

Publisher: Asian Economic and Social Society

ISSN: 2225-4226

Volume 3 No. 3 March 2013.

Journal of Asian Business Strategy



The Game Theory, Morality and the ‘Game of Life’

Amaresh Das† and **Biruk Alemayehu** (College of Business,
Southern University at New Orleans, New Orleans, USA)

Citation: Amaresh Das and Biruk Alemayehu (2013). The Game Theory, Morality and the (Game of Life). Journal of Asian Business Strategy, Vol. 3, No. 3, pp. 39-47.



Author(s)

Amaresh Das and Biruk Alemayehu

College of Business, Southern University at New Orleans, New Orleans, USA

The Game Theory, Morality and the ‘Game of Life’

Abstract

Game theory as long as it is a study of *interdependent rational* choice, can be used to explain, to predict and evaluate human behavior in contexts where the outcome of action depends on what other agents choose to do. Game theory can be made relevant to ethics and can be used in moral and political philosophy, including economics, business and public policy areas.

Keywords: Interpersonal, comparison of utility, Pareto inefficiency, maxi-min strategy

Introduction

One of the most common objections to an economist’s policy suggestions, whether in the classroom or in Washington, is “That would not be fair!” The economist’s usual response is that fairness is very important, but it is the business of philosophers, not economists as if philosophy had nothing to do with it. But, economists do have something to say to philosophers about fairness and that game theory, in particular, is important if we are to understand what lies at the root of being human. This is not as far-fetched as it may sound.

Game theory is often used to identify the ‘functions’ of morality. Can one, in fact, take this ‘functionalist’ approach? Yes, morality is supposed to correct problems of threatening what economists call ‘Pareto- inefficiency’ that would be the result of unfettered individual rational action. On the functionalist account the moral agent then seems *ipso facto* to be irrational. This then begs the question ‘why be moral? It can be used to describe the problem(s) that would occur in the absence of morality and inferences about the remedial or the ameliorative function of morality can, therefore, be drawn from this description.. Functionalism precludes the answer. Functionalism appears to seek explanations of the emergence and persistence of moral norms and practice. Moral theorists are not really interested in such explanations. Rather,

they usually seek to understand morality with the aim of ascertaining what we should do or what we are obligated to do. It is morality as a guide to action and to life and that is the principal interest of the moral philosopher. Morality here is normative - a source of guidance. There is a difference between determining the function of morality and determining whether a particular set of norms and practices are, in fact, the ones we should follow.

Escape from solipsism is a necessary first step for ethics and game theory asks how people make decisions in light of their opinions about how other people will behave. This goes directly to the question of what we humans have in common with each other. From such a beginning we can move by small steps to the question. First, say, who will buy the next round of drinks and eventually, whether the rich should pay taxes to subsidize the poor? Economists can learn a good deal of philosophy from it and in the same way philosophers can also learn from the question a good deal of game theory. Some of the prevailing philosophic concepts that have to be discarded in the context are Kant’s categorical imperatives, Bentham’s interpersonal comparison of utility, and Rawls’ principle of maintaining the welfare of the least well off.

Moral philosophy: how can Kant, Bentham and Rawls help?

In *Theory and Practice* Kant sets out his social contract justification of a civil state. The categorical imperative is the central philosophical concept in the moral philosophy of Immanuel Kant. It may be defined as a way of evaluating motivations for action. Kant's categorical imperative denotes an absolute, unconditional requirement that asserts its authority in all circumstances, both required and justified as an end in itself. It is best known in its first formulation:

“Act only according to that maxim whereby you can, at the same time, will that it should become a universal law”.

