

Game Theory and Philosophy

For the past several weeks, we have generalized the basic 2 person 2 strategy game to apply to show how cases with large numbers of players and large number of strategies can be analyzed. We also have explored cases in which no "deterministic" equilibrium existed in order to demonstrate the possibility of "stochastic" equilibria (with mixed strategies).

For the most part, we have used economic and political examples to illustrate how such game theoretic models can be used to analyze real world situations of interest to social scientists. [Last week's lectures by Brian Skyrms showed how game theoretic models can further extended to analyze cultural evolution.]

This week, I thought that we should spend a little time showing how game theory can be employed to analyze problems posed by philosophers. In many cases, a bit of game theory would have made their arguments much easier to understand.

I. From Plato's Republic: on the Importance of Justice (360 BCE)

- i. [Book 2] If you, please, then, I will revive the argument of Thrasymachus.
 - a. And first I will speak of the nature and origin of justice according to the common view of them.
 - a. Secondly, I will show that all men who practice justice do so against their will, of necessity, but not as a good.
 - a. And thirdly, I will argue that there is reason in this view, for **the life of the unjust is after all better far than the life of the just**--if what they say is true, Socrates, since I myself am not of their opinion.
 - a. But still I acknowledge that I am perplexed when I hear the voices of Thrasymachus and myriads of others dinning in my ears; and, on the other hand, **I have never yet heard the superiority of justice to injustice maintained by any one in a satisfactory way.**
 - a. ... because he has no misgivings about injustice and at every contest, whether in public or private, he gets the better of his antagonists, and gains at their expense, and is rich, ... And thus, Socrates, gods and men are said to unite in making the life of the unjust better than the life of the just.
- ii. [Book 2, continued]
 - a. Once more, Socrates, I will ask you to consider another way of speaking about justice and injustice, which is not confined to the poets, but is found in prose writers.
 - a. The universal voice of mankind is always declaring that justice and virtue are honorable, but grievous and toilsome; and that the pleasures of vice and injustice are easy of attainment, and are only censured by law and opinion.
 - a. They say also that honesty is for the most part less profitable than dishonesty; and they are quite ready to call wicked men happy, and to honor them both in public and private when they are rich or in any other way influential, while they despise and overlook those who may be weak and poor, even though acknowledging them to be better than the others.
- iii. [Book 8] Let us place **the most just by the side of the most unjust**, and when we see them we shall be able to compare the **relative happiness or unhappiness** of him who leads a life of pure justice or pure injustice.
 - a. The inquiry will then be completed.

- a. And we shall know whether we ought to pursue injustice, as Thrasymachus advises, or in accordance with the conclusions of the argument to prefer justice.
- iv. Plato's analysis of justice, and its role in government and in personal life is long and complex, but a good deal of insight in to the basic problem--in the context of a "republic" of like-minded persons can be discovered from game theoretic models.
- v. The "Justice Dilemma"

the "Justice Dilemma"		
	Player B's Choice	
Player A's Choice	Just	Unjust
Just	(3,3)	(1,4)
Unjust	(4,1)	(2,2)

- a. Notice that most of the essential features of the problem are captured in this modified PD game. Injustice is the dominant strategy, and an Unjust strategy profits from the strategies of the Just (4>1).
- a. Note also, that in a community of like-minded individual, that a just society is superior to an unjust society (3,3) > (2,2).
- a. [Recall that the types of governments analyzed by Plato attract certain types of individuals, and that Plato's ideal society attracts those predisposed to just behavior. From book 8, "Him who answers to aristocracy, and whom we rightly call just and good, we have already described. ... Then let us now proceed to describe the inferior sort of natures, being the contentious and ambitious, who answer to the Spartan polity; also the oligarchical, democratical, and tyrannical."
- a. The point here is not to say that a hundred pages of sophisticated discussion by Plato can be completely captured in one game matrix, but rather to say that a good deal of that argument can be represented in a such a matrix.
 - ◆ We can, however, show in less than one page that a society of the just, such as Plato argues will be attracted to the Republic, will be better off (relatively more happy) than a society of the unjust.
- a. In a longer analysis, we can also address whether such a community would be robust to "invasions" of the unjust--using techniques like those developed by Brian Skyrms last week.
 - ◆ We leave such repeated and/or evolutionary games for later in the course, but it is intuitively obvious that in the above game, this will depend upon the ability of the "guardians" to exclude and/or punish the unjust.
 - ◆ [from Book 3] "and so, their country being their mother and also their nurse, they are bound to advise for her good, **and to defend her against attacks**, and her citizens they are to regard as children of the earth and their own brothers."

