CLASSROOM GAMES: A PRISONER'S DILEMMA

Charles A. Holt and Monica Capra*

Abstract

Game theory is often introduced in undergraduate courses in the context of a prisoner's dilemma paradigm, which illustrates the conflict between social incentives to cooperate and private incentives to defect. We present a very simple card game that efficiently involves a large number of students in a prisoner's dilemma. The extent of cooperation is affected by the payoff incentives and by the nature of repeated interaction. The exercise can be used to stimulate a discussion of a wide range of topics such as bankruptcy, quality standards, or price competition.

Keywords: prisoner's dilemma, game theory, experimental economics, classroom experiments.
JEL codes: C72, C92

The prisoner's dilemma is an important paradigm that illustrates the conflict between social incentives to cooperate and private incentives to defect. We present a very simple card game that quickly and conveniently involves a large number of students in a prisoner's dilemma. The extent of cooperation is often affected by the payoff incentives and by the nature of repeated interaction, even in finite-horizon situations where standard theory predicts uniform defection. We have used this game to stimulate discussion of a wide range of topics such as bankruptcy, quality standards, imposition of trade barriers, provision of public goods, price competition, etc. Appropriate classes include: principles, intermediate microeconomics, game theory, experimental economics, and topics classes where the dilemma arises, for example, environmental economics, public economics, law and economics, managerial economics, and industrial organization.

PROCEDURES

Playing cards are convenient for classroom games because they allow the instructor to

* We wish to thank Sanem Erucar, Matthew Moore, and Roger Sherman for helpful suggestions. This research was funded in part by the National Science Foundation (SBR-9617784 and SBR-9818683).

implement the experiment quickly, even in large classes. Cards also permit students to keep their choices secret until asked to reveal them. In addition, there is no need to collect and match "decision sheets." Other devices can substitute for cards, but we have found ordinary playing cards to be inexpensive, durable, and easy to handle. To conduct a classroom prisoner's dilemma, all you need is a single deck of playing cards and copies of the instruction and record sheet that are provided in the Appendix. The instructions will fit on a single page, in 10 point type, but you may have to adjust the margins a little.

Begin by giving each student a copy of the instructions and two playing cards, one red card (Hearts or Diamonds) and one black card (Clubs or Spades). The instructions, which are read aloud, explain that each person will be paired with another person in the room after they make their card play choices. The pairing is done by the instructor, who points to two people selected spontaneously at random, and asks them to reveal their decisions. Choices determine earnings in a very simple and intuitive manner: playing a red card increases one's own earnings by $2, and playing a black card does not change one's own earnings but increases the partner's earnings by $3.[1] Only the card color (not the number) matters, so it is best to use a deck for which the back side is neither red nor black. To reduce confusion, you may write the connection between card colors and earnings on the blackboard as "Red: your earnings increase by $2; Black: your earnings do not change but the partner's earnings increase by $3." This procedure yields a prisoner's dilemma shown in Table 1, where the first number in each payoff pair is the payoff for the row player and the second number is the payoff for the column player.[2]

A convenient way to proceed is to call on all students in a given row of desks to make their decisions. Ask them to play either their red card or their black card, and to hold it to their chests. In this manner you can guarantee that no one else sees their choices and you can see

---

[1] The choices can also be interpreted as: playing a red card "pulls" two dollars to oneself, and playing a black card "pushes" three dollars to one's partner. This push/pull explanation is from Aumann (1987). Dixit (personal communication) has written computer programs in which participants make decisions by deciding whether to push or pull a playing card image on screen.

[2] The same game can be explained in a number of different ways. For example, the game in Table I can be described equivalently as: "playing a red card increases one's payoff by $6 and reduces the other's payoff by $4, and playing a Black card increases one's payoff by $4 and reduces the other's payoff by $1." This is an interesting way to frame the same game, which if it affects behavior, may provide some topics for class discussion.

