Game Theory

2.5 Application: Morality

George Mason University, Spring 2018

Gauthier [1991]

"Morality faces a foundational crisis. Contractarianism offers the only plausible resolution of this crisis. These two propositions state my theme. What follows is elaboration."

MacIntyre: "In the actual world which we inhabit the language of morality is in...[a] state of grave disorder...we have—very largely, if not entirely—lost our comprehension, both theoretical and practical, of morality."

Harman: "Moral hypotheses do not help explain why people observe what they observe. So ethics is problematic and nihilism must be taken seriously...An extreme version of nihilism holds that morality is simply an illusion...In this version, we should abandon morality, just as an atheist abandons religion after he has decided that religious facts cannot help explain observations."

Gauthier [1991]

Morality is supposed to justify our choices and actions.

However, it is not clear what grounds moral principles.

Moreover, Expected Utility Theory is also supposed to justify our choices and actions.

Deliberative justification based on EU Theory is more basic than *moral justification* since the former mode "relates to our deep sense of self."

"Deliberative justification does not refute morality. Indeed, it does not offer morality the courtesy of a refutation. It ignores morality, and seemly replaces it. It preempts the arena of justification, apparently leaving morality no room to gain purchase."

If you can't beat them, join them: morality must find its place within EU Theory.

Gauthier [1991]

Recall that individual rationality can lead to suboptimal outcomes in many interactive choice situations.

	confess	do not confess
confess	-5,-5	0,-10
do not confess	-10,0	-1,-1

Gauthier [1991]

The role of morality is to constrain individuals so that we can end up in the (Pareto) optimal outcomes.

	confess	do not confess
confess	-5,-5	0,-10
do not confess	-10,0	-1,-1

Gauthier [1991]

"Each person can see the benefit, to herself, of participating with her fellows in practices requiring each to refrain from the direct endeavor to maximize her own utility, when such mutual restraint is mutually advantageous. No one, of course, can have reason to accept any unilateral constraint on her maximizing behavior; each benefits from, and only from, the constraint accepted by her fellows. But if one benefits more from a constraint on others than one loses by being constrained oneself, one may have reason to accept a practice requiring everyone, including oneself, to exhibit such a constraint. We may represent such a practice as capable of gaining unanimous agreement among rational persons who were choosing the terms on which they would interact with each other. And this agreement is the basis of morality."

Gauthier [1991]

"Morality is not to be understood as a constraint arising from reason alone on the fulfillment of nonrational preferences. Rather, a rational agent is one who acts to achieve the maximal fulfillment of her preferences, and morality is a constraint on the manner in which she acts, arising from the effects of interaction with other agents."

Example: Assisting one's fellows.

Gauthier [1991]

Gauthier is not the first to attempt to derive morality from rationality. The novelty in his approach is really in how he addresses Hobbes' Foole who asks: Granting that it is rational to agree to certain constraints on individual choice, why is it rational to adhere to these constraints in any particular choice situation where you would be better off by breaking them?

Hobbes' answer: a supreme sovereign is needed to detect and punish cheaters.

Gauthier's main ideas are these:

- Accepting a moral practice is a matter of having a certain kind of disposition.
- *Constrained maximizers* who are disposed to comply with moral practices can expect to do better than *straightforward maximizers* who are not disposed to be moral.

Gauthier [1991]

"In plausible circumstances, persons who are genuinely disposed to a more rigorous compliance with moral practices than would follow from their interests at the time of performance can expect to do better than those who are not so disposed. For the former, constrained maximizers as I call them, will be welcome partners in mutually advantageous cooperation, in which each relies on the voluntary adherence of others, from which the latter, straightforward maximizers, will be excluded. Constrained maximizers may thus expect more favorable opportunities than their fellows. Although in assisting their fellows, keeping their promises, and complying with other moral practices, they forgo preference fulfillment that they might obtain, yet they do better overall than those who always maximize expected utility, because of their superior opportunities."
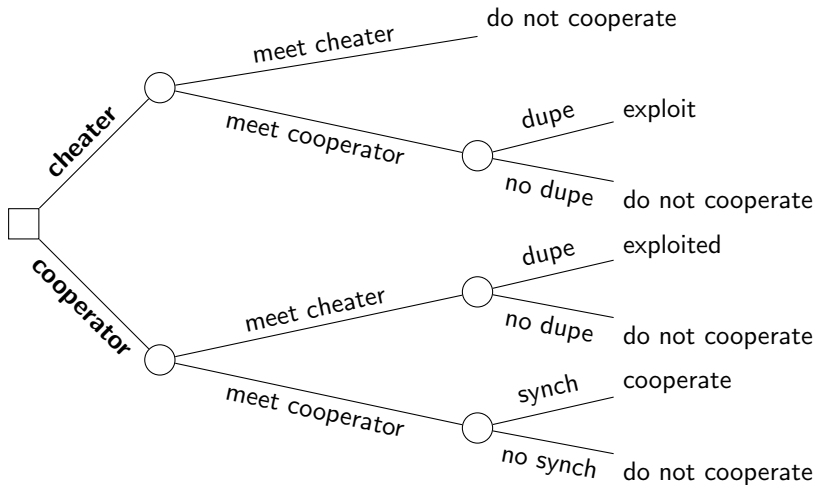
$u(\text{exploit}) = 1$
$u(\text{cooperate}) = u'$
$u(\text{do not cooperate}) = u$
$u(\text{exploited}) = 0$

