

## **When the lottery is fairer than rational choice**

### **Interview with Jon Elster**

By Florent GUÉNARD & Hélène LANDEMORE

**Jon Elster explains why he has gradually distanced himself from rational choice theories to favour chance and the luck of the draw, less liable to be unjust when deliberation is plainly impossible. Video interview.**

**La Vie des idées : Jon Elster, you are a professor at the Collège de France where you hold the chair of Rationality and Social Sciences. You are renowned for your work on the theory of rational choice, notably developed in French in your essay *Le Laboureur et ses enfants* (The Ploughman and his Children). Once an exponent of the theory of rational choice, you have grown more critical over the years and notably so in your latest book on the philosophy of social sciences, *Explaining Social Behaviour*, where you go so far as to describe some versions of this theory of rational choice as pure science fiction. How would you explain your evolution on this issue?**

**Jon Elster :** Well, there has indeed been an evolution, but perhaps not as sharp as you might think, since, if you take a look at the subtitles of my three books in English on the subject, the first includes the word “irrationality”, the second the formula “subversion of rationality” and the third the phrase “limitation of rationality”. Thus from the outset, I was very aware both of irrational behaviours and of the inadequacies of the theory of rational choice. But as time went by, I became more and

more aware of two phenomena: on the one hand, people behave irrationally' and this to a surprising degree considering that we were able to land a man on the moon (which is undeniably an achievement of rationality) and on the other hand the theory of rational choice is very often unspecific insofar as it does not just prescribe what the agent must say (or do). And it is precisely in this context that I refer to science fiction since some practitioners of rational choice theory appear to assume that agents have an almost infinite capacity for instant complex calculations such as these authors themselves perform over pages of mathematical appendices. There we have science fiction: the notion that people are pocket calculators with an almost infinite capacity for the instant resolution of differential equations.

**La Vie des idées :** To the point that you have been highly critical of the very hypothesis of rationality, and, in your lectures at the Collège de France, you have more recently turned your attention to another significant hypothesis, that of self-interest. Thus, in the second phase of your critique of homo oeconomicus, you dispute the fact that people act exclusively to maximise their interest however narrowly or broadly understood; at the same time, it is difficult to find examples of disinterested behaviours and to prove that these behaviours are genuinely disinterested. Have you got any in mind?

**Jon Elster :** Let us take a few concrete examples. Starting with the person who pushes a fifty euros note in a deserted church charity box. This, outwardly, is a disinterested act, thrice anonymous since neither friends and associates, nor the charity leaders, nor the beneficiaries are aware of the donation. Therefore it looks disinterested. Now, maybe, God sees it. Therefore it is conceivable that the person is buying their salvation. That amounts to Simony, so there is no point in it, therefore, it is irrational, but people are sometimes, as I have just described, irrational. There is also, but it is a deeper problem, the question raised by Kant who was, I think influenced on this point by La Rochefoucauld: the fact that even when foregoing external plaudits (from friends etc.), it remains possible that the action has the object of obtaining internal plaudit, from within. And that is a possibility that must never be excluded. Thus it would appear, according to Kant, that we should be - perhaps not cynical, busying ourselves with detecting interested motivation always and everywhere, - but merely agnostic: it is impossible to know, whatever the case, if a

given behaviour was indeed motivated by purely disinterested rationale. I will, however set forth an other scenario: Take a Kamikaze fighter whom we suppose to be an atheist, with no inbuilt suicidal tendencies and who carries out his action in complete anonymity in order to advance the cause of his country. To be sure, Kant, or indeed La Rochefoucauld might say that this act of self-sacrifice had no other end than, at the ultimate hour, supreme self-gratification – but it does seem somewhat far-fetched. So I think that there are acts such as kamikaze acts, or equally suicidal missions undertaken by WWII soldiers who gave their life for democracy. I think you've got to be blithely cynical to ignore such phenomena.

**La Vie des idées : You don't think that love of the fatherland is a form of self-interest?**

**Jon Elster :** It is always possible to define interest in an all-embracing way so that the argument that people act out of self-interest becomes circular, but in its standard definition, love of the fatherland, love of the community is not interest, It is, I think, in common parlance and philosophical terms alike, an example of disinterestedness.

