CENTRAL EUROPEAN POLITICAL SCIENCE REVIEW

CONTENTS

John C. Harsányi and the Game Theory –
25 Years of Nobel Price and 100 Years ago was born

INTRODUCTION

The **No. 78 of Central European Political Science Review** is dealing with the game theory in the social science. For the past twenty-five years eleven game-theorists have won the Nobel Prize. Among the first was **John Charles Harsányi with John Nash and John Selten**.

Harsányi was named the "Hungarian Beautiful Mind" or "One of The Martians" (by **Einstein**). Generally accepted that game theory was established by **John Neumann,** but **John Harsányi** topped it off. In the current issue of CEPSR in his excellent chapter **Prof Jon Elster**, as one of the good friends of the brilliant thinker clearly stated: "I have enormous respect for **Harsányi**". For other side we can say that he was an essential creator of humanity also because he made ethics a part of human actions again, and with this he could possibly save rationality itself in the 21th century, giving it a new meaning through simultaneously making humanism a creating force.

John C. Harsányi was born 100 years ago (May 29, 1920 in Budapest) and died 20 years ago (Aug. 9, 2000 in Berkeley), and 25 years ago has got the Nobel-price (Dec. 1994). This is Harsányi's most important message for the 21th century. This is the reason why we dedicate and recommend the **Vol. 20 No. 78 of CEPSR to** the Hungarian-American game theory pioneer to **J. C. Harsányi.** This issue contain six important studies about game theory.

The **Central European Political Science Review** is a refereed academic journal that publishes theoretical, policy, and empirical research articles on Central European countries across the whole spectrum of political science, history, social science, environment protection, and culture, all with emphasis on Central European politics, public policy, political culture and foreign policy in comparative perspectives. However, **CEPSR** also welcomes manuscripts on different aspects of contemporary social development, when these relate closely to Central European politics, political economy, political culture, reform and opening, development, the military, law and legal system, foreign relations, and other important issues of political significance.

The editors of **Central European Political Science Review** welcome articles, conference reports and book reviews for publication, which focus on Central Europe or has a message for this region.

János Simon
Editor-in-chief

Jon Elster

Some Personal Reflections on Harsányi

Over the years I have read many writings by John Harsanyi, and much benefited from what I could understand. I also met him personally several times in the 1970s, and had stimulating talks with him, one of which I shall discuss later. In this short paper, I shall first discuss some game-theoretic issues on which he wrote extensively and about which I have no claim to expertise, and then at greater length consider some socio-psychological issues on which he wrote little but about which I know more.

I believe the first writing by him I came across was his demonstration of the equivalence of the Zeuthen and Nash solutions to two-person bargaining games. From the point of view of the philosophy of the social sciences, this result was remarkable. It seemed – and still seems - miraculous that a dynamic, process-oriented model of bargaining should yield the same prediction as an axiomatic model, especially as the solution – to maximize the product – not the sum - of the utility gains of the two parties seems so counterintuitive. To my knowledge, Harsanyi never commented on the equally remarkable fact that the Rubinstein solution to the two-person bargaining problem also converges to this outcome when the time interval between successive offers and counteroffers goes towards zero.

In his writings on bargaining, as in his work on game theory more generally, Harsanyi was always, and rightly, concerned with the problem of *indeterminacy* in strategic interaction. Indeterminacy threatens to undermine the predictive power of game theory, an issue with which Harsany was much concerned. His book with Reinhart Selten, *A General Theory of Equilibrium Selection in Games*, was a heroic attempt to prove that even in games with multiple equilibria, none of which dominates the others, there always exists a unique solution which will be the point of convergence chosen by rational actors. Harsanyi must have read with mixed feelings the comments by Robert Aumann in the Preface he wrote for the book: "Although the theory selects a unique equilibrium, as a theory it need not be unique." Indeed, their solution concept may be seen as representing *one of several reflective equilibria* in the sense of Rawls (1971).

