Econ 201 Lecture 20

Resolving Prisoner's Dilemmas and Other Commitment Problems

In games like the prisoner's dilemma, the hockey helmet game, the ultimatum bargaining game, and the satellite office game, players have trouble arriving at the outcomes they desire because they are unable to make credible commitments. Thus, if both players in the prisoner's dilemma could somehow reach a binding agreement to remain silent, each would be assured of getting a shorter sentence.

Hence the logic of the underworld code of Omerta, under which the family of anyone who provided evidence against a fellow mob member would be killed. Likewise, the helmet rule results in a better outcome for hockey players by committing them to wear helmets in circumstances in which they would otherwise choose not to do so.

Example 20.1. Will the restaurateur pay the waiter extra to provide good service?

The restaurateur wants his waiter to provide good service so that customers will enjoy their meals and come back in the future. If the waiter provides good service, the owner can pay him $100 per day. But if the waiter provides bad service, the most he can pay the waiter is $60 per day. The waiter is willing to provide bad service for $60 per day, and for $30 extra would be willing to provide good service. The owner's problem is that he cannot tell whether the waiter has provided good service. What will happen?

<table>
<thead>
<tr>
<th>Restaurant</th>
<th>Pay waiter $100/day</th>
<th>Pay waiter $60/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide good service</td>
<td>Second best for each</td>
<td>Best for owner, worst for waiter</td>
</tr>
<tr>
<td>Provide bad service</td>
<td>Best for waiter, worst for owner</td>
<td>Third best for each</td>
</tr>
</tbody>
</table>

Each side has a dominant strategy: Restaurant- pay $60/day; waiter- provide bad service. Outcome is lower-right cell and that is inefficient. The tip is a solution to this commitment problem. The Mafia's code of silence, hockey helmet rules, tips for waiters— all work by changing the material incentives facing the relevant decision makers. But it is not always practical to changes material incentives in precisely the desired ways.

Example 20.2. Will Sylvester leave a tip when dining on the road?

Sylvester has just finished a $100 steak dinner at a restaurant on Interstate 81 some 500 miles from home. The waiter provided good service. If Sylvester cares only about doing the best for himself that he can, will he leave a tip?

Unlike case of restaurant with local patrons, waiter in this restaurant has no way to penalize the diner in the future if he leaves no tip.

<table>
<thead>
<tr>
<th>Diner</th>
<th>Leave 15% tip</th>
<th>Leave no tip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide good service</td>
<td>Second best for each</td>
<td>Best for diner, worst for waiter</td>
</tr>
<tr>
<td>Provide bad service</td>
<td>Best for waiter, worst for diner</td>
<td>Third best for each</td>
</tr>
</tbody>
</table>

Dominant strategy for the waiter: provide bad service. Dominant strategy for diner: leave no tip. Again a worse outcome for each than if waiter had provided good service and diner had tipped.

Moral Sentiments as Commitment Devices

If commitment problems cannot be solved by altering the relevant material incentives, it may nonetheless be possible to solve them by altering people's psychological incentives. For example, feelings of guilt when they
cause harm to others, feelings of sympathy for the interests of their trading partners, feelings of outrage when they are treated unjustly, and so on may lessen the incentive to behave in opportunistic and mutually destructive ways.

**Example 20.3. In a moral society, will the business owner open a satellite office?**

The owner of a thriving business wants to start up a satellite of that business in a distant city. If she hires someone who manages the new office honestly, she can afford to pay a weekly salary of $1000— a premium of $500 over what the manager would have otherwise been able to earn-- and still earn weekly financial return of $1000 for herself. The owner's concern is that she will not be able to monitor the behavior of the satellite manager, and that this person would therefore be in a position to embezzle heavily from the business. The owner knows that if the satellite office is managed dishonestly, the manager can earn $1500, while causing the owner a financial loss of $500 per week. If the owner believes that the managerial candidate comes from a society whose members have been strongly conditioned to behave honestly, will she open the new office?

Suppose that the effect of the conditioning is to cause the managerial candidate to be willing to pay $10,000 to avoid the guilt he would feel if he managed dishonestly. The game tree for the game then looks like this:

In this case, the best choice for the owner at point B will be to open the satellite office, because she knows that at point C the manager's best choice will be to manage honestly. The irony, of course, is that the honest manager in this example ends up richer than the selfish manager in the similar example we considered in the last lecture.

**Are People Fundamentally Selfish?**

Without question, self-interest is an important human motive. Yet narrow self-interest is surely not the only human motive.

If you were in the second player's position in the ultimatum game, would you accept a proposal of $99 for the first player and only $1 for you?

