

Robert Jones
Phil 355

The Game Theory approach to Justice and Redistribution of Goods

Game theory is a branch of mathematics that studies strategic situations involving a number of players. Typically the goal of this study is to examine possible actions and outcomes in an attempt to maximize a player's returns. We have seen very similar ideas applied over the course of the semester in political philosophy, most notably in the writings of John Rawls, Robert Nozick, and David Gauthier. While each of these authors has a different political philosophy, each seems to be building their argument from the same basic position.

In his essays on Justice, Rawls outlines an idea of justice as an equal opportunity for individuals to pursue their life goals, and calls for a redistribution of wealth to ensure that social and economic inequalities are arranged so that they benefit the least advantaged of society, and continue to encourage fair equality of opportunity. To govern this, Rawls introduces his Difference Principle, the idea that changes to this redistribution structure must help the most disadvantaged groups to be considered an improvement.

Rawls' method of arriving at this structure looks very much like a game theory approach, and Rawls' general set-up for this society is done from what he calls the Original Position. From this position, individuals meet as players considering life or society as a strategic game. They only have knowledge of the outcomes as allowed from Rawls' Veil of Ignorance, and they agree upon an idea of The Right that will supposedly maximize their outcomes, or in this case, opportunity to pursue life goals. The Veil of Ignorance is a set of limitations on the knowledge that individuals have from the original position, removing those ideas or concepts that would lead to bias or otherwise affect

their decision-making in the original position. Individuals under the veil are stripped of their knowledge of race, gender, religion—everything but general facts about human nature and a thin idea of the good.

So Rawls has set up what can loosely be described as a non-zero sum game where the players (our people in the original position) have imperfect knowledge (the understatement of the century), and are trying to maximize their returns. The structure they will agree upon, he argues, is the structure that constitutes a just society. And up to this point, Nozick and Gauthier actually seem to agree with Rawls' approach. Where they disagree is in the decisions that each author feels individuals will make from the Original Position.

Rawls believes that individuals in the Original Position are unlikely to gamble with their eventual utility, arguing that from the veil of ignorance, those individuals that are incapable of knowing what the circumstances of their birth will be (race, gender, location, etc.), will choose an approach that seeks to maximize utility to the minimum positions in society, so that regardless of their plight, they still possess a near-equal opportunity to pursue their life goals. Because of this, they will commit to a plan that includes the redistribution of primary goods in a way that benefits the least advantaged members of society. Rawls is arguing that people in the Original Position will be influenced by their lack of knowledge and 'play it safe,' forgoing the possibility of higher utility for a guaranteed minimum level.

Nozick on the other hand, disagrees with this conceptualization of humanity. It is clear from his ideas on Libertarianism that Nozick believes that individuals in the Original Position will seek to maximize the utility of certain situations. He believes that

these people will gamble with regard to the circumstances surrounding their lives in an effort to achieve even greater utility in the process. The disagreement here is over the fundamental nature of human beings and the methods they will employ in their decision-making with regards to the outcome of the game.

Gauthier raises a different set of criticisms (although as we will see, these are based around similar criticisms of human nature and the possibility of cooperation). Gauthier claims that people are inherently self-interested, and he argues that mutual agreement in the original position, while beneficial to everyone in theory, are not beneficial in practice if players defect. That is, if people break or ignore the guidelines of Rawls' [non-binding] contract, then it is no longer in the self-interests of any particular individual to abide by the contract. Choices from the original position then, will reflect the self-interests of individuals in the position, and they will only rely on the bare minimum amount of cooperation among individuals.

These concerns over human nature seem to miss the point, however. It's rather un-charitable to claim that Rawls ever actually intended for a group of individuals to realize his set-up for justice (he must have been aware of the impossibility of such a task), and so the argument of what humans are likely to choose may be irrelevant. The purpose of the original position is not to speculate on the nature of humanity but rather to demonstrate that the set of outcomes that maximizes overall utility is the set that involves wide redistribution of wealth. And given the coercive power of the government in redistributing primary goods, while cooperation may be necessary for achieving an idea of justice, Rawls doesn't actually seem to be trusting to it in the real-world scenario. In

fact, he seems to understand that cooperation is unlikely to be guaranteed, and hence redistribution of goods is determined by social institutions and not voluntary.

In the text on left-wing libertarianism, Kymlicka outlines and discusses the prisoner dilemma, and how the best dual outcome (each criminal choosing not to confess, and getting one year in prison as a result) is completely dependent on cooperation, while the safer outcome (each prisoner confessing and getting 5 years) works independently of the choice of the other prisoner. The argument for maximizing the minimum level of society is similar: individuals in the original position forego the maximum possible benefit in order to play it safe, as the level of cooperation that Nozick suggests (voluntary charity that benefits the less-off members of society), seems much less practical or equally contrary to human nature. Rawls understands that while individuals agree to cooperate in the original position, they also concede that such cooperation is not guaranteed out of the position, and set up coercive power to compensate, making Gauthier's criticisms of the contract slightly less biting.

Rawls' approach does suffer from some problems, however. Just like in the prisoner's dilemma, playing it safe with regards to redistribution and cooperation can often lead to as many worse outcomes as better ones. Massive redistribution requires a minimum level of efficiency, and it may not always be the case that providing a service or good to everyone is feasible or desirable. We've certainly already discussed the need to have unequal power with regards to law enforcement and authority. But in a similar fashion to the 'double-confession' outcome of the prisoner's dilemma, it may be the case that we trade off something like good healthcare for 80 percent of citizens in favor of poor healthcare for all citizens.

Ultimately though, these concerns are not strong enough to warrant a dismissal of Rawls' approach to justice. Rather, minor adjustments are needed to account for the inconsistency between Rawls' conception of the Original Position and the nature of human beings before the game-theory approach to justice can be properly justified. Once we've accomplished that however, and allowed for the difficulty of cooperation, the approach not only yields a stronger, more practical idea of justice and redistribution, but it yields an approach that stands up to harsher criticism.