The issue of indeterminacy goes back to the inception of game theory. In 1713, in the second edition of *Essay d'Analyse sur les jeux*

de hazard, Rémond de Montmort describes a two person zero-sum game between a father and his son that, in normal form, can be stated thus:

Father chooses

		Odd	Even
	Odd	1, -1	0, 0
Son chooses			
	Even	0, 0	4, -4

Commenting on this game, which has no equilibrium in pure strategies, Montmort wrote that, "Such questions are very simple, but I believe they are insolvable. If that is the case, it is a great pity, for this problem often arises in ordinary life, as when each of two people who have some business together wants to adjust his behavior to that of the other" (Montmort 1713, p. 406).

In the first half of the twentieth century, the "insolvable" problem was solved, at least apparently, by Emile Borel and John von Neumann, by the introduction of mixed strategies. However, mixed strategy-equilibria are equilibria only in a weak sense, since each player is indifferent among the all the pure strategies to which the mixed strategy assigns a non-zero probability. Knowing this, the other player has no reason to choose her equilibrium strategy, and the whole equilibrium unravels.

Harsanyi proposed several responses to this conundrum. One was that the lack of a good reason for the agent to conform to the solution in such cases points to a flaw in game theory as traditionally conceived. In his substitute solution concept, only "centroid" or equiprobabilistic mixed strategies are allowed. This proposal reflects the idea that when there are several strategies that would give the agent equally good payoffs, one is chosen at random by "what amounts to an unconscious chance mechanism inside (the agent's) nervous system|" (Harsanyi 1977, p. 114.) Here rational choice is replaced by a purely causal mechanism.

Another and, I believe, more influential response was the reinterpretation of mixed strategies in terms of incomplete information. In the summary on Wikipedia, "we can think of the mixed strategy

equilibrium as the outcome of pure strategies followed by players who have a small amount of private information about their payoffs".

Because I have not wrapped my mind around the latter idea sufficiently to see whether it eliminates the indeterminacy, I shall instead recount a discussion I had with Harsanyi at Stanford in the summer of 1977. At the time, I was working on a paper on "Some unresolved problems in the theory of rational behavior", later published as Ch. 3 of *Ulysses and the Sirens* (Elster 1979/1984). It was natural, then, to ask what him he thought these problems might be. As I did not take notes at the time, the following reconstruction from memory of his answer, using a stylized example, which he did not provide, will have to do:

> There are 100 elected members to the city council, which has a quorum of 40. Assume that all members derive some utility from the quorum being reached, to preserve a viable local democracy, a utility that exceeds the direct and opportunity costs of showing up. The quorum is taken at the beginning of the meeting. Because all want to go back watch an important soccer match on TV, nobody stays after the quorum is taken, hence they have no substantive interest in any decisions that might be reached. There is a large number of equilibria in pure strategies, in each of which exactly 40 members attend. Which equilibrium, if any, will be selected?

As any theory, rational-choice theory may in fact fail in two distinct ways: by *not yielding unique predictions* and by *observed behavior not matching the predictions* when they *are* unique. As I have argued, Harsanyi was deeply concerned with the problem of *indeterminacy* in rational-choice theory. Indeterminacy arises not only in strategic interaction, but also, for instance, in gathering information. As Sidney Winter (1964-65, p. 252) wrote, "The choice of a profit-maximizing information structure itself requires information, and it is not apparent how the aspiring profit maximizer acquires this information or what guarantees that he does not pay an excessive price for it"; see Elster (1989), Ch. 1.3 for other examples.

By contrast, Harsanyi did not pay much attention to the second source of failure, *irrationality*. To my knowledge, the only place where he discusses irrational behavior is in the article "Rational-choice models of political behavior vs. functionalist and conformist theories" (Harsanyi 1969). I shall discuss what I take to be the

central claim of that paper shortly, but at this point I shall only briefly consider his, equally brief, discussion of irrationality.