where $0 < u < u' < 1$

$Pr(\text{meet cooperator}) = p$
$Pr(\text{dupe}) = q$
$Pr(\text{synch}) = r$

$EU(\text{cheater}) = u + pq(1 - u)$
$EU(\text{cooperator}) = u + pr(u' - u) - q(1 - p)u$

In the special case where $p = 1$,

$EU(\text{cheater}) = u + q(1 - u)$
$EU(\text{cooperator}) = u + r(u' - u)$

So $EU(\text{cooperator}) > EU(\text{cheater})$ iff $\frac{r}{q} > \frac{1-u}{u'-u}$

Gauthier [1991]

Gauthier's theory is a social contract theory in the spirit of Rawls.

We have not explicitly agreed to our existing moral practice. Rather, moral principles "are those that would secure our agreement ex ante, in an appropriate premoral situation. They are those to which we should have agreed as constituting the terms of our future interaction, had we been, per impossible, in a position to decide those terms. Hypothetical agreement thus provides a test of the justifiability of our existent moral practices."

Gauthier [1991]

Objection. Why is it rational to dispose oneself to accept the constraints that would be agreed upon in a premoral original position? Why isn't it rational to dispose oneself to accept the constraints that are *actually* in play in our society? After all, the latter disposition seems more pertinent to mutually advantageous interaction with one's fellows.

Reply. Constraints that would not secure agreement ex ante are unstable upon reflection on the existing moral order. Individuals whose prospects would be improved by renegotiation can make a strong appeal.

Objection. If there are situations where $EU(\text{cheater}) > EU(\text{cooperator})$, why not agree in the original position to cooperate conditional on society being such that r is high, q is low, and so on? Why agree to cooperate across the board? In the original position, you do not know whether society will be structured in a way that rewards constrained maximization.

Objection (Smith [1999]). Constrained and straightforward maximization are not the only options. Once other kinds of dispositions are taken into account, it is not clear that we should be constrained maximizers.

Smith [1991]

Ex. Fishermen.

"You and I are two fishermen inhabiting adjoining properties on a dangerous coastline. Hidden sandbars often cause our boats to run aground and our catch to be lost. Each of us can expect two such accidents in the coming year, one on our own sandbar, and one on our neighbor's sandbar. If either of us erected a lighthouse, it would prevent any accidents on the adjacent sandbar. The cost to each of us of a single accident is $500, whereas the cost per year of erecting and maintaining a lighthouse is $600."

Smith [1991]

	build	do not build
build	-$600,-$600	-$1100,-$500
do not build	-$500,-$1100	-$1000,-$1000

Smith [1991]

	build	do not build
build	-$600,-$600	-$1100,-$500
do not build	-$500,-$1100	-$1000,-$1000

Smith [1991]

In Smith's reconstruction of a simple version of Gauthier's argument where dispositions are transparent, each fisherman decides whether to be a constrained maximizer or straightforward maximizer.

CM: Forming the intention to build if you build, or not build if you do not; and then actually building if I expect you to build, or not building if I expect you not to build.

SM: Forming the intention not to build whatever you do; and then actually not building whatever I expect you to do.

Smith [1991]

	CM	SM
CM	-$600,-$600	-$1000,-$1000
SM	-$1000,-$1000	-$1000,-$1000

Smith [1991]

	CM	SM
CM	-\$600,-\$600	-\$1000,-\$1000
SM	-\$1000,-\$1000	-\$1000,-\$1000

Though ⟨CM, CM⟩ and ⟨SM, SM⟩ are both Nash equilibria of this game, CM dominates for each player.

Smith [1991]

However, there are other options besides CM and SM.

A fisherman might decide to be an *unconditional* or *radical cooperator*.

UC: Building one's lighthouse whatever one's partner does.

RC: Building one's lighthouse if and only if one's partner has chosen unconditional cooperation UC.

CM is not the best choice if your partner chooses UC or RC.

Smith [1991]

	CM	SM	UC	RC
CM	-$600,-$600	-$1000,-$1000	-$600,-$600	-$1000,-$1000
SM	-$1000,-$1000	-$1000,-$1000	-$500,-$1100	-$1000,-$1000
UC	-$600,-$600	-$1100,-$500	-$600,-$600	-$600,-$600
RC	-$1000,-$1000	-$1000,-$1000	-$600,-$600	-$1000,-$1000

Though $\langle CM, CM \rangle$ and $\langle SM, SM \rangle$ are still the only Nash equilibria of this game, CM no longer dominates for either player.