all, can you tell us a little bit about your research in game theory?

Jörgen Weibull: When I approached game theory many years ago I was delighted to find such a beautiful machinery for analyzing so many things. And I've tried to extend the methodology of game theory itself, and also to use it. One area is political economy, where we try to understand movements in the political sphere, how parties position themselves and how voters behave. Another area which I've worked on quite a lot is called evolutionary game theory; it is in line with Darwinian theory of natural selection, but it can be extended to cultural evolution. This branch of game theory has also become much used in computer science, and nowadays researchers use many of the methods from

evolutionary game theory when they run algorithms in artificial intelligence.

Jean-François, do you want to add anything or react to this presentation by Jörgen?

Jean-François Laslier: Game theory is mainly a mathematical set of results, and it's widely used by people doing theory in different disciplines, including economics. It has specific links with the study of rationality, and that's a strong point in economic theory because this discipline is taking seriously the rationality of various agents, but it has also a huge importance in the study of fields where rationality is not at stake, or at least is very different from economic rationality, for instance in biology. And so evolutionary game theory has been developed by people who were initially working in biology

or at least in theoretical biology, and Jörgen's work in evolutionary game theory is partly within this framework: ideas that come from evolutionary biology but are meaningful for the abstract and general theory of evolution, for any evolving system.

Jörgen, one of your last papers was about this very point, the use of game theory for studies related to evolutionary theory and social and moral philosophies. You speak of «Homo Moralis». Can you explain this work?

Jörgen Weibull: One reason I've been working in evolutionary game theory is that it seems that many questions in social science may be better phrased in evolutionary game theory terms than in standard game theory terms. Here I am thinking about rationality and the many times you have to limit the

interpretation so that you only think of a few interacting agents, with full rationality; while in evolutionary game theory, the setting is more like a population or a large group of boundedly rational individuals who now and then randomly meet to trade or do things together. We try to understand how that evolves in the population. The question that one typically asks is the following: Is a certain habit or norm for how people deal with things, for example how you buy or sell things, a stable state of affairs? What happens if a few individuals in the population start to behave in a different way, either by chance, intention or mistake? Will they be successful or not, in terms of material outcomes, in that society? Once you begin to see things like this, the next question is: Are there behaviors that are stable in the sense we just defined, that is, even if a few individuals would switch to some other behavior, they would not outcompete those who are residents, who play in “the normal” way?

Now, arriving at your question on moral philosophy, what I have done in recent years, in collaboration with professor Ingela Alger (CNRS, IAST and TSE) is to change the focus from evolution of behaviors to evolution of preferences, or, if you like, value systems. Imagine that evolution takes place on that sphere. It is then as if nature endowed us not only with cognition and perception, but also with values, tastes and preferences, and delegated to

us to choose behaviors adapted to the interaction at hand. While traditional evolutionary game theory looked at behaviors – asking which strategies will be evolutionarily stable – we ask which preferences or values will be evolutionarily stable.

That project has given some new results which were surprising to us and also, I think, to others. By phrasing the question “What preference systems are stable in an evolutionary sense”, we wonder if they would be resistant against “invasion” of other values or other preferences in the population. We phrase this question in a mathematical way inspired by classical evolutionary game theory. Without any preconception about the answer, we just asked what evolutionary stability of preferences would lead to. We let the mathematics decide. First, we got long and complicated equations that we initially didn’t understand, but, finally, we could see that the equations boiled down to simple expressions, containing two terms. One term can be interpreted as representing pure self-interest, to do what is in an individual’s best material interest, without regard to others. The other term that came out in these equations can be interpreted as a representation of Immanuel Kant’s categorical imperative in moral philosophy. This categorical imperative says that you should only act in such a way that you would wish that your behavior could become a general rule for what people should do in that situation.

For example, I shouldn’t throw garbage in the streets unless I think it would be a good thing if everybody would throw garbage in the streets. This deontological moral reasoning is very different from standard economics reasoning. And it turns out it has quite strong impacts on behavior and can actually explain, I think, some of the puzzles and paradoxes in economics. For example, why do people behave in environmentally friendly ways, even when their individual behavior is too insignificant to have any noticeable affect the environment (say, the climate)?

This line of research is an example where mathematics forces us to new and sometimes surprising insights. In this case, it is the logic of the mathematical definition of evolutionary stability of preferences that decides. And then you try to understand what it leads to. It’s quite exciting, I think.

Indeed, it is all very exciting and makes us see things differently. Jean-François, maybe you want to add some ideas to the theories developed by Jörgen?

Jean-François Laslier: This is a wonderful example of how the analytical strengths and analytical precision of mathematical thinking can find links between things that are apparently quite distant and thereby touch to many different domains. In particular here, it brings these tools to the philosophical reflection, philosophical thinking about

what is morality, what is self-interest, what are preferences, and so on and so forth. This may be not intended at first, but is a very striking feature of this intellectual story.

Jörgen, to conclude, would you have another example of the use of game theory to share with us?

Jörgen Weibull: Actually, there is also another theme distinct from this, which I think also is fascinating, and that's trying to understand political life. Especially if we look at representative democracies. Then it appears important to better understand our democratic institutions, their stability and whether they are able to aggregate the values and the judgments of the voters in a good way. And apparently, many contemporary democracies have difficulties with this. It might be that some voters feel alienated and don't think that there is any political party that represents their viewpoint and interests.

Together with a Ph.D. student in Stockholm, Atahan Afsar, and Jean-Francois, we have an ongoing research project where we allow for two or

more political parties that may position themselves in multiple dimensions, not only on the traditional left-right political spectrum, but also on a spectrum from liberal to authoritarian, another spectrum from environmental or non-environmental, and so forth. We try to understand the interplay between several arenas: competition between politicians and political parties for the citizens' votes, and coalition formation between political parties in parliament: which government formations would be stable and which would not? And here, game theory again comes in. There are some wonderful tools in game theory which can be used. One set of tools is from John Nash. If you remember *The Beautiful Mind*, the film about John Nash, who got the economics Nobel Memorial Prize in 1994. He laid the foundations for an approach, called noncooperative game theory, which is excellent for analyzing competition between political parties for votes. But there is also another Economics Prize laureate, Lloyd Shapley, mathematician and game theorist from the US, who developed ideas how to think about coalition formation,

more precisely, who will form a coalition with whom, and how will they divide the power. In the present context, the concern is about political parties who first compete for the citizens' votes in an election, then negotiate after the election to form a compromise government. This is an important and fascinating process that one can see at work in many countries that use proportional representation. Coalition formation is a typical subject of cooperative game theory. These tools, from cooperative and non-cooperative game theory, can be combined in the analysis of political economy, and researchers are working on this. In this project, we try to contribute with some new insights, to better understand the multi dimensionality of ideologies, and also the potential alienation and abstention of voters. When people don't vote because of alienation, we get unrest and instability in society. I think it's an important role for us, as researchers, to help as best as we can to understand our democratic institutions, and maybe also come up with suggestions for how to improve them.

Antoine Beillevaire, responsible for promoting the Chairs at PSE, conducted the interview.

- ★ **Jörgen Weibull** is professor at the Stockholm School of Economics.
- ★ **Jean-François Laslier** is professor at the Paris School of Economics and senior researcher at the CNRS.