Given the focus of his article, Harsanyi limits himself to irrational *political* behavior, which he mainly understands as *symbolic* action, undertaken for the psychological satisfaction it provides to the participants rather than for its probable impact on policy. He asserts that "there is *per se* nothing 'irrational' in performing such symbolic actions", except when "people indulge in such symbolic actions at very high costs to themselves, which seem to be quite out of proportion to the psychological satisfactions they can possibly obtain in this way". As an example, "opposition parties may organize a protest demonstration against government policies even though they do not expect it to have any effect on the government's action" and the protesters risk being beaten up by the police. An action that has "very high" costs and no instrumental value must be irrational. The action of the protesters may, he suggests, be the result of rational manipulation by a leader, who hopes to get participants locked into positions from which they will later find it hard to back down, but that fact does not make *their* action rational.

This analysis is distinctly unsatisfactory. It is surprising that Harsanyi, an economist, does not provide a *metric* for assessing when costs are "out of proportion" to psychological satisfactions. More important, he ignores what I believe is the most important aspect of political irrationality (and of irrationality in general), viz. *irrational beliefs* about factual matters or about means-end causality. Some of these beliefs exemplify the "cold" or unmotivated irrationality studied by behavioral economists, whereas others illustrate "hot" or motivated irrationality studied by psychologists. As an example of the former, some objections to free trade may be due to the fact that many people do not understand the principle of comparative advantage (Kemp 2007). As an example (to which I shall return) of the latter, in an emotional state people tend to overestimate or underestimate risk compared to their risk assessments in a non-emotional state (Lerner and Kelner 2001).

I have enormous respect for Harsanyi. I propose this criticism of a very marginal aspect of his work only to show that his penetrating analyses of one possible failure of rational-choice theory – indeterminacy – went hand in hand with virtual neglect (as far as I know) of another, and in my opinion more important, source: irrationality. Because of the pervasive importance of these two failures, I have

come to the conclusion that the core concept of the social sciences should be *choice*, not rational choice (Elster 2015). I shall now consider some other aspects of the article. Harsanyi first proposes a devastating criticism of Talcott Parson's functionalism, a criticism with which I agree wholeheartedly. He also states his adherence to the principle of methodological individualism, which also has my full agreement. He then goes on to discuss the use of rational-choice theory in the explanation of political behavior, and especially the appropriate motivational assumptions for such explanations. One central claim is that in politics, people act on a combination of *interest and low-cost altruism*. People are willing to make small sacrifices of their self-interest for the sake of the common good, but not big ones. In "politics as usual", the claim does not seem implausible, even though it would be hard to verify it with great precision. However, not all politics has this character.

By "altruism", Harsanyi meant something like the public or general interest. I shall use the term *reason*, because I want to rely on a richer set of motivational assumptions proposed by the 17th century French moralists and adopted, incidentally, by the framers of the American constitution. These writers distinguished among *reason, interest,* and *passion*. Reason is the pursuit of the long-term public good. Interest is the pursuit of the good of a proper subgroup of society, be it a group or an individual. Passion, or emotion, is the pursuit of goals motivated by episodic emotions, such as anger, fear, or enthusiasm. These motivations can differ in strength. In a phrase of which Harsanyi would have approved, James Madison wrote that "the mild voice of reason, pleading the cause of an enlarged and permanent interest, is but too often drowned, before public bodies as well as individuals, by the clamors of an impatient avidity for immediate and immoderate gain" (*Federalist* No. 42). I wonder, though, what Harsanyi would have made of a maxim by La Bruyère: "Nothing is easier for passion than to overcome reason; its greatest triumph is to conquer interest." La Bruyère probably took it for granted that interest could easily overcome reason, but added that interest could itself be overcome by passion.

If we add the passions to reason and interest, what is their importance in politics? In politics as usual, evolving in modern democracies around elections and assemblies, they probably have a relatively marginal place, except for some "hot" issues about which citizens feel very deeply. Examples include the *indignation* of

Turkish politicians at the claim that the massacre of Armenians was a genocide, the *resentment* of many American voters of what they perceive as the *contempt* in which they are held by the elites, the visceral *fear* of immigrants in many European countries, and *anger* on both sides of the debates over abortion in many countries.