Table 1. A Prisoner's Dilemma with Low Gains from Cooperation

		Column Player	
		black	red
Row Player	black	(3, 3)	(0, 5)
	red	(5, 0)	(2, 2)

(Data: 17 percent cooperative choices in one-shot games, 58 percent cooperative choices in repeated matchings, omitting the final period.)

when all the students have made their card choices. Then, you can pick pairings at random by saying: "You and you, please reveal your cards." At first, you will need to remind students how to calculate earnings, but soon you will be able to proceed quickly with the next pairing. Students record their own earnings for the first period on the first row of the record sheet. If there is an odd number of people in the row of desks, one person will not play in that period. This procedure is repeated for each of the remaining rows.

Table 2. A Prisoner's Dilemma with High Gains from Cooperation

		Column Player	
		black	red
Row Player	black	(8, 8)	(0, 10)
	red	(10, 0)	(2, 2)

(Data: 58 percent cooperative choices in one-shot games.)

As noted in the instructions, payoffs change in the second period of the game. The increase in the partner's payoff from playing a black card is increased from $3 to $8. This increases the gains from cooperation, without changing the incentives to defect, as shown in Table 2. Announce the new payoffs by rewriting the setup on the blackboard as "Red: your earnings increase by $2; Black: your earnings do not change but your partner's earnings increase by $8."[3] Then go back to the first row, ask students to make a choice and hold the card against their chests, then match them with someone else, etc. If you choose pairings nonsystematicaly and spontaneously, it is unlikely that you will inadvertently match the same people together when you return to this row in the second period.[4] Another interesting variation, which can be implemented during this classroom game, is to match each person with the same partner for the final three periods after announcing this treatment aloud. This is best done a row at a time: after all have been matched once, the same pairings can be used a second time, and then a third time, before moving on to the next row.

With a principles class of size 24, it took less than 25 minutes to read instructions and complete five periods, which left plenty of time for discussion. For a larger class, you could conduct the period 1 game for several rows, the period 2 game for several other rows, and the period 3 repetition for the remaining rows.[5] With more than about eighty students, it is better to bring some people to the front of the classroom and let the others watch. Speed is important to prevent the audience from losing interest. Specifically, in very large classes, time can be saved by using fewer participants and treatments.

It increases interest to announce in advance that you will pick one person at random, *ex post*, and pay a percentage of that person's earnings.[6] Based on a payout rate of 10 percent, the student selected from the class described above received $2.10. It is possible to do this exercise with purely hypothetical payoffs, but even a small monetary payment makes it easier to deal with questions like: "What am I supposed to do? Should I be trying to get the most money for

---

[3] Defection may be common with some types of students, for example, cohorts in a business school that compete with each other in other classes. Therefore, it may be necessary to increase the gains from cooperation even in period 1 to ensure some diversity of decisions.

[4] Of course it is possible to use a random device to match people, like assigning numbers to students and drawing pairs of numbered ping pong balls from a bucket, but to do so takes more time than it is worth in a classroom experiment.

[5] To avoid confusion about payoffs, you should ask the students who are only playing the single-period game to circle the first row in the record sheet. Similarly, students who are only playing the high incentive to cooperate setup should circle the second row, etc.

[6] A possible selection method using 10-sided dice is described in the instructions. An alternative is to ask people to write down their birthday dates (not the year) and choose the person whose date is the closest to some prominent date, like the beginning of spring break, or at the University of Virginia, Thomas Jefferson's birthday.

myself?"

DISCUSSION

The game is clear enough for someone to realize the conflict between the gains from cooperation (choose the black card) and the private incentive to defect (choose the red card). Nevertheless, it is important to let students articulate this problem. One way to proceed is to let students in one of the rows of desks engage in a brief discussion between periods two and three, before the final three repeated matchings. At least some students will agree on choosing the black card, or will try to encourage others to cooperate. Discussion is likely to reduce defection, but some defection usually persists, especially if you do not permit students to renew the discussion after period 4. Avinash Dixit, in private communication, indicates that those who were more vocal about cooperating were generally more likely to defect in a prisoner's dilemma.[7]

The prisoner's dilemma is an important paradigm that economists use in the analysis of a wide variety of strategic situations.[8] This classroom exercise can be tied to many different applications, such as price-competition, bankruptcy and public goods. For example, in the case of price competition, each seller would be better off cutting price, but they would both benefit if they could (illegally) agree to raise price. It helps to start discussion with an application like this, covered on an intuitive (i.e., nonnumerical) level. Moreover, this is an example where the prisoner's dilemma is not bad for all; it benefits consumers to the extent it induces price competition. In fact, antitrust laws are designed to prevent firms from cooperating in setting higher prices. Other applications will be discussed later.