II. Hobbes and Locke: On the Usefulness and Limits of a Social Contract

- A. About 2 thousand years after Plato attempted to characterize an ideal society grounded in justice, Hobbes (1651) and Locke (1689) undertook a fairly similar--if somewhat less ambitious--enterprise.
- i. They attempted to explain what a proper government, grounded in the assent of the people, would look like.
 - ii. They both begin with a state of nature (as Aristotle does in his analysis of the state), and use that unpleasant circumstance to motivate the emergence of government as a means of advancing common interests (as also mentioned in Plato and Aristotle).
- B. Hobbes regards the natural state to be war of "every man against every man."
- a. [*Leviathan*, Ch. 13] in the nature of man, we find **three principal causes of quarrel. First, competition; secondly, diffidence; thirdly, glory.**
 - a. Hereby it is manifest that during the time men live without a common power to keep them all in awe, **they are in that condition which is called war**, and such a war as is of every man against every man.
 - a. **In such condition** there is no place for industry, because the fruit thereof is uncertain: and consequently no culture of the earth; no navigation, nor use of the commodities that may be imported by sea; no commodious building; no instruments of moving and removing such things as require much force; no knowledge of the face of the earth; no account of time; no arts; no letters; no society; and **which is worst of all, continual fear, and danger of violent death; and the life of man, solitary, poor, nasty, brutish, and short.**
 - i. However, all is not hopeless. There are also forces that incline men toward peace:
 - a. The **passions that incline men to peace** are: fear of death; desire of such things as are necessary to commodious living; and a hope by their industry to obtain them.
 - a. And **reason** suggesteth convenient **articles of peace upon which men may be drawn to agreement**. These articles are they which otherwise are called **the laws of nature**, whereof I shall speak more particularly in the two following chapters.
 - ii. Unfortunately, peace is not easily obtained. Conventions based on trust will not be enough to produce peace.
 - a. [Chapter 14] If a **covenant be made** wherein neither of the parties perform presently, **but trust one another**, in the condition of mere nature (which is a condition of war of every man against every man) upon any reasonable suspicion, **it is void**:
 - a. but if there be a common power set over them both, with right and force sufficient to compel performance, it is not void.
 - a. [Obtaining peace requires a common power to enforce the peace.]
- C. The problems addressed in Hobbes and Plato may appear to be quite different.
- i. Hobbes' problem is in a sense more desperate than the problem addressed by Plato--peace is a matter of life and death, whereas the ideal society might be regarded as simply an improvement on already comfortable lives within existing imperfect societies.
 - ii. Plato attempts to characterize the ideal society in order to understand justice.

- iii. Hobbes attempts to understand anarchy (the state of nature) so as to explain irrevocable transfers of authority from the people to a sovereign.
- iv. However, it turns out that in game theoretic terms the problems are analytically surprisingly similar.

I. Anarchy and the Social Dilemma		
	Person B's Choice	
Person A's Choice	Peace	Attack
Peace	(3,3)	(1,4)
Attack	(4,1)	(2,2)