**La Vie des idées : Is there a perfect symmetry between the rationality hypothesis and the self-interest hypothesis ?**

**Jon Elster :** they are of course two component of standard rational choice theory, or of the economical choice (*homo oeconomicus*) but I think that the rationality hypothesis is somehow more fundamental in that there is not only the rational choice theory but also the fact that we, each and everyone of us, adopt rationality as a norm. It is a personal norm, we are neither proud of nor happy with our occasional lapses from this norm, so that upholding the personal norm of rationality acts as a sort of permanent counterforce to irrational drifts.

But I can see nothing similar when considering self-interest and disinterestedness. I don't think you can say that disinterestedness represents a permanent counterforce to self-interest. It is perhaps in this sense, therefore that it can be said that rationality is indeed something deeper than self-interest.

**La Vie des idées : What definition of reason or rationality do you propose in replacement of the theories of instrumental rationality (the economists' instrumental theory) that you criticise?**

**Jon Elster :** You say “reason or rationality”. Now, in my inaugural lecture, I explained at length the distinction between reason and rationality. Rationality is the norm of an individual's action's efficiency; reason is a more normative conception of impartial behaviour, impartial in both senses of an impartial treatment of individuals and an impartial treatment of moments in time.

Thus it is impossible to replace the economist's notion of rational choice by the notion of reason since the first is explanatory whereas the second is normative. And neither would I replace the economists' explanatory theory with a new notion of rationality, I would simply state that I would restrict the field of rational explanations to the situations where theory and behaviour concur in that individuals are able to internalise the theoretical norms, that is up to a point beyond which they are not able to make the very fine distinctions and very elaborate calculations advanced by the most refined theory.

Thus a quite minimal theory of rational choice seems essential to me. And my favourite example of the application of a minimalist but very fertile theory of rational choice is Thomas Schelling's idea in his attempt to explain a paradox known since antiquity: what is it that sometimes brings generals to burn their bridges or admirals their ships? It seems absurd. Why not keep all the means at one's command, why give up the use of some of them? Thomas Shelling has shown, by means of a very simple example in game theory that it can be rational to give up some facilities with a view to obtain a strategic advantage.

Or let us take another ubiquitous example in contemporary social sciences: the prisoner's dilemma and the fact that rational individual behaviour may give rise to collective phenomena perhaps not irrational but, say suboptimal. This had long been generally accepted; it is now precisely understood. And so, there is a whole range of similar examples which partly reveal the explanatory strength of rational choice

theory in this more modest version, but perhaps more importantly its conceptual power since it is only today that the nature of an interaction structure is truly understood through the use of game theory. I think game theory was the most significant advance in 20<sup>th</sup> Century social sciences, not so much for its explanatory power than for this conceptual power: it makes it possible to distinguish between interaction structures which, though outwardly similar, are, under closer scrutiny, seen through the lens of game theory, very different indeed.

**La Vie des idées : Over and above your work on rational choice and the philosophy of science you have taken particular interest in the question of justice and equity in three distinct areas: collective bargaining, scarce goods allocation (such as transplant organs) and finally retributive justice (and the complexities it brings up in the case of political transition, such as for instance what happened in former East-European countries, post-communist transitions, or even simply, after the Allies victory in France when the fate of collaborationists had to be decided). Can you explain the distinction between these three forms of justice and in which way it seems useful to you in order to understand the motivations of social actors?**

**Jon Elster** : I will first take you back a little to explain what brought me to take an interest in all these questions... Like everybody in my generation, I have been deeply influenced by John Rawls' Theory of Justice, published in 1971, but I thought, after some hesitations, that I did not really have either the inclination or the talent required to develop normative theories under my own steam. And for reasons I can't quite reconstruct, I turned my attention to the perceptions people held of justice as factors driving their behaviour. What I mean is that to explain people's choices, it often makes sense to take their conception of justice into account since sometimes they choose this action rather than that which at first appears to them more in keeping with justice.

Thus we have both a vast spectrum of conceptions of justice, liable to have a causal and therefore explanatory strength and a broad spectrum of fields of application, such as the three you have just mentioned, and which are incidentally quite different in practical terms.