Kant expressed extreme dissatisfaction with the popular moral philosophy of his day, believing that it could never surpass the level of hypothetical imperativesⁱⁱ: a utilitarian says that murder is wrong because it does not maximize good for those involved, but this is irrelevant to people who are concerned only with maximizing the positive outcome for themselves. Consequently, Kant argued that hypothetical moral systems cannot persuade moral action or be regarded as bases for moral judgments against others, because the imperatives on which they are based rely too heavily on subjective considerations. He presented a deontological system, based on the demands of the categorical imperative as an alternative. According to Kant,

An original contract by means of which a civil and thus completely lawful constitution and Commonwealth can alone be established.... [does not have to be assumed to] actually exist as a fact.....Such an assumption would mean that we would first have to prove from history that some nation, whose rights and obligations ever been passed down to us, did in fact perform such a [contract] and handed down some authentic record in legal instrument, orally or in writing before we could regard ourselves as bound by pre-existing civil constitution.

One of the first major challenges to Kant's reasoning came from the French philosopher Benjamin Constant who asserted that since truth telling must be universal according to

Kant's theories, one must (if asked) tell a known murderer the location of his preyⁱⁱⁱ.

Apologists of the Bentham's interpersonal comparison of utility have argued that comparability of mental states such as utility is never possible but believes, however, that human beings are able to make some interpersonal comparisons of utility because they share some common backgrounds, cultural experiences, etc. In the example cited by Sen (1976), it should be possible to say that Emperor Nero's gain from burning Rome was outweighed by the loss incurred by the rest of the Romans. Sen, Harsanyi and others thus argue that at least partial comparability of utility is possible, and social-choice theory proceeds under that assumption. Sen proposes, however, that comparability need not be partial. Under Sen's theory of informational broadening, even complete interpersonal comparison of utility would lead to socially suboptimal choices because mental states are malleable. A starving peasant may have a particularly sunny disposition and thereby derive high utility from a small income. This fact should not nullify, however, his claim to compensation or equality in the realm of social choice.

Social decisions should accordingly be based on malleable factors. Sen proposes interpersonal comparisons based on a wide range of data. His theory is concerned with access to advantage, viewed as an individual's access to goods that satisfy basic needs (e.g. food), freedoms (in the labor market, for instance), and capabilities. We can proceed to make social choices based on real variables, and thereby address actual position, and access to advantage. Most importantly, Sen's method of informational broadening allows social choice theory to escape the objections of those like Robbins, which looked as though they would permanently harm social choice theory.^{iv}

According to Harsanyi (1977), social preferences are established on conditions that personal preferences are corrected and censored by an Impartial Observer. Goodwin (1986) points out this issue and proposes an alternative, thanks to 'laundering preferences'; it consists in encouraging individuals to

modify their preferences to 'launder' them. Goodwin emphasizes that in a social context individuals are ready to correct their preferences by themselves. They will express only their public-oriented, ethical preferences while suppressing their private-oriented ones. It brings us back to Sen's (1976) resolution in which he maintains there is a difference between individual preferring a social state to another and an individual wanting his/her preferences be taken into account in social choice. Sen for the first time introduced the concept into the social preference choice theoretical framework with a condition of liberalism and the notion of decisiveness: individuals must be decisive – their preferences must be acknowledged by society – over some pairs of social states, which belong to their private sphere. Sen shows that the condition of liberalism and a weak Pareto principle lead to an impossibility of a Paretian liberal. But, Sen's formal analysis has no need to distinguish between decisive pairs that enable an individual to take decisions that are 'personal' to him/her and those that are not.^v The Kantian ideal of a hypothetical contract as the moral foundation for a liberal conception of justice has been further developed by John Rawls in his *A Theory of Justice*. Rawls like Kant argues that

Principles of justice... are the principles that free and rational persons who are concerned to advance their own interests would accept in an initial position of equality.

According to Rawls, the principles of justice that would be derived in the original position are the following

A special conception of justice with a principle of equal political liberty, a principle of equal opportunity and a principle requiring that the distribution of economic goods works to the greatest advantage of the least advantaged.

A general conception of justice with a principle requiring that the distribution of all social goods works to the general advantage of the least advantaged.