- A 50-50 split is the most common offer.
- Highly one-sided offers are almost always rejected.

Do you tip in restaurants away from home?

- Tipping rates in restaurants patronized mostly by out-of-town diners are essentially the same as those in restaurants patronized mostly by local diners.

Other examples of behavior inconsistent with narrow pursuit of self interest:

- Family feuds
- Falklands war
- Rescues
- Bone marrow donation
- Return of lost wallets

**Is Honesty the Best Policy?**

My wife and I used to gather with several friends and our families for dinner at the Glenwood Pines on occasional Friday evenings. Our children loved going there because after eating they got to go into the bar with a pocketful of coins to play the pinball machines while their parents linger at the dinner table. One evening, my son Chris, then 9 years old, came back to the table with a furrowed brow to report that one of the machines had eaten several of his quarters.

“What did you do about it?” I asked him.

“I told the bartender, and he gave me some extra quarters,” he responded.

“So what’s the problem?” I asked, noticing that he still seemed troubled.

“I think the bartender gave me one more quarter than I lost,” he said.

“What do you think you ought to do?” I asked.
“Tell the bartender,” he replied resolutely.

As we left, Chris went up to the bartender and handed him a quarter, explaining what had happened. Several people sitting at the bar chuckled at this, and one told Chris that he was foolish to have given back the quarter.

Is honesty the best policy? Although armchair and professional philosophers have debated this question for millennia, no clear consensus has emerged. The answer seems to depend, after all, on whose perspective we have in mind by the term “best.” If we take the perspective of society as a whole, for example, most of us would say that honesty clearly is the best policy. (If forced to choose between living in a society in which everyone was honest, including you, or one in which no one was honest, would it really be a difficult decision?)

Disagreement begins, however, once we take the perspective of the individual. Cynical comments like the one directed at my son reflect the view-- by no means uncommon-- that although it would be nice if everyone were honest, we live in a competitive world and must seize our opportunities when they arise. Others counter that honesty pays dividends even at the individual level. Thus, they argue that there is always a chance that dishonest persons will be caught and punished; and they add that, quite apart from the material payoffs involved, the honest person typically enjoys greater peace of mind.

To this, the cynics respond that there are plenty of opportunities in which the chances of being caught and punished are negligible. For instance, the bartender would never have known if my son had kept the extra quarter, and even if he had known, it would have been too costly for him to do anything about it. Moreover, the cynics add, peace of mind doesn’t pay one’s bills.

The emerging science of evolutionary psychology contributes a fresh new twist to this tired debate. Evolutionary psychologists begin with the proposition that the human central nervous system is best understood as the product of Darwinian natural selection. Thus, they argue, the details of our cognitive, appetitive, and emotional repertoires are intelligible only when viewed as adaptations-- features of the organism that enhance its ability to survive and leave offspring.

This perspective might appear to place evolutionary psychologists squarely on the side of the cynics in the honesty debate. But although some evolutionary psychologists do take essentially this position, I will argue that the logic of Darwinian theory is in fact more consistent with the opposite view-- that honesty promotes not only the interests of the group, but those of the individual as well.

Why a Superficial Reading of Darwin Suggests that Honesty is Not the Best Policy

The number of descendants that an organism leaves behind depends in part on its ability to acquire food and other resources necessary to sustain life. Any characteristic that enhances that ability thus tends-- by definition-- to be favored by natural selection. Some traits, such as intelligence or keen eyesight, are beneficial both to the individuals who have them as well as to the larger populations in which they reside. Other traits, however, pose a conflict between the individual and group.

Consider, for example, the tendency for the hackle's on a dog's neck and back to rise when he and a rival are about to fight for the same mate. This mechanism serves the individual dog's purposes because, by making him appear to be larger, it increases the likelihood of his being able to intimidate his rival. (Evolution saw to it that dogs know better than to fight an opponent who is significantly larger.) From the perspective of dogs as a group, however, the hackle-raising mechanism is largely wasteful, for when all dogs raise their hackles, their rank ordering by apparent size is the same as if none had done so. The bodily resources required to sustain the hackle-raising mechanism could have been put to better uses-- say, by supporting sharper vision or a keener sense of smell.

One of Darwin’s central insights was that selection pressure is almost always more intense at the individual than at the group level. Thus, even though dogs as a species would fare better if none had hackle-raising mechanisms, any individual dog that lacked this mechanism would pay a prohibitive reproductive penalty. In evolutionary terms, his ability to see or smell better is of little consequence, after all, if it comes at the expense of being able to secure a mate.

