Models and Simulations: Discussion Questions Week 3

Topic: Game-theoretic approaches to morality.

Readings: Gauthier. "Morality and Advantage." Also, Gauthier *Morals by Agreement*. pp. 1-16.

Discussion Question 1: Consider Gauthier's. "Morality and Advantage." Suppose one interprets the phrase "system of principles" in "the thesis" to mean "a strategic profile in a game." In other words, suppose the thesis that Gauthier considers is the claim that an action is moral if and only if it is part of a strategic profile possessing certain properties.

1. Characterize types of strategic profiles that are "moral" according to "the thesis" using the following game theoretic terms: best response, Nash equilibrium, and Pareto dominance. To do so, you will first need to use these terms to characterize what Gauthier means by "disadvantageous" acts and "advantageous for everyone." You may need to ignore Gauthier's discussion of "no system" in his definition of "advantageous for everyone."

2. Can a moral strategic profile be a Nash equilibrium of a game?

3. Which, if any, of the strategic profiles in the Stag Hunt are moral according to the thesis?

4. In a one-person game (i.e. a decision problem with no uncertainty), are any actions moral according to the thesis? Why or why not?

5. Can you write down a payoff matrix that has more than one "moral" strategic profile?

6. According to the thesis so interpreted, does an individual's intentions matter in characterizing whether or not his actions are moral? How is this related to Gauthier's discussion of the "prudent but trustworthy"s man?

7. Gauthier claims that the thesis also fails to capture the fact that morality requires considerations of fairness. Construct a payoff matrix for a

prisoner's dilemma in which (i) the only strategic profile in which the payoffs are equal for the two players is not a moral strategic profile and (ii) in the remaining strategic profiles, including the "moral" one, the payoffs between the two players are grossly unequal.

Discussion Question 2: Explain Gauthier's claims that the "prudent but trustworthy man" does not behave in a way that we (intuitively) consider to be moral, even if he chooses to cooperate in prisoner's dilemma like situations.

1. Explain Gauthier's argument for this claim.

2. Open Ended: Suppose a "system of principles" in the thesis Gauthier considers is interpreted as "a decision rule." How would you translate the thesis using game-theoretic and decision-theoretic terms? What might the analog of "no system" be?

3. In what ways does the "prudent but trustworthy" man differ from the "constrained maximizer" that Gauthier discusses on page sixteen of *Morals by Agreement*?

4. What is the difference between *ex ante* agreement and *ex post* compliance? Why does Gauthier claim that Rawls' justification for the principles of justice fail to explain *ex post* compliance with said principles?

5. How does Gauthier's theory of constrained maximization solve the problem of *ex post* compliance?

6. Gauthier's conclusion in "Morality and Advantage" (1967) is that the question "Why should I be moral?" has no satisfactory answer. Does he still accept that conclusion in *Morals by Agreement* (1986)?