Rawls' principle of maximizing the welfare of the least well-off rests heavily on the claim that morally arbitrary factors (for example, the family one is born into) shouldn't determine one's life chances or opportunities. Rawls' claim that departures from equality of what he calls primary goods — "things which a rational man wants whatever else he wants" are justified only to the extent that they improve the lot of those who are worst-off under that distribution in comparison with the previous, equal, distribution. Rawls is also keying on an intuition that a person does not morally deserve their in born talents; thus that one is not entitled to all the benefits they could possibly receive from them; hence, at least one of the criteria which could provide an alternative to equality in assessing the justice of distributions is eliminated. His position is at least in some sense egalitarian with a proviso that equality is not to be achieved by worsening the position of the least advantaged. An important consequence here, however, are those inequalities can actually be just on Rawls' view, as long as they are to the benefit of the least well off?

Rawls holds that his principles of justice should be chosen in the original position because persons so situated would find it reasonable to try to secure for themselves the highest minimum payoff. In effect they would want to follow the conservative dictates of the 'maximin strategy' and maximize the minimum payoff. Rawls describes his reasoning as follows:

Now looking at the situation from the standpoint of one person selected arbitrarily, there is no way for him to win special advantages for himself. Nor, on the other hand, are there grounds for his acquiescing in special disadvantages.

Since it is not reasonable for him to expect more than an equal share in the division of social goods and since it is not rational for him to agree to less, the sensible thing for him to do is to acknowledge as the first principle of justice one requiring equal distribution.

Some egalitarian critics have raised concerns over Rawls' emphasis on primary social

goods. For instance, Sen has argued that we should attend not only to the distribution of primary goods, but also how effectively people are able to use those goods to pursue their ends. In a related vein one wonders why health care shouldn't be treated as a primary good, and some of Rawls subsequent work has addressed this question, arguing for a right to health care within a broadly Rawlsian framework.

Taking lessons from moral theory

We can use tools from game theory to build them back up. The categorical imperative cannot be justified by the argument that 'if everybody behaved Kantian way, where would we be?' But reciprocity in repeated games may rescue something very like it. Interpersonal comparison of utility is disdained in modern welfare economics but evolutionary biology shoe why people should be able to know each others' levels of happiness. Rawls principle is hard to justify as the outcome of rational decisions in his original position but perhaps it can be revived using the idea of multiple equilibria in games. How is this all to be done?

We can begin with 'the game of life' in which we live our lives maximizing utility by our choice of behavior towards other people. Societies are organized around conventions and we can think of these as equilibria of the game of life. Not love and duty but reciprocity is the cement of society. A number of different conventions are possible equilibria and some equilibria are preferable to others.

Once a convention is established, we all obey it from self-interest as Nash equilibrium. Roughly we can distinguish two strategies. First, there are those who will model morality as the result of one-time choice of a very large collection of agents, the moral community. Secondly, there may be those who will approach morality as the result of a series of repeated small-scale interactions. Here morality is interpreted as the outcome of bargaining process. This is an old idea in moral and political philosophy: it is the idea of *social contract*.

Within the bargaining theory, there may be two approaches that seek to answer the question posed here. First, there is the traditional axiomatic approach as developed in the context of cooperative games. The axiomatic approach assumes that once rational agents have come to an agreement, they will comply with it.

The task of the theorist is to consider the bargaining area and determine which outcome would satisfy a number of reasonable requirements of a rational outcome of the negotiations. Things such as the names of the parties concerned should not matter for the result, whereas their preferences do matters.^{vi}

An axiomatic theory of the games that endows players with degrees of belief would provide axioms, specifying players' probabilistic belief and a formal model of such theory is pretty complex.

Recently some have presented many simpler and manageable axiomatic theories that employs (weak or strong) knowledge operator. In these cases, in order to maximize one's expected payoff at a node, a player has to know what to expect at following nodes, which is the same as saying that the player has to know that the following players act rationally.

From the point of view computing equilibrium all the theories are equivalent. They differ however, in the way that handles deviation. Criteria of choice among them might thus be the extent of the revisions that a deviation induces^{vii}. This intuition is the driving force of the so-called Nash (1950) program.