Before you write the payoff table on the board, students should be given a chance to discuss incentive effects. One way to begin is to ask them how behavior changed when the

---

[7] Avinash Dixit of Princeton University, in correspondence dated May 1, 1995, describes a computer program that implements a prisoner's dilemma. He observed that some of the participating students began talking among themselves prior to entering the decisions on their computers.

[8] The importance of the prisoner's dilemma paradigm is indicated by the fact that it is often used to introduce game theory in introductory classes. Indeed, many students may get the incorrect impression that game theory is nothing more than the prisoner's dilemma.

payoffs were changed in period two. Cooperation in our class increased from 17 percent in period one to 58 percent in period two. If the data from your class also show the same trend, you should ask students for an explanation. The discussion should focus on the bigger gain from cooperation in the latter setup, although the two-dollar incentive to defect is the same in both Tables 1 and 2. This can be followed with a question about how to change the payoffs in a manner that would reduce cooperation, (i.e., reduce the payoffs from choosing the black card or raise the payoffs from choosing the red card).

At this point it is useful to help students construct the payoff table. In advanced classes, the table can be used to illustrate the notion of a Nash equilibrium in which neither person has a unilateral incentive to deviate. In Tables 1 and 2, the red/red decisions constitute a Nash equilibrium since neither person can gain by deviating (choosing black). (Stress that the profitability of deviations from a Nash equilibrium are evaluated assuming that deviations are unilateral). The next step is to relate payoff changes from diagonal movements in the table to gains from mutual cooperation. For example, the difference between the noncooperative (2,2) outcome and the cooperative (3,3) outcome in Table 2 is much less than the corresponding diagonal difference in Table 2. In contrast, the gains from unilateral defection are measured by downward movements for the row player and rightward movements for the column player. These gains from defection are the same ($2) in each table.

The discussion of payoff tables can be tied back into the duopoly price competition example.[9] The payoffs in Table 1 could represent firms' profits for high or low price choices, assuming a unitary elasticity of demand. Suppose that the cooperative (black card) decision represents choosing a price of $4, and that both firms sell a single unit when they choose this price. With a constant marginal cost of $1, the profits will by 4×2 — 1 = $3 for each firm, as shown in the upper left box of Table 1. With prices of $2, each firm sells 2 units, which yield profits of 2×2 — 2 = $2 each, as shown in the lower right box. If only one firm chooses the low price of $2, it will sell all 4 units and earn 2×4 — 4 = $4, and the other earns nothing, which

[9] Also, see Seiver (1995) for a classroom oligopoly price game presented in matrix form.

corresponds to the off-diagonal boxes, with the 5 replaced by a 4 in each.[10]

A comparison of Tables 1 and 2 shows that the \$2 gain from defection is held constant as the (diagonal) gain from cooperation is increased from \$1 in Table 1 to \$6 in Table 2. With a different class, we held the gain from cooperation constant at \$1 and increased the gain from defection from \$2 to \$7. This change caused the rate of cooperation to fall from 22 percent to 11 percent, as we switched from the first period setup to this new one. A higher value from choosing the red card corresponds to greater profits for a unilateral price cut in the duopoly interpretation, perhaps resulting from a lower marginal cost or less product differentiation.

Repetition will almost always increase cooperation; in our class it increased from 17 percent in the one-period matching to 58 percent in the two initial periods of the repeated matching (with the first period payoffs).[11] Interesting discussion may arise when you ask explanations from students who defected in period one and cooperated in the repeated-pairing periods. Students may mention that they could increase earnings if both partners cooperate, and that the repetition allowed them to signal their intentions to cooperate. Repetition is a key element in many economic applications, such as the oligopoly pricing example.