I want to go beyond politics as usual, however, to discuss the importance of passions in large-scale political change, and specifically in constitution-making. I have been studying this question, including the 1989-90 Hungarian process, about which John Schiemann can tell you much more, for some forty years now. I am currently working on a large-scale comparison between the Federal Convention of 1787 and the first French constituent assembly of 1789-1791.

Constitution-making can involve many emotions, but I shall focus on *enthusiasm*. Although it is not studied by academic psychology, we all know this emotion, at least in its more trivial forms, such as enthusiasm for soccer or stamp-collecting. I shall follow the definition of the Oxford English Dictionary, which yields a less trivial idea: "Rapturous intensity of feeling in favor of a person, principle, cause, etc.; passionate eagerness in any pursuit, proceeding from an intense conviction of the worthiness of the object". In the present context, the object will be a *political principle or cause*, such as national independence or the establishment of democracy. Immanuel Kant asserted that without this emotion, "nothing great in the world has even been done". He added, crucially, that although enthusiasm can induce the adoption of *goods ends*, it tends to undermine the choice of *good means* for these ends.

Let me cite what I wrote some thirty-odd years ago, "the emotions which in some sense provide a meaning and a sense of direction to life, also prevent us from going steadily in that direction" (Elster 1985). I made that statement to characterize Karl Marx, in whom the passionate commitment to the cause of the working class was a source *both of energy and of wishful thinking*. The idea can also be applied to collective efforts, such as constitution-making. In fact, I have argued (Elster 2017) for the following propositions:

> (1) Constitutions tend to be made in times of crisis
> (2) Crises tend to go together with strong passions
> (3) Strong passions tend to undermine rational belief formation
> (4) Only strong passions can generate the political will needed for constitution making

(5) Therefore, constitution making tends to be flawed

(6) Therefore, constitutions tend (weakly) to be flawed.

I shall not present the extensive empirical evidence for proposition

(1), but simply remind you of the many constitutions made in revolutions or counterrevolutions, financial crises, or defeat in war. The only counterexample I have found in Western democracies was the making of the new Swedish constitution in 1969.

Proposition (2) is hardly controversial, and well documented by the same cases. The emotions in questions include enthusiasm - often combined with anger - and fear.

Proposition (3) is supported by a large body of psychological literature, which has demonstrated the existence of several mechanisms by which emotions can induce irrational beliefs. The two most important are *biased belief formation* and *information-poor belief formation*.

Whereas anger can induce irrationally optimistic risk assessments, fear can lead to irrationally pessimistic ones (Lerner and Keltner 2001). Historical evidence suggests strongly that enthusiasm is similar to anger in this respect. I should, though, that anger and enthusiasm can also *induce risk-taking by reducing risk aversion*, not only by causing excessive optimism. When risk-taking behavior is caused by irrational optimism, it is a form of *foolhardiness*, but when caused by lessened risk-aversion, we should probably speak of *courage*. In practice, it may be hard to tell which of the two mechanisms are at work. In his book on the Hungarian transition, John Schiemann (2005, pp. 41-49) explains the cautious attitude of the older generation by risk-aversion, but perhaps it could also be explained by a pessimistic cognitive bias.

Emotions tend to induce *urgency*, a desire to act sooner rather than later. If there is an external danger, this tendency may be rational, or at least adaptive, but it can be triggered even when there is no hurry, as illustrated by the proverb "Marry in haste, repent at leisure". Urgency or, as one might also call it, *inaction-aversion*, causes agents to act before gathering the optimal amount of information. Their decisions will not be biased, but poorly informed. In a constitution making context, urgency can make an assembly ignore its own rules about proceeding slowly, for instance by violating a rule that a motion cannot be adopted on the day it is proposed.

Proposition (4) is a restatement of Kant's dictum. For people to sacrifice their petty, partisan, and short-term interests for the sake of

the common good, they have to be swayed by a strong emotion. As one of the French framers, Clermont-Tonnerre, said in 1789, "Anarchy is a frightening yet necessary passage, and the only moment one can establish a new order of things. It is not in calm times that one can take uniform measures."