This program aims at evaluating axiomatic solutions by checking whether the outcome of a negotiation game leads to the same outcome. The success of the Nash program is crucial for the plausibility of the classic axiomatic theories of the Social contract^{viii, ix}. In deciding which equilibria are fair, we should look to an imaginary 'game of morals' identical to the game of life except that at any point a player can call for a return behind the veil of ignorance to reshuffle everyone's position in

society – knowing, however, that someone else might then call for another reshuffling.

A fair outcome is a fixed point in the game of morals, a set of conventions from which no one would appeal. How can we be sure that there is always one unique solution or are bargaining problems to some extent underdetermined? The plurality of bargaining solution concepts that are discussed in bargaining theory is a bad omen in this regard. There are reasons to doubt that the game-theoretic approach to bargaining can really help us predict the outcome of the negotiations of rational agents. Both the axiomatic approach and the non-cooperative game approach proceed from the assumption that there is a unique, rational outcome of such negotiations. While this may be plausible in some situation, it is far from obvious that this is always the case.

As a matter of fact in recent days, many, notably, Bicchieri (1990), Basu (1990), proceed to distinguish between the game theorist's and the player's own 'theory of the game'. The latter is theory that is sufficient for each player to infer a certain sequence of moves, whereas the former is intended as a justification of such a sequence of moves. They then inquire what happens when a theory of the game is augmented with information that a move outside the inferred solution has occurred. It is shown that a theory that is sufficient for the players to infer a solution and still remains consistent in the face of deviation must be 'modular'. By this they mean that players have distributed knowledge of it. Furthermore, whenever the theory of the game is 'group knowledge' (common knowledge) among the players (i.e., it is the same at each node); a deviation from the solution gives rise inconsistencies and, therefore, forces a revision the theory at later node. On the contrary, whenever a theory of the game is modular, a deviation from equilibrium play does not induce a revision of the theory^x

Insights from game theory

Game theory that regards morality as the intended result of a complex large scale bargaining process between fully informed

and fully rational agents, one can move away from all the established assumptions.^{xi}

By the same token, an evolutionary game theorist writing about ethics and morality have shown that among not-so-fully rational agents many of the norms of coordination and cooperation can emerge that are usually the object of inquiry of the more traditional moral theories, that is, evolutionary theorists can show that evolution favors not only the emergence of patterns of behavior that conform to moral standards but also favor the development of cognitive heuristics that have all the characteristics of moral reasoning.^{xii}

Regardless of the merits of the different approaches we have discussed here, there are some remarkable insights that the application of game theory offers to the moral theorist. There may be many games with multiple equilibria. This is especially the case with iterated plays of particular games such as the prisoner's dilemma. One of the implications of this fact is that insofar as these games are helpful representations or models of our social interactions, we have reason to expect such indeterminacy in the game of life. Also, game theory makes it clear that in any sufficiently large population we can expect determinate mixes of behavioral dispositions.

The big question then remains to be spelled out big time is: whether everything that is relevant for moral theory about the agent can be captured by the rather one-dimensional picture of rational man as proposed by different versions of prevailing game theory. The agent is supposed to be completely characterized by his preference rankings over outcome and his beliefs at each stage of the game. However, morally important distinctions – e.g., between important differences in character - have no place in this characterization. What about many repeated prisoner dilemma games where it of course pays to have a 'reputation' to be cooperative? Ordinarily, reputation is generally believed and associated with a person's character. But in a game theory, a reputation can simply a history of the game that is, the outcome of player's *moves* in similar games. There is a relevant difference between the two cases.

One important feature that is missing here is, call it, moral education (for the old-fashioned) or design (for the modern). In *The Republic*, Plato moves from trying to explain why Glaucon should behave virtuously to how society could be redesigned to make future Glaucon behave virtuously. Economists do not want to make the transition from analyzing equilibria of the existing game to changing the rules of the game. One of the game theory's most profound lessons is that a player can benefit from new rules which reduce his pay-offs on out-of-equilibrium paths – the pangs of conscience, for example. Game theorists acknowledge that people in the original positions might choose a society in which education alters their preferences but somehow ignores it. Incorporating knowledge or education into the framework of analysis is crucial unless economic education improves enough that can produce an entire society of citizens who understand the importance of playing fair instead of having to take it on faith.