Superficially, at least, the same Darwinian logic appears to work against the evolution of honesty. For the purposes of this discussion, suppose we define an honest person as someone who keeps a promise even though it would be advantageous for her to break it. (The argument I am about to construct can easily be adapted to other, more comprehensive, definitions.) Suppose further that there are many situations in which enforceable promises would be mutually advantageous to all parties involved. For the sake of concreteness, imagine a situation in which the owner of a business wants to hire a consultant to advise him about strategy for entering what he perceives to be a profitable new market. His concern is that in order for the consultant to advise him, he must give the consultant access to valuable confidential information about his business-- information that would be worth millions of dollars to the owner's competitors.

If the owner could find someone who could credibly promise to preserve the confidentiality of the information, both the owner and consultant would benefit. But if the owner thinks that people are fundamentally dishonest, he will predict that the consultant will reveal the confidential information to the highest bidder. And this implies that the owner's best option is not to hire the consultant in the first place.
To see how Darwinian forces appear to work against the evolution of honesty in situations like these, imagine an environment in which situations like the one described above were common, and in which an initial population contained some individuals who were genetically predisposed to be honest and others predisposed to be dishonest. If the proportion of honest people were high enough to begin with, it would have paid owners to hire consultants on the chance that they would turn out to be honest. It is true, in this case, that an honest consultant would have earned more than someone who was not hired to consult at all. But a dishonest consultant would have received a still higher payoff, by assumption.

The implication is that dishonest individuals will leave more offspring than others, causing dishonesty to proliferate. In the end, honest persons would appear destined for extinction, even though a population consisting only of dishonest persons would do worse than one consisting only of honest persons.

A Friendly Amendment

As even cynics must concede, however, abundant evidence contradicts the claim that everyone is dishonest. How might the impulses that drive such behavior have survived the ruthless culling of natural selection?

A simple change in the assumptions of our consultant story suggests one possible answer. In that story, owners had no practical means to distinguish honest consultants from dishonest ones. But suppose that honest individuals bore some identifying marker-- say, a red birthmark in the shape of an "H" on their foreheads. Owners would then be able to hire honest consultants and avoid dishonest ones. As a result, individuals with the honesty trait would receive higher payoffs, causing honesty to proliferate. This time it is the dishonest individuals who appear destined for distinction.

The assumption of a red birthmark is fanciful, of course, but there do in fact appear to be statistically reliable signals that distinguish honest persons from dishonest ones. The key to understanding the logic of these signals is to recognize that honest behavior is motivated not by rational calculation about advantage, but by emotion. Thus, the person who keeps a promise does so not because she calculates that she will be better off if she keeps it, but because she feels sympathy to the interests of the promisee, or because she would feel guilty if she broke her word.

As Darwin himself first pointed out in his 1872 book, The Expression of Emotion in Man and Animals, such emotions have characteristic signatures that are visible to all. For example, when my son reported to me that the bartender had inadvertently given him an extra quarter, his concern was evident not just in the words he chose, but also in the expression upon his face.

The stick-figure face below represents my attempt to capture the essence of that expression-- the eyebrows elevated at the bridge of the nose and slanting downward toward the edge of the face, the furrows at the center of the brow. This expression is produced by a complex combination of facial muscle contractions-- principally, of the pyramidal muscles at the bridge of the nose and the corrugator muscles at the center of the brow. This combination is extremely difficult to summon willfully.1 (If you are skeptical, sit before a mirror and try it!) Yet the expression appears spontaneously on the faces of subjects who are experiencing the emotions of sadness, concern, or perhaps even guilt. Subjects living in any culture on earth can reliably identify the emotions that correspond to this simple drawing.

In addition to facial expressions, there are other statistically reliable signals of affect. Thus, posture and other elements of body language, the pitch and timbre of the voice, the rate of respiration, and even the cadence of speech are systematically linked to underlying emotional states.2 Because these linkages are beyond conscious

---

control in most people, it is difficult to conceal the experience of certain emotions. And it equally difficult to feign
the characteristic expressions of these emotions on occasions when they are not actually experienced. We are
therefore able to employ such clues to form estimates of the emotional makeup of others, and to form judgments
about their character.  

If we were able to make perfectly reliable character judgments, we could always avoid dishonest persons in
ventures that require trust. Such persons would therefore earn lower payoffs than honest persons, and, according to
Darwinian theory, would eventually be driven from the population. Evolution has been going on for billions of
years. That so many dishonest persons remain among us suggests that perfectly reliable character judgments are
either impossible or at least costly.

