Game theory and institutions

Ken Binmore *

Economics Department, Bristol University, Bristol, BS8 1TN, UK

A R T I C L E   I N F O

A B S T R A C T

Binmore, Ken—Game theory and institutions

This short paper is a much abbreviated summary of an attempt to treat justice as a kind of institution, offered in the hope that it will serve as a case study in how game theory can serve as a useful intellectual framework for the study of human institutions. *Journal of Comparative Economics* **38** (3) (2010) 245–252. Economics Department, Bristol University, Bristol, BS8 1TN, UK.

© 2010 Published by Elsevier Inc. on behalf of Association for Comparative Economic Studies.

1. Rules of the game?

How best can game theory be employed to advance our understanding of economic institutions? North (2005) has been very influential in arguing that an institution should be characterized in terms of the rules that govern its operation. These rules are to be understood as including, not only formal legal rules, but the informal social norms that govern individual behavior and structure social interaction within the institution.

For such "rules of the game" to correspond to rules of the game as understood in game theory, they need to be unbreakable. That is to say, the players must not have strategies available to them that can result in a rule being violated. Can a social norm have such a character? The answer depends on the circumstances.

For example, it may be that the punishment that would be inflicted by one's fellows for falling foul of a social norm may be so severe that nobody would ever contemplate taking the risk while in their right mind. In modeling the institution, one might then sensibly specify obedience to the social norm as being one of the rules of the game. By doing so, one loses the opportunity to study the social norm itself, since its characteristics are imposed on the model by fiat, but studying the reasons for the evolution and survival of this particular social norm may not be the purpose of the analysis.

For example, when considering how best to regulate an industry, one may write a model in which the agents are assumed to honor the requirements of the criminal law. Of course, we can all cheat and steal if we want to, but if the monitoring and enforcement mechanisms put in place by society are sufficiently effective, then nobody will actually cheat or steal, and so the result is the same as if the agents were literally incapable of cheating and stealing. In brief, when adequate enforcement mechanisms are in place, then it makes good sense to interpret North's advice as implying that social norms (like observing the law or maintaining political correctness) should be treated as though they were as strictly binding on the agents as the rules of a formal game.

* Fax: +44 1600 860691.
   E-mail address: k.binmore@ucl.ac.uk

However, it is commonplace for models to be constructed in which the need to locate an actual enforcement mechanism in real life to justify assuming unbreakable rules in a model is not even considered. The most flagrant examples arise in constitutional studies when appeals are made to the theory of mechanism design in proposing ideal constitutions.[1] But there is no external enforcement agency by definition when we are talking about the whole constitution of an independent nation. The rules written into the constitution must therefore be self-enforcing if they are to operate successfully.

One only has to compare the constitution of the USA with that of the old USSR to see why. The latter document is a wondrous list of utopian aspirations, but entirely useless to the citizens of the USSR because those holding power held its provisions in contempt. The constitution of the USA, on the other hand, was written by people who understood that the best a constitution can do is to help in coordinating behavior on one of the many equilibria in the actual game of life that would be played by American citizens in the future. It is admittedly only a pretence that the Supreme Court merely reinterprets constitutional provisions rather than rewriting them to bring them up to date with current opinion, but the survival of the essence of their creation is nevertheless a tribute to the hard-headed realism of the founding fathers of the American Republic.

In brief, a constitution is not a substitute for the unbridled exercise of power. One cannot make power go away by writing words on a piece of paper. The best a constitution can do is to focus attention on a particular way of balancing power. To overlook this point is to fail to understand constitutional issues completely.

The game theory solution to the problem that arises when the rules of an institution are not enforced by some incorruptible external agency is to move to a larger game—the game of life—in which the institution is regarded as being embedded. This larger game must have rules that are genuinely unbreakable, like the laws of physics, because if the players had strategies whose implementation resulted in the rules being broken, then they would not be the genuine rules of the game. The rules of the institution to be studied then have a lesser status. The players can break them if they want to, but if the institution is stable, the rules will not be broken, because obeying them is part of the behavior required by an equilibrium of the game. That is to say, an institution is not treated as a game itself, but as part of the description of an equilibrium within a larger game of life.

2. Equilibrium selection problem

The rest of this paper outlines my theory of justice with a view to offering a case study of how an institution can be seen as an equilibrium in an unvarying game of life (Binmore, 2007, 2005, 1998). I am interested in the institutions that collectively determine the social contracts of societies—especially the ancestral hunter–gatherer societies within which our sense of fair play presumably first evolved. I think that we evolved the capacity to entertain fairness norms because they allowed our species a quick and efficient way to solve the coordination problems that inevitably arise when a group is faced with a new situation. For example, how should a novel source of food be shared without fighting or other wasteful conflict? If I am right, then fairness can be seen as evolution's solution to the equilibrium selection problem that arises in certain games with multiple equilibria.

The idea that fairness arose as a solution to an equilibrium selection problem sometimes creates confusion because critics do not know whether to classify the theory as neoclassical or behavioral. In fact, it belongs to neither category. It differs from the tradition in neoclassical economics in recognizing that almost any realistic game has many Nash equilibria. Even perfectly competitive markets have multiple Nash equilibria if they are repeated every day. It then ceases to be true that there is a necessary trade-off between equity and efficiency. Neoclassical arguments to this effect rely on models that have only one equilibrium. More generally, I think that it is only because traditional neoclassical economics largely confined its attention to models with a unique equilibrium that it managed to get so far without paying any serious attention either to fairness norms or to all the many other institutions without which our society would fall apart.

My theory differs from the standard behavioral approach in not treating fairness as a property of utility functions. People doubtless do have other-regarding preferences to a greater or lesser degree. How else are we to explain why most of us give money to charity? However, I think that a theory of fairness which says that people play fair because they like playing fair is painfully naive. Nor am I satisfied that the empirical support claimed for this approach from experimental work is anywhere near adequate (Binmore and Shaked, 2010). In my theory, individuals may or may not have other-regarding preferences, but even if they were all entirely selfish, they would still be stupid to ignore the fairness norms that operate in their society.[2]

2.1. Social contracts

A social contract is the set of common understandings that allow the citizens of a society to coordinate their efforts.

The common understandings or conventions that make up a social contract are many and various. They range from the arcane table manners we employ at formal dinner parties to the significance we attach to the green pieces of paper we carry round in our wallets bearing pictures of past presidents. From the rules that govern how we drive our cars in heavy traffic to the meaning of the words in the language we speak. From dietary and sexual taboos to the standards of integrity and

---

[1] The word *mechanism* in this context refers to those rules of a game that do not relate to the personal preferences and beliefs of individual agents.

[2] The behavioral literature often blurs the distinction between a social preference and a social norm, but it is important in my work to make a sharp distinction between a payoff function and an equilibrium selection device.