Some related topics of discussion include end-game effects, which should come up in a game-theory class. In the repeated matchings, the rate of cooperation declined from 58 percent in the two initial periods to 33 percent in the final period. You can ask students to explain why the rate of defection increased in the final period of the three-period pairing. Those who started cooperating in the repeated game hoping to signal others to do so too, would probably realize that in the last period they could gain by defecting. Students may provide other reasons for defection in the final period, for example, there is no chance that defection will be punished later, or defection in the final period is intended to punish someone who defected in an earlier period. In a game-theory class, point out that the Nash equilibrium is to defect in the final period, and by the usual backward-induction arguments, in all periods.

Once you finish with the discussion of the results of your classroom experiment, you

---

[10] A lower marginal cost of 1/2 for each of the third and fourth units would produce Table I exactly.

[11] By coincidence, the effect of repetition was exactly the same, in the aggregate, as the effect of the increase in the gains from cooperation in period two.

should talk about applications so that students realize the practical relevance of this exercise. For instance, some economic interactions have a natural final period, as is the case when a tourist is not likely to make a repeat purchase, and a seller may have an incentive to offer a low-quality product. Different institutions and practices have developed to help sellers get out of a prisoner's dilemma where sellers have a private incentive to offer low quality, even though all sellers would earn more money if they could enforce high quality standards that will increase consumer demand. Some of these quality-assurance institutions include chain stores, franchise industries, industry quality standards, warranties and guarantees, and better business bureaus.

Bankruptcy negotiations provide another interesting example of a prisoner's dilemma. When the value of a debtor firm is less than the total amount of the creditors' loans, each creditor has a private incentive to liquidate its loan in order to recover the money. If all lenders try to liquidate, the firm will have to sell its assets, often at prices that do not allow all creditors to recover their investments. In this case, creditors could be better off if they allowed the firm to reorganize and continue operation, especially when the scrap value is less than the firm's value in operation. In this environment, the role of bankruptcy law is to help creditors escape the prisoner's dilemma and let the debtor firm reorganize instead of liquidate.

Finally, the provision of public goods, or the lack of it, is another application of the prisoner's dilemma. Because public goods are nonexcludable, each person has a private incentive not to contribute to the provision of such a good, even if everyone would be better off if all contributed. The role of government here is to help citizens avoid prisoner's dilemma outcomes in which key services like clean water, defense, roads, are not privately provided.[12] In fact, the same general approach (using playing cards) can be used to set up interesting n-person public goods games. For example, playing Red might mean keeping $6 (not contributing to the public good), and playing Black might mean increasing each other person's earnings by $2 (contribute to a pubic good). Free riding is a dominant strategy since you because $6 by contributing, regardless of what the others do. The socially optimal outcome, in contrast, is for all to contribute as long as there are at least 4 others who would get the $2 benefit from a contribution.

---

[12] Davis and Holt (1993, Chapter 6) offer a more precise discussion of the relationship between a prisoner's dilemma and the voluntary provision of a public good.

This variation can lead to interesting discussions of how cooperation rates may depend on group size. Class experiments involving the voluntary provision of a public good are discussed in Holt and Laury (1996).

By discussing a variety of applications, students will learn to spot cases where the prisoner's dilemma appears, that is, where there is a dichotomy between the individual incentive to defect and the social incentive to cooperate.[13]

FURTHER READING

There is a long tradition of prisoner's dilemma experiments in economics and social psychology. It is well known that behavior in these games is responsive to incentives and repetition. For example, Sherman (1971) applies this dilemma to oligopoly and notes that the structure of prisoner's dilemma payoffs can be related to market demand elasticity and cost conditions. Increases in the payoff differences along the diagonal would represent a less elastic market demand, which increases the gains from cooperation. Moreover, a high marginal cost can encourage cooperation in repeated oligopoly games because firms might find it more profitable to restrict output (increase prices).