- a. Hobbes' state of nature is a war of every man against every man.
 - ◆ Notice that the dominant strategy in this game is attack.
 - ◆ And, thus, the Nash equilibrium can be thought of as a war of every man against every man (Attack, Attack).
- a. Notice also, that a common interest exists in peace.
 - ◆ Both players would be better off if they gave up war and agreed to peace.
 - ◆ ((3,3) is not only Pareto superior to (2,2), it makes both parties better off.
 - ◆ The two parties have a **common interest** in peace.
- a. However, peace is not a Nash equilibrium of this game.
 - ◆ That is to say, both parties have an incentive to renege on the peace agreement.
 - ◆ Thus, an unenforceable covenant based only on trust tends to return to war, as argued by Hobbes.
- v. To escape from the social dilemma--the war of every man against every other--Hobbes argues requires the creation of an outside authority that can change the payoffs of the game--"to compel performance" of the peace treaty.
- vi. He also argues that, because peace is in the interest of all, and is so valuable, that essentially everyone would agree to create such a power--the Leviathan.
 - ◆ **For the laws of nature, as justice, equity, modesty, mercy, and, in sum, doing to others as we would be done to**, of themselves, without the terror of some power to cause them to be observed, **are contrary to our natural passions**, that carry us to partiality, pride, revenge, and the like.
- vii. The "oath" taken to conclude a Social Contract [from ch. 17]

♦ **I authorize and give up my right of governing myself to this man, or to this assembly of men, on this condition;** that thou give up, thy right to him, and authorize all his actions in like manner.

- a. This done, the multitude so united in one person is called a COMMONWEALTH; in Latin, CIVITAS.
- a. This is the generation of that great LEVIATHAN, or rather, to speak more reverently, of that mortal god to which we owe, under the immortal God, our peace and defense. For by this authority, given him by every particular man in the Commonwealth, he hath the use of so much power and strength conferred on him that, by terror thereof, he is enabled to form the wills of them all, to **peace at home**, and mutual aid against their enemies abroad. ..

D. Hobbes does not literally mean that Leviathan will "force" each person to do their duties under the covenant of peace, but rather that it will enforce the law by punishing those who violate the peace. Leviathan thus transforms the game into:

Leviathan Solution to the Social Dilemma		
	Person B's Choice	
Person A's Choice	Peace	Attack
Peace	(3,3)	(1,4 - F)
Attack	(4 - F,1)	(2 - F, 2 - F)

- i. Note that as long as the penalty (fine) is greater than 1, the dominant strategy to the game under Leviathan is now Peace, and the new Nash equilibrium is (peace, peace).
- ii. In this manner, Leviathan can solve the problem addressed by Hobbes.

E. Whether a powerful government will actually perform this duty is not addressed by Hobbes, only its potential utility in obtaining peace.

- i. However, Hobbes argues that the value of peace is so great that the transfer of power (sovereignty) to the Leviathan is unlimited, and cannot be revoked.
- ii. This idea was later challenged by Locke, Rousseau, and Montesquieu (as well as numerous real world revolutionary movements, including those in the Netherlands, the United States and France).

F. Locke, for example, notes that some forms of social contract can be ruled out, because individuals would never have agreed to them. [From the Treatises on Government, pg. 310]

- i. But though men, when they enter into society, give up the equality, liberty, and executive power they had in the state of nature, into the hands of the society, to be so far disposed of by the legislative, as the good of the society shall require; yet it being only with an intention in everyone the better to preserve himself, his liberty, and property (*for no rational*

creature can be supposed to change his condition with an intention to be worse) the power of the society or legislative constituted by them, can never be supposed to extend farther than the common good; but is obliged to secure every one's property, by providing against those three defects above mentioned that make the state of nature so unsafe and uneasy.

ii. In terms of the games noted above, one can imagine people who voluntarily agree to move from the first game to the second, but **not to every game that might be constructed by Leviathan.**

- ♦ This reasoning implies that the contract cannot be unconditional, if it is to be a voluntary agreement as Hobbes argues.

Leviathan Problem that is Worse than the Social Dilemma		
	Person B's Choice	
Person A's Choice	Peace	Attack
Peace	(3 - T, 3 - T)	(1 - T, 4 - F - T)
Attack	(4 - F - T, 1 - T)	(2 - F - T, 2 - F - T)