If you take for instance local justice and the example of transplant organs allocation, I think that there, often, the disinterested concern for justice or equity is quite pure. That is the case, for instance when US medics suggest changing the kidney transplant allocation system in order to advantage the African American population which on the one hand is more prone to kidney failure and on the other is less compatible with the kidneys available. There, I think, we really have quite a pure and disinterested concern for equity since the medics themselves have nothing to gain from it.

Then again, a strategic use of equity norms is also frequently observable in collective bargaining. I am not suggesting that the trade unions that call on this or that equity norm be hypocritical as I consider conscious hypocrisy to be rather rare. But I think that the unions – as do the rest of us I suspect – very often gravitate towards a conception of equity or justice that fits in more or less closely with their interest. Not too well, because it should not be too conspicuous, but all the same, to some extent. I have been able to observe, for instance, in collective bargaining in Sweden that trade unions had up their sleeves tens, perhaps hundreds of reasonable-sounding equity - or justice norms, on which they could accordingly rely when arguing with employers.

And finally, yet another, though different example of justice as a smokescreen for something else: transitional justice. In the countries that had been occupied by Germany post 1945 and also in East-European countries post 1989, the aspiration was for a justice that would be distinct from the desire for revenge. It was very striking in both instances, as it was paramount not to imitate the regimes that had just been toppled and whose disregard for the Law had been a glaring feature. People wanted to observe the legal process and obtain revenge at the same time. Not an easy thing to do. Nevertheless, measures in no small part motivated by passions and emotions, by the desire for revenge were dressed up as acts of justice. Whereas in collective bargaining, as I said earlier, justice was, to some extent a smokescreen for interest, here, justice was a smokescreen for passions. This can be verified in a very simple way: by noting that the sentences meted out for a crime in 1945 were much more severe than those for the very same crime in 1948, which fits in with a significant feature of emotions: they are short-lived, whereas Justice, in principle, is timeless.

**La Vie des Idées : Regarding this question of your theories's practical implications on the issue of transitional justice, you have suggested that in countries where pretty much every citizen had been corrupted by the regime, it would be resolutely better to turn to the future and shun any justice turned towards the past insofar as it would be arbitrary to select a few individuals and impossible to punish everyone adequately and justly.**

**Is it not, in a way, too much to ask of human nature, given the importance you also allow for human passions etc.?**

**Jon Elster** : You are absolutely right: I made this proposal in 1992 that secret police archives should be burnt and even the land registry too in order to stop the property restitution process, and nobody listened to me. Accordingly there were trials, guilty verdicts, and vast transfers and property redistribution.

The reason I wanted the archives and land registries burnt was that a full rectification and retribution was impossible and that incomplete rectification and redistribution were unfair and arbitrary.

For instance, favouring property restitution, a tangible thing, sidelined the huge intangible sacrifices and suffering born by a large part of the populations. And likewise the selection for retribution of those who had collaborated with secret police services released those who had worked full time for these secret services since they did not figure in the archives (they ran them). So that all this seemed so arbitrary to me that I proposed to turn a collective back on the past and look to the future. And like many leaders in the region, I took a leaf from the Spanish experience. When, in the 70s, the Spanish transition took place, Franco's erstwhile partisans, the Socialists and the Communists came to an agreement foregoing transitional justice, no doubt partly because they were dealing with a remote past, and also because there were guilty parties on both sides, so it was in nobody's interest to keep this past alive.

**La Vie des idées : An other example of the practical implications of this theory of justice is that in the event of child custody litigation, you think it would be fairer for parents to draw lots rather than spend years in legal action. Could you develop this?**

**Jon Esler :** In most Western countries, the Law states that child custody must be determined according to the highest interest of the child. So, it is not a question of justice towards the parents but exclusively of the good of the child. Therefore, in principle the best for the child would be to be with the parent best able to look after him or her. However, the identification of the best-suited parent takes a long time and causes much suffering. Countless examples show that the animosities, hostilities, hatreds between parents arising from the court case and the resulting suffering for the child are considerable and therefore, I think that in a great many cases, it would have been much wiser simply to decide from the outset, without any litigation, by tossing a coin.