More recently, Cooper, DeJong, Forsythe, and Ross (1996) report data from experiments that show more cooperation in finitely repeated prisoner's dilemma games than in one-shot games (even though the Nash equilibrium is to defect). They observe that cooperation is positive and declines over time in each case, which is consistent with previous studies (see Dawes 1980 and Roth 1988 for discussion and references). There was not universal defection in the final period of the finitely-repeated prisoner's dilemma games, possibly resulting from the positive feelings (reciprocity) that had developed during the initial cooperative phase (Dawes and Thaler 1988).

---

[13] See Capra et al. (1998) for an experimental study of behavior in a particularly interesting generalization of the prisoner's dilemma known as the "traveler's dilemma."

APPENDIX

We are going to play a card game in which everybody will be matched with someone on the opposite side of the room. I will now give each of you a pair of playing cards, one red card (Hearts or Diamonds) and one black card (Clubs or Spades). The numbers or faces on the cards will not matter, just the color. You will be asked to play one of these cards by holding it to your chest (so we can see that you have made your decision, but not what that decision is). Your earnings are determined by the card that you play and by the card played by the person matched with you. If you play your red card, then your earnings in dollars will increase by $2, and the earnings of the person matched with you will not change. If you play your black card, your earnings do not change and the dollar earnings of the person matched with you go up by $3. If you each play your red card, you will each earn $2. If you each play the black card, you will each earn $3. If you play your black card and the other person plays his or her red card, then you earn zero and the other person earns the $5. If you play red and the other person plays black, you earn the $5, and the other person earns zero. All earnings are hypothetical, except as noted below.

After you choose which card to play, hold it to your chest. We then tell you who you are matched with, and you can each reveal the card that you played. Record your earnings in the space below. (Option: After we finish all periods, I will pick one person with a random throw of dice and pay that person 10 percent of his or her total earnings, in cash. All earnings for everyone else are hypothetical. To make this easier, please write your name: ____________________ and the identification number that I will now give each of you: _________. Afterwards, I will throw a 10-sided die twice, with the first throw determining the "tens" digit, until I obtain the ID number of one of you, who will then be paid 10 percent of his or her total earnings in cash.) Any questions?

To begin: Would the people in the row that I designate please choose which card to play and write the color (R or B) in the first column. Show that you have made your decision by picking up the card you want to play and holding it to your chest. Everyone finished? Now, I will pair you with another person, ask you to reveal your choice, and calculate your earnings. Remember to keep track of earnings in the space provided below. Finally, please note that in period 2 you will be matched with a different person, and payoffs will change. In period 3 you will be matched with a different person and payoffs change again, but you get to play with him/her in the last three periods.

period	your card (R or B)	other's card (R or B)	your earnings
1	_____	_____	_____
2	_____	_____	_____
3	_____	_____	_____
4	_____	_____	_____
5	_____	_____	_____

References

Aumann, R. 1987. Game theory. In J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave*. London: MacMillan Press, 1987.

Capra, C. M., J. K. Goeree, R. Gomez, and C. A. Holt. 1998. Anomalous behavior in a traveler's dilemma? forthcoming in the *American Economic Review*.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross. 1996. Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior 12*(February): 187-218.

Davis, D. D., and C. A. Holt. 1993. *Experimental economics*. Princeton: Princeton University Press 1993.

Dawes, R. M.. 1980. Social dilemmas. *Annual Review of Psychology*. *31*, 169-93.

Dawes, R. M. and R. H. Thaler. 1988. Anomalies: Cooperation. *Journal of Economic Perspectives 2*(Summer): 187-97.

Holt, C. A., and S. K. Laury. 1997. Voluntary provision of a public good. *Journal of Economic Perspectives 11*(Fall): 209-15.

Roth, A. E. 1988. Laboratory experimentation in economics: A methodological overview. *Economic Journal 98*(393): 974-1031.

Seiver, D. A. 1995. A simple game theory experiment for teaching oligopoly. *Classroom Expernomics 4*(Fall): 1-3.

Sherman, R. 1971. Empirical oligopoly. *Kyklos 24*(1): 30-49.