

# MW24.2 Experimental Economics (SS2020)

## Cooperation Games: Prisoner's Dilemma

Olexandr Nikolaychuk

### Prisoner's Dilemma

⇒ originally introduced by Melvin Dresher and Merrill Flood (1950) to test the Nash Equilibrium predictions

	C	D
C	$a, a$	$d, c^*$
D	$*c, d$	$*b, b^*$

exa)

	C	D
C	2, 2	0, 3*
D	*3, 0	*1, 1*

C: "cooperate"

D: "defect"

$$c > a > b > d$$

$\{D; D\}$  → dominant strategy equilibrium  
→ Nash Equilibrium (NE)

### Predictions:

- one-shot → defect (NE)
- repeated finite → defect (SPNE)
- repeated infinite → cooperative play *can* be sustained as equilibrium play

⇒ Folk Theorem [Friedman, 1971]<sup>1</sup>:

- \* sufficiently patient players
- \* grim trigger strategies

### Behavioral Data:

⇒ significant share of subjects exhibit cooperative behavior [Dawes and Thaler, 1988]<sup>2</sup>

---

<sup>1</sup>James W. Friedman. A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, 38(1):1–12, 1971

<sup>2</sup>Robyn M. Dawes and Richard H. Thaler. Anomalies: Cooperation. *The Journal of Economic Perspectives*, 2(3):187–197, 1988

## Axelrod [1980] Competition(s)

⇒ How should one play a *repeated* Prisoner's Dilemma?

Some sample strategies:

- always defect
  - always cooperate
  - equiprobable randomization
  - match what the opponent plays, e.g., on average
  - grim trigger (e.g., cooperate until defected)
  - “tit-for-tat” (i.e., start by cooperating, then copy what the opponent does)
- } unconditional; perhaps, not the smartest

### First Tournament:

- 14 strategies from leading scientists + RANDOM
- each strategy plays against every other (and itself) for 200 rounds × 5 times

### Second Tournament:

- 62 strategies + RANDOM
- infinitely repeated play (~200 rounds × 5 times)

### Ecological Tournament:

- second tournament in an “evolutionary” setting
- all strategies equally represented with the more successful replacing the less successful over 1000 generations [Fig. 1, p. 400]

⇒ TIT-FOR-TAT is the best throughout (Anatol Rapoport) [Table 2. p. 384]

- \* nice
  - \* provokable/retaliating
  - \* forgiving
- } properties of all successful strategies

(!) *not* a “solution” to Prisoner's Dilemma, though

- \* would only come in 4<sup>th</sup> if played the top 50% strategies only
- \* cannot detect RANDOM
- \* won't exploit when given the opportunity
- \* there is no best rule independent of the environment (i.e., the distribution of opponent strategies)

## How Do Human Subjects Play Prisoner's Dilemma?

⇒ Two major perspectives:

### Reputation Building [Kreps et al., 1982]<sup>3</sup>

- (some) players have the *belief* that their opponent is *not* rational but rather is playing some conditionally cooperative strategy (e.g., tit-for-tat)
- cooperation then is more beneficial as that probability  $\rightarrow 1$

⇒ *selfish* players will cooperate in early rounds!

⇒ defection is still dominant in the last round as well as in one-shot games!

### Altruism Theories

- (some) players are not strictly selfish but benefit from cooperation in a manner *not* reflected in the payoff matrix

a) pure altruism:

$$u_i = \pi_i + \delta \cdot \pi_j \text{ s.t. } \delta > 0,$$

where  $u_i$  is own utility, and  $\pi_i$  and  $\pi_j$  are own and opponent's payoffs, respectively

b) duty/"warm glow":

$$u_i = \pi_i + \delta \text{ s.t. } \delta > 0 \text{ if one chooses to cooperate, and } 0 \text{ otherwise}$$

c) reciprocal altruism:

$$u_i = \pi_i + \delta \text{ s.t. } \delta > 0 \text{ if both players choose to cooperate, and } 0 \text{ otherwise}$$

⇒ a) and b) can support cooperation even in one-shot games by making cooperation either a *best response* or *dominant* strategy

exa)

	C	D
C	$a + \delta, a + \delta$	$d + \delta, c$
D	$c, d + \delta$	$b, b$

\* both players can exhibit "warm glow" altruistic behavior potentially

\* based on the actual value of  $\delta$ , cooperation can be:

— *dominated* strategy  $\Leftrightarrow \delta < \min(b - d, c - a)$

— *best response* strategy  $\Leftrightarrow \min(b - d, c - a) < \delta < \max(b - d, c - a)$

— *dominant* strategy  $\Leftrightarrow \delta > \max(b - d, c - a)$

---

<sup>3</sup>David M Kreps, Paul Milgrom, John Roberts, and Robert Wilson. Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2):245 – 252, 1982

## Cooper et al. [1996]: Reputation Versus Altruism

⇒ reputation building versus “warm glow” altruism

- one-shot treatment (OST): 20 rounds; perfect stranger matching; 40 subjects
- repeated treatment (RT):  $2 \times 10$  periods; partner matching; 30 subjects
- between-subject design

(!) last 10 rounds from OST; 10 rounds of practice (one-shot) in RT

[Table 1 and Fig. 1, p. 199]

⇒ cooperation rates are positive and generally declining over time in both treatments

⇒ cooperation rates in RT are higher than in OST

⇒ neither theory can describe all of the data

[Fig. 2, p. 201]

⇒ most of cooperative play in OST comes from the subjects who do not cooperate all the time → best response altruism

⇒ 12.5~15% are altruists (i.e., cooperate more than 50% of the time); 62.5~85% are selfish

[Fig. 3, p. 205]

⇒ actual cooperation rates in RT follow a concave pattern while reputation building predicts a convex one and altruism predicts a constant level after the initial drop from period one

⇒ only 25% of subjects behave in accordance with reputation building on the individual level (e.g., defection in the last period, no cooperation following defection)

## Suggested Literature

- Charles A Holt. *Markets, games, & strategic behavior*. Boston Pearson Addison Wesley, 2007 [Chapters 3.1–3.2]
- Robert Axelrod. More effective choice in the prisoner’s dilemma. *The Journal of Conflict Resolution*, 24(3):379–403, 1980
- Russell Cooper, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. Cooperation without reputation: Experimental evidence from prisoner’s dilemma games. *Games and Economic Behavior*, 12(2):187 – 218, 1996
- \* James Andreoni and John H. Miller. Rational cooperation in the finitely repeated prisoner’s dilemma: Experimental evidence. *The Economic Journal*, 103(418):570–585, 1993