Proposition (4) states a necessary, not a sufficient condition. Strong passions can lead to disastrous outcomes. I assume that many of the (courageous or foolhardy) participants in the Tahrir Square demonstrations in 2011 were moved by enthusiasm, but the outcome of the process they set in motion was not democracy, but an autocracy even more repressive than the one they protested against. On May 16 1791, the French framers, "drunk with disinterestedness" (Lebèque 1911, p, 261), declares themselves ineligible to the first ordinary legislature, so that no one should suspect them of making self-serving decisions, The outcome – an assembly made up of inexperienced deputies - opened the door to the rule of the Jacobin clubs and to the Terror.

Proposition (5) follows directly from Propositions (3) and (4). This is not the occasion to discuss Proposition (6), except to mention that a constitution that is the result of a flawed process may be flawed compared to some abstract standard, yet superior to the constitution it replaces. The American constitution is an example.

Let me return to Harsanyi's article. The "conformist" theories mentioned in the title correspond, but only approximately, to what are currently referred as theories of *social norms.* Harsanyi frames the issue as one of resistance to changes in social values, arguing that "the main opposition will come from individuals whose personal interest would suffer from a shift to new values and new patterns of behavior". However, the conformist model seems to collapse into the rational-choice model, because, as he writes, "possible sanction against nonconformity is simply one of the many elements a rational individual will have to include in his cost-and-benefit calculus when he is choosing between acceptance of some new values and loyalty to the old".

This argument raises the possibility of an infinite regress, however, once we ask about the motivation of the sanctioning individuals. To sanction or punish another person for what she says or does often involves a cost or a risk, since she might retaliate in one form of another. To abstain from sanctioning might seem costless, unless there are third parties who sanction non-sanctioners. *But why they*

would they do that? Empirically, the chain of sanctions of non-sanctioners quickly runs out of steam. People do not usually frown upon others when they fail to sanction people who fail to sanction people who fail to sanction norms violators. Instead, we must conclude that people sanction norm-violations because they think *they are wrong*. If a person expresses disapproval of someone urinating in the street, it is not because he thinks that others will express their disappproval of *him* if he doesn't.

In my opinion, the two main pillars of the social sciences are history and psychology. I believe some of the weaknesses in Harsanyi's view of human motivations can be traced back to his neglect of these disciplines. At the same time, this selective attention allowed him to focus in depth on the subset of motivations represented by rational choice, notably in strategic interactions. This trade-off between breadth and depth is probably inevitable.

Literature

Elster, Jon 1984. *Ulysses and the Sirens*, Cambridge University Press

Elster, J. 1985. "Sadder but wiser? Rationality and the emotions", *Social Science Information* 24, 375-406.

Elster, J. 2015. *Explaining Social Behavior*, rev. ed., Cambridge University Press

Elster, J. 2017. "Emotions in constitution-making", *Scandinavian Political Studies* 40, 133-56.

Harsanyi, J. 1969. "Rational-choice models of political behavior vs. functionalist and conformist theories", *World Politics* 21, 513-38.

Harsanyi, J. 1977. *Rational behavior and bargaining equilibrium in games and social situations*, Cambridge University Press

Kemp, S. 2007. "Psychology and opposition to free trade", *World Trade* 6, 25-44.

Lebègue, H. 1911. *Thouret*, Paris: Alcan

Lerner, J. and Keltner, D. 2001. "Anger, fear, and risk", Journal of Personality and Social Psychology 81, 146-59.

Montmort, R. de 1713. *Essay d'analyse sur les jeux de hasard*, 2nd. ed., Montpellier

Rawls, J. 1971. *A Theory of Justice*, Cambridge MA: Harvard University Press

Schiemann, J. 2005. *The Politics of Pact-Making*, London: Macmillan

Winter, S. 1954-1955. "Economic 'natural selection' and the theory of the firm", *Yale Economic Papers* 4, 225-72.