Consider the implications of this last possibility— that if we incur the costs of scrutinizing potential trading
partners, we can make accurate character judgments about them. Would it pay to incur these costs? That would
depend on how likely it is that a randomly chosen trading partners is dishonest. If that likelihood is high, then it
pays to be vigilant, just as it pays to install an expensive security system in a house located in a high-crime
neighborhood. But if the overwhelming majority of one’s potential trading partners are honest, then extreme
vigilance will simply be wasteful.

These observations suggest a tendency for populations to gravitate toward a stable mix of honest and
dishonest persons. Because populations that consist almost exclusively of honest persons would discourage
vigilance in the choice of trading partners, opportunities would be ripe for dishonest persons in those populations.
And this would mean that the share of dishonest persons would grow. Conversely, because populations with only a
small minority of honest persons would strongly reward vigilance in the choice of trading partners, honest persons
would avoid interacting with dishonest persons, and the resulting higher payoffs to honest persons would cause their
share of the population to grow. At some intermediate mix of the two types, these countervailing forces will be in
balance, and the composition of the population will therefore tend to stabilize.

The Problem of Mimicry

For honest individuals to be able to survive in competition with dishonest individuals, there must, as noted,
be some means by which honest individuals can identify, and interact selectively with, one another. But if there is
advantage in being honest and perceived as such, there is even greater advantages in appearing honest. After all, a liar who appears trustworthy will have better opportunities than one who glances
about furtively, sweats profusely, speaks in a quavering voice, and has difficulty making eye contact. Indeed, he
will have the same opportunities as a genuinely honest person, but will get a higher payoff because he exploits them
to the fullest.

The behavioral clues we employ to reach character judgments are obviously far from perfect. Even
experienced professional polygraph experts cannot be sure when someone is lying. If the ability to mimic the
signals of trustworthiness were perfect, the mechanism that sustains the evolution of honest individuals simply
could not work. Fortunately for honest individuals, however, instances of perfect mimicry do not appear to exist in
nature.

Consider the monarch butterfly has a whose foul taste that protects it from predators, who have learned to
associate the objectionable taste with the monarch’s distinctive wing markings. The similar wing markings of the
viceroy butterfly provides it with a measure of protection against the same predators, even though the viceroy has
not incurred the cost of producing the bad taste.

The protective cover provided by the monarch depends, of course, on the continued presence of
sufficiently many monarchs in the environment to keep predators on guard against the foul taste. But if the viceroy
were able to mimic the monarch's wing markings perfectly and without cost, the protective power of these markings
would soon decay— for if the viceroy’s wing markings were a perfect copy of the monarch’s, it would be just as
likely as the monarch to escape predation, even though it hadn't expended the bodily resources to manufacture the
foul taste. The viceroy's share of the population would grow because of this advantage, and predators would eventually lose their incentive to avoid the wing markings.

But this has not happened, leading us to conclude that perfect mimicry either takes time or entails
significant costs. The fact that the bearer of the genuine trait has the first move in this game will often prove a
decisive advantage. The monarch's wing markings are themselves evolving, and it is more difficult to hit a moving
target than a stationary one.

Similar logic applies to those who would mimic emotional traits. If the signals we use for detecting these
traits had no value, we would have long since ceased to rely on them. The inevitable result is an uneasy balance
between people who are really honest and others who merely pretend to be.

--

\footnote{Observable expressions of emotion are of course not the only reliable clues to character. For a discussion of the role of reputation and other factors, see chapter 4 of my Passions Within Reason.}
The popular impression of Darwin's contribution is that nature is "red in tooth and claw," that only the opportunist can survive the ruthless pressures of natural selection. Even many sophisticated Darwinians cling to the belief that honesty in one-shot encounters with strangers cannot survive these pressures.

This view collapses, however, if people are able to make reasonably accurate character assessments. In the end, the question of whether people have this ability is an empirical one. Elsewhere, Tom Gilovich, Dennis Regan, and I have shown that even on the basis of brief encounters involving strangers, experimental subjects are adept at predicting who will cooperate and who will defect in prisoner's dilemma games.\(^4\) Thus, in one version of our experiments the base rate of defection was only 26 percent, but the accuracy rate of predicted defections was more than 56 percent. It seems reasonable to expect that predictions regarding others whom we know well would be even more accurate.

If you lost an envelope containing $1000 in cash at a crowded concert, can you think of someone you feel sure would return it to you if he or she found it? If so, then you accept the central premise of my friendly amendment to the traditional Darwinian account of honesty. As long as it is possible for honest individuals to identify at least some others who are also honest, and to interact selectively with them, such individuals can survive in competitive environments. If this does not quite imply that honesty is the best policy at the individual level, it does say that honesty is a policy that individuals can live with.