Conclusion

Economists who use game theory in their work may be functionalists, contractarians, or even evolutionary theorists. All these approaches represent different combination of game theory and ethics. The contractarians with their emphasis on fully rational agents and bargaining represent a more traditional use of game theory.

The new evolutionary approach, on the other hand, with its emphasis on bounded rational agents and repeated interactions, is a new arrival with a certain promise. What is needed however is a kind of synthesis of these approaches. Game theory in economics, as well as in other public policy areas tends to suggest that people are honest only to the extent that they have economic incentives for doing that.

This is purely a *homo economics* assumption which is not all true and which must have to confront reality. Again, no society is viable without some norms or rules of conduct or a knowledge operator where strictly economic incentives may not be present or when a man

may not be able to shine as a 'rational fool'. Preferences as rankings have to be supplanted by some richer concepts involving culture, education or commitment that may be at variance with the structure of all traditional game theories. The traditional game theory with its emphasis on *fully* rational agents and bargaining can be said to have been overburdened with extra structure.

References

- Basu, K. (1990). On the Non-existence of a Rationality Definition for Extensive Games. *International Journal of Game Theory*, 5(2): 33-44.
- Bicchieri, C. (1990). *Rationality and Coordination*, Cambridge University Press
- Binmore, K. (1994). *Playing Fair Game Theory and the Social Contract*, 1, Cambridge, MIT Press.
- Binmore, K. (2005). *Natural Justice*: New York, Oxford University Press
- Rubinstein, A. (1982). Perfect Equilibrium in a Bargaining Model. *Econometrica*, 50(1): 97-109.
- Gauthier, D. (1967). Morality and Advantage, *Philosophical Review*, 76(4): 460-475.
- Goodwin, R. E. (1986). Laundering Preferences' in Elster and Hylland (eds), *Foundations of Social Choice Theory*, 75-101, Cambridge University Press, Cambridge.
- Gibbard, A. (1974). A Pareto-consistent Libertarian Claim. *Journal of Economic Theory*, 7(1): 388-410.
- Harsanyi, J. C. (1977). Behavior and Bargaining Equilibrium in Games and Social Situations, Cambridge University Press, Cambridge.
- Harsanyi, J. C. (1955). Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility. *Journal of Political Economy*, 63(2): 309-321.
- Hirofumi, U. (1999). Global Warming as a cooperative game. *Environmental Economics and Policy Studies*, 2(1): 80-95.
- Rawls, J. (1971). *A Theory of Justice*, Cambridge, Harvard University Press.

- Skyrms, B. (1996). *Evolution of the Social Contract*, Cambridge, Cambridge University Press.
- Skyrms, B. (2004). *The Stag Hunt and the Evolution of Social Structure*, Cambridge, Cambridge University Press
- Sen, A. (1976). Liberty, Unanimity and Rights. *Economica*, (4), 217-245.
- Sen, A. (1977). Rational Fools, A Critique of the behavioral Foundations of Economic Theory. *Philosophy and Public Affairs*, 6(4): 317-344.
- Verbeek, B. (2007). The Authority of Norms. *American Philosophical Quarterly*, 44(3): 245-258.

Notes

ⁱ Even if moral norms and practices serve to bring about Pareto-superior outcomes not realizable through uncoordinated individual rational action, no explanation of the existence and persistence of morality are provided unless it is shown that this function somehow motivates human action or in some other way is causally effective in bringing about mutually beneficial outcomes.

ⁱⁱ Hypothetical imperatives apply to someone dependent on them having certain ends:

- if I wish to quench my thirst, I must drink something;
- If I wish to acquire knowledge, I must learn.

ⁱⁱⁱ Kant denied that such an inference indicates any weakness in his premises: not lying to the murderer is required because moral actions do not derive their worth from the expected consequences. He claimed that because lying to the murderer would treat him as a mere means to another end, the lie denies the rationality of another person, and therefore denies the possibility of there being free rational action at all. This lie results in a contradiction in conceivability and therefore the lie is in conflict with duty.