It is clear that, there again, I have not met with much support since the proposal seems absurd. Nevertheless, I had the following and rather interesting experience. When I was at the University of Chicago, I took part in a radio programme where I explained my proposal and people phoned in to ask questions. A gentleman asked: “Professor Elster, do you really think that children’s fate should be left to the cast of a dice?” And I answered that I realized it would never happen. But the following call was from a gentleman who told me that his wife and he had applied to adopt a child and that their demand had been rejected by drawing lots. And he added: “we felt much happier than if our parenting skills had been assessed and found wanting”. In such a case a lottery’s impartiality leaves people’s self respect unscathed. Whereas knowing that we have been judged and rejected, is so very harsh. At least in this case, not one of custody but of adoption, which is not that very different, a lottery system is used and works.



**La Vie des idées : I wonder if one could see an interesting application of the prospect theory here as it would seem that people have less regret for things that they have not been granted by fate; on the other hand, they are much more reluctant to lose something to fate....**

**Jon Elster** : Undoubtedly. Losing custody of an existing child, compared to not obtaining to care for an adopted child, the former loss will be felt much more grievously. So this may explain why it is easier, seems more acceptable to resort to drawing lots in the latter instance rather than the former. That's conceivable.

**La Vie des idées: you have also written, particularly in French on Leibniz, Marx, Tocqueville. Why these authors and how did they contribute to your intellectual journey?**

**Jon Elster** : Academic and personal both. I was entranced by Leibniz. By chance at the time, in the early 70s, I was interested in the history of biology and I found that it all began with Leibniz and monads (cell equals monad); at this, I gave into Leibniz-mania for about two years, I wrote a book about Leibniz and I never returned to him after that. It was a momentary but fortunate mania!

My book on Marx started with my thesis in Paris, and then I wrote a book on him in English 13 years later; it fulfilled a personal and political project, since in Norway at the beginning of the 60s I was a socialist, my parents were socialists, all my friends were socialists and I wondered whether I could base my instinctive socialism on somewhat more solid theoretical grounds. And so I set to studying Hegel first, since it is impossible to understand Marx without Hegel, and therefore did a mini-thesis on Hegel in French. Then I did my thesis on Marx, in Paris under Raymond Aron's supervision. And after these 10 or 15 years, I realised that no, Marxism did not provide the theoretical basis I was seeking. So my own Marxism was a bit like capitalism in Russia: it wilted before it had blossomed. It was, if you will, like personal failure of some sorts. But it was interesting to work on Marx, it has taught me a lot.

With regards to Tocqueville, that simply had to do with the extreme intellectual excitement I felt when, at the end of the 70s, I reread *Democracy in America* (which I had read in college – without grasping much of it); When I reread *Democracy in America*, in the light of other social sciences, history texts I had read in between, I was struck by the brilliance of Tocqueville’s reasoning – and of his style, of course. So it was genuinely sort of blinding revelation that I have only experienced with Tocqueville, Thomas Schelling and Paul Veyne. Those are the three occasions in my life when I really had that experience of getting so worked up as I read that I have to stand up and walk around my room to calm down. So I have just finished a book in English on Tocqueville which will come out in about a year’s time.

**La Vie des idées : You seem to be particularly in tune with French literature, French authors. Tocqueville but also Proust and a number of 18<sup>th</sup> century moralists. Where does this stem from? Why? Did the themes you study lead you to dwell on this literature?**

**Jon Elster** : It may well have been chance. Rather than La Rochefoucauld and Proust, it could have been Hume and Jane Austen. It’s the years I have lived in France... I don’t know if that’s a very good answer but I guess it is the clarity of expression, the depth of intuition (in Pascal or Stendhal for instance)...

**La Vie des idées : Especially the psychology...**

**Jon Elster** : Especially the psychology. And I have had this slightly odd, indeed surprising experience, in my lectures at the Collège de France: I have found that the hypotheses are to be found in the work of the moralists (the questions, that is) while the answers rest with contemporary economics and psychology. And so there is this kind of to-and-froing between the 17<sup>th</sup> and the 21<sup>st</sup> Centuries, which might be somewhat unsettling for the audience, I don’t know.

**Translated from French by Françoise Pinteaux-Jones**

Published in [laviedesidees.fr](http://laviedesidees.fr), November 19th 2008

© [laviedesidees.fr](http://laviedesidees.fr)