- a. To illustrate this limit, suppose that Leviathan imposed taxes, **T**, on all the possible game outcomes, in addition to imposing fines, **F**, sufficient to solve the original social dilemma.
- a. We demonstrated above that $F > 1$ will avoid war.
- a. However, notice that if $T > 1$, the "taxes" will make both persons worse off under Leviathan than they were in the original state of nature.
 - ♦ Peace may still be generated, but the payoff associated with it is now below the original payoff in the state of war.
 - ♦ For example, if $T = 1.5$, the payoff associated with peace is $3 - 1.5 = 1.5$ which is lower than the payoff associated in the original social dilemma, $1.5 < 2$.
- iii. Although peace obtains, free individuals would not give up "war" for such a peace, even though war is unattractive.
 - a. How likely this is, depends in part on the relative magnitude of the payoffs from peace and war.
 - ♦ If the payoffs of peace are very high compared to war, say 3 vs. -200, as implied by Hobbes' analysis, then for a wide range of T's, there will still be advantages to adopting or creating a Leviathan.
 - ♦ Recall that Locke's characterization of the state of nature is much less dire than that of Hobbes, and that Rousseau's is even less dire than in Locke. In such settings, the risk of an "overtaxing" Leviathan is greater.
 - a. If peace is not as advantageous relative to the equilibrium in the state of nature, than many more "tax" policies by Leviathan will be problematic.

- iv. The **game matrix implies** that the acceptable range of social contracts requires Leviathan to be bound to the range where the "tax" is less than "3-2" in our original representation of anarchy.
 - a. How this can be accomplished is not clear in Locke or Rousseau, although Montesquieu suggests that dividing the power of government may increase liberty by constraining government.
 - a. (Note that one can think of Plato's process of educating and selecting "guardians" as an attempt to produce a leviathan--an unconstrained and powerful government--whose policies will lie in this range--e.g.whose policies will **only enhance** the welfare of the individuals living in their city-state.)

III. Rawls, Fairness, and the Social Contract

- A. Rawls like Locke and Rousseau is interested in the kinds of contracts that might be acceptable to individuals in a state of nature, but does not characterize the state of nature in any detail.
 - i. Instead, he simply assumes that a social contract is to be worked out, and must be accepted by unanimous consent.
 - ii. He attempts to determine what kinds of contracts would be accepted and concludes that:
 - a. "What must be shown is that a certain arrangement of the basic structure, certain institutional forms, are more appropriate for realizing the values of liberty and equality when citizens are conceived as such persons, that is (very briefly) as having the requisite powers of moral personality that enable them to participate in society viewed as a system of fair cooperation for mutual advantage.
 - a. So to continue, the two principles of justice read as follows:
 - ♦ Each person has an equal right to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with a similar scheme for all.
 - ♦ Social and economic inequalities are to satisfy two conditions: first, they must be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they must be to the greatest benefit of the least advantaged members of society." [Rawls, 1985:227]
 - iii. Rawls does not really characterize institutions, although he implicitly accepts modern democratic institutions--rather he leaves the institutions a bit implicit and analyzes the kinds of "end results" and "distributional formulas" that might be agreed to.
- B. In effect, Rawls is interested in the problem of choosing **which game to play**, insofar as a social contract tends to determine a payoff structure.
 - i. Individuals are free to choose, but the institutions of distributional justice affect the game payoffs that they might achieve.
 - ii. In the games above, the games were assumed to be symmetric, because Hobbes, Locke, and Rousseau all assume that everyone is essentially equal in a state of nature.
 - iii. This is not a reasonable assumption for Rawls, because in this case the degree of symmetry will be a consequence of the social contract--rather than a preexisting characteristic of the game setting (state of nature).

- C. To simplify a bit, suppose that there are just two types of people and three types of social contracts, each framing a particular "post contract" society. The social contract can be interpreted as the "rules of the game" played in the society that emerges after the contract is adopted.

		Game #1 (payoffs: A, B)	
		Person B's Choice	
Person A's Choice		Work	Shirk
	Work	(1, 1)	(1, 1)
	Shirk	(1, 1)	(1, 1)

		Game #2 (payoffs: A, B)	
		Person B's Choice	
Person A's Choice		Work	Shirk
	Work	(3, 3)	(2,1)
	Shirk	(1,2)	(1, 1)

		Game #3 (payoffs: A, B)	
		Person B's Choice	
Person A's Choice		Work	Shirk
	Work	(2.5, 5)	(2,0)
	Shirk	(0,4)	(1, 1)

- i. Suppose that players examine these three games and rank order them according to the predicted equilibria of the game.
- ii. (Note that the relevant comparisons are based on anticipated equilibria, rather than nonexistent outcomes. Why?)