^{iv} Additionally, since the seminal results of Arrow's impossibility theorem many positive results focusing on the restriction of the domain of preferences of individuals have elucidated such topics as optimal voting. The initial results emphasized the impossibility of satisfactorily providing a social choice function free of dictatorship and inefficiency in the most general settings. Later results have found natural restrictions that can accommodate many desirable properties.

^v Formally Sen makes no distinction between a man deciding whether to sleep in a prone or supine position and a religious leader dictating whether one should do so. Gibbard (1974) investigates the issue: he uses a Cartesian product structure to describe individual rights and points out the internal inconsistency caused by an extended condition liberalism. This result is called Gibbard's paradox (or Gibbard's First Liberalist Claim).

^{vi} Harsanyi (1955), Rawls (1971) and Gauthier ((1967) (2008)) all have used axiomatic approaches to justify this version. Gauthier not only uses bargaining theory to determine, as Rawls and Harsanyi sought to do, the content of fundamental moral principles; he also tries to show that rational agents will act morally.

^{vii} Any theory of the game that employs a monotonic logic is sufficient for the players to infer a solution becomes inconsistent when augmented with information that an off-equilibrium move has been played. In order to preserve consistency a revision of the theory is in order. When revising a theory of the game, it matters how much the players know i.e. it matters whether the theory is a common knowledge, group knowledge, or distributed knowledge among players.

^{viii} Such theories regard morality as the result of (hypothetical) negotiations between ideally rational agents but do not bother to spell out exactly how the parties reach the result. See Binmore (1994), Rubinstein (1982). Consequently, if there is not at least the promise of such a detailed analysis, as is promised by the Nash program, the result they present lacks plausibility.

^{ix} There are, of course, other issues, psychological and sociological. These complex issues underlying choice have recently been forcefully brought out by a number of penetrating studies dealing with consumer decisions and production decisions. It is very much an open question as to whether these behavioral characteristics can be at all captured within the formal limits of consistent choice on which the maximization approach depends. In this context I cannot resist the temptation of quoting a passage from a work by Sen (1977):

A person's choices are considered 'rational' in this approach (choice consistency approach) if and only if these choices call all be explained in terms of some preference relation consistent with the revealed preference definition, that is, if all his choices are explained as the choosing of 'most preferred' alternatives with respect to a postulated preferences relation. The rationale of this approach seems to be based on the idea that the only way of understanding a person's real preference is to examine his actual choices, and there is no choice –independent way of understanding someone's attitudes towards alternatives.

^x An axiomatic theory of the game that endows players with degrees of belief would thus have to provide axioms specifying players' probabilistic beliefs, and a formal model for such a theory would be much more complex and cumbersome than the kind of model we provide. A much simpler and manageable axiomatic theory of the game would be a theory that employs a (weak or strong) knowledge operator. In this case, in order to maximize one's expected payoff at a node a player has to know what to expect at following nodes, which is tantamount to saying that the player has to know the following players act rationally.

^{xi} We can illustrate, à la, Skyrms (2004) this as follows: Rousseau describes the state of nature as one that resembles the so-called Stag Hunt: Imagine two hunters who can choose to hunt for hare. Their chance of calculating a hare are not affected by the actions of others. However both prefer to have venison for dinner but if they were to hunt for stag, they will only be successful if the other does so as well.

	#2	
	stag	hare
	3	2
stag	3	0
#1		

hare	0	2
	2	2

Suppose #1 and #2 coordinate on (Hare, Hare) This equilibrium is strictly Pareto-inferior to (stag, stag). Whereas contraction choice would have it that (stag, stag) is the correct norm to settle upon, evolutionary game theory teaches us that it is unlikely that the Pareto-efficiency equilibrium will be selected in a process of repeated interactions. What is more the Pareto-efficient equilibrium is unstable: occasional deviations from this equilibrium will lead the population as whole to coordinate on (hare, hare) rather than (stag, stag).

^{xii} See Skyrms (1996, 2004), Binmore (1994, 2005), Hirofumi (1999)