- iii. In this case Players of type A will prefer Game 2 to Game 3 to Game 1, and Players of type B will prefer Game 3 to game 2 to game 1.
- iv. Unfortunately, because the games are symmetric, and because players know their interests, there will be no social contract. Game 1 can be rejected, but no agreement over games 2 and 3 emerges.
- v. [Veil of Ignorance] If players did not know whether they were type A or type B players, their calculations would be different--and it is quite possible that an agreement would emerge.
 - a. The veil of ignorance induces a symmetry in the games that was not present without the veil.
 - a. Suppose that half the people are of type A and half are of type B, and initially assume--in contrast to Rawls--that the players are risk neutral.
 - ♦ In this case, the players will rank order the games according to the expected payoffs, and both A and B will **agree about rank** order of the games: #3 (with and expected payoff of 3.5) > #2 (with an expected payoff of 3) > #1 (with a payoff of 1).
 - ♦ If both players are very risk averse, **as assumed by Rawls**, they will agree that #2 > #3 > #1, and #2 will be unanimously agreed to.
 - a. In this way, the veil of ignorance can produce agreements that would be impossible in a state where everyone understood exactly how the proposed "rules of the game" would affect each person's payoff.
- vi. [Note, however, that if people have different degrees of risk aversion, no agreement may be reached, even from behind a veil of ignorance.]

IV. Problems

- ♦ [Quote from Hume, 1739, pg 345]

But government extends farther its beneficial influence; and not contented to protect men in those conventions they make for their mutual interest, it often obliges them to make such conventions, and forces them to seek their own advantage, by a concurrence in some common end or purpose. **There is no quality in human nature, which causes more fatal errors in our conduct, than that which leads us to prefer whatever is present to the distant and remote, and makes us desire objects more according to their situation than their intrinsic value.**

Two neighbors may agree to **drain a meadow**, which they possess in common; because 'easy for them to know each others mind; and each must perceive, that the immediate consequence of his failing in his part, is the abandoning the whole project. But 'very difficult, and indeed impossible, that a thousand persons should agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expense, and would lay the whole burden on others.

Political society easily remedies both these inconveniences. Magistrates find an immediate interest in the interest of any considerable part of their subjects. They need consult no body but themselves to form any scheme for the promoting of that interest. And as the failure of any one piece in the execution is connected, though not

immediately, with the failure of the whole, they prevent that failure, because they find no interest in it, either immediate or remote. Thus bridges are built; harbors open' ram-parts rais' canals form' fleets equip' and armies discipline every where, by the care of government.

- A. Each of the three subtexts in color can be analyzed using game theory.
- B. Construct game theoretic representations of these "problems" beginning with the green quote.
 - i. Brian Skyrms argues that the "green" quote describes a "Stag Hunt," whereas James Buchanan argues that it describes a PD game or Free Rider problem.
 - ii. Develop a game matrix to characterize both versions of the game.
 - iii. Which, if either, seems the appropriate game to represent the essential features of "drain a meadow" game?
- C. The "Magistrate Game" is evidently an asymmetric game, in which only one person decides the level of a public service or good.
 - i. Use a 1x3 game matrix to represent Hume's theory of the magistrate.
 - ii. Show what happens if the magistrate's interests are opposed to those of its citizens rather than aligned with them.
- D. The under counting of the "future relative to the present problem" seems to contradict the normal assumptions of rationality.
 - i. Explain why.
 - ii. Now suppose that there is an asymmetric game analogous to the Magistrate game in which the present self controls the strategy variable, and the future self has no influence over current strategies. Use this representation and a 3x1 game to illustrate Hume's time discounting problem.
 - iii. Is there always such a conflict? Is such behavior necessarily irrational? Discuss.
- E. Find another famous philosophical excerpt that game theory can be used to clarify, illustrate, or to deepen understanding.