Backward Induction in Finitely Repeated Prisoner’s Dilemma: Experimental Evidence

Matthew Embrey  Guillaume Fréchette  Sevgi Yuksel

Maastricht  NYU  NYU

Zurich 2013
Motivation

- The prisoner’s dilemma is a canonical game in social sciences.
- We know little about how people behave in finitely repeated PDs.
- Results are inconclusive on whether or not subjects learn backward induction:
  - some studies find cooperation to decline with experience,
  - others find that first defection occurs later on in the interaction.
- Folk wisdom: As the horizon of the repeated game increases cooperation rates increase.
  - Backward induction is more difficult as there are more steps of reasoning required.
**Cooperation decreases with experience**

**Dal Bo (2005)**

PD1, Horizon = 4
COOPERATION INCREASES WITH EXPERIENCE

Andreoni & Miller (1993)
Horizon = 10

Mean cooperation

Mean time to first defection

Andreoni & Miller (1993)
Prior Work

- Studies differ in the parameters of the game: number of matches, the horizon, stage game payoffs; but rarely within studies.
- Data is analyzed differently
  - Round 1 behavior
  - Last round behavior
  - Mean cooperation rate
  - Mean time to first defection
WHAT WE DO

- Meta-analysis of the prior experimental research.
- Conduct a new experiment to understand how the ability to perform backward induction varies with the environment.
  - Document how this varies with the parameters.
**OUR CONTRIBUTIONS**

Statistical:

- Measures used in the literature (mean cooperation rate, mean time to first defection) can be misleading.
Our Contributions

Statistical:

- Measures used in the literature (mean cooperation rate, mean time to first defection) can be misleading.

Design:

- Establish that having many matches is important.
- Conduct the first study varying payoffs and horizon in a between subject design.
Our Contributions

Statistical:
- Measures used in the literature (mean cooperation rate, mean time to first defection) can be misleading.

Design:
- Establish that having many matches is important.
- Conduct the first study varying payoffs and horizon in a between subject design.

Conceptual:
- Propose that the impact of the horizon confounds other factors.
- Horizon affects play:
  - not because it requires more steps of backward induction,
  - but because it increases the value of cooperation which can be captured by the basin of attraction.
**META: PAPERS AND PARAMETERS**

1. Selten and Stoecker (1986)
   - $H = 10$, 25 matches

2. Andreoni and Miller (1993)
   - $H = 10$, 20 matches, 1 session

   - $H = 10$, 2 matches, 3 sessions

   - $H = 10$, 10 ($\times 6$) matches

   - $H = \{2, 4\}$, $\{5,8,9,10\}$ matches, 4 sessions, within design

   - $H = 10$, 20 matches, 4 sessions

7. Friedman and Oprea (2012)
   - $H = 8$, 8 matches, 3 sessions, within design
## Meta: Stage Game Parameters

<table>
<thead>
<tr>
<th>Original</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>D</td>
<td>T</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\frac{R-P}{R-P} = 1)</td>
<td>(\frac{S-P}{R-P} = -\ell)</td>
</tr>
<tr>
<td>(\frac{T-P}{R-P} = 1 + g)</td>
<td>(\frac{P-P}{R-P} = 0)</td>
</tr>
</tbody>
</table>

\(T > R > P > S\)
Meta: Stage Game Parameters

Finitely repeated PD experiments
Normalized payoffs

DB2005  AM1993
CDFR1996  FO2012
BMR2006  SS1986
HN2001

Normalized payoffs
**Meta: Stage Game Parameters**

**Between Subjects Designs Only**

Finitely repeated PD experiments

Normalized payoffs

- AM1993
- CDFR1996
- BMR2006
- SS1986
- HN2001
META: STANDARD ANALYSIS

- Payoffs matter: high $g$ and high $\ell$ decrease cooperation.
- With experience: last round cooperation rate close to zero.
- Average and round 1 cooperation, round to first defection:
  - when $H$ is small, they decrease;
  - when $H$ is large, they increase.
## Meta: Standard Analysis

<table>
<thead>
<tr>
<th>Experiment</th>
<th>H</th>
<th>g</th>
<th>ℓ</th>
<th>Average</th>
<th>Cooperation rate (%)</th>
<th>Mean round to first def.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>DB2005</td>
<td>2</td>
<td>1.17</td>
<td>0.83</td>
<td>14</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>within sub.</td>
<td>2</td>
<td>0.83</td>
<td>1.17</td>
<td>25</td>
<td>32</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.17</td>
<td>0.83</td>
<td>33</td>
<td>44</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.83</td>
<td>1.17</td>
<td>31</td>
<td>37</td>
<td>22</td>
</tr>
<tr>
<td>FO212</td>
<td>8</td>
<td>4.00</td>
<td>4.00</td>
<td>33</td>
<td>43</td>
<td>33</td>
</tr>
<tr>
<td>within sub.</td>
<td>8</td>
<td>2.00</td>
<td>4.00</td>
<td>38</td>
<td>43</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.33</td>
<td>0.67</td>
<td>40</td>
<td>43</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.67</td>
<td>0.67</td>
<td>44</td>
<td>50</td>
<td>69</td>
</tr>
<tr>
<td>BMR2006</td>
<td>10</td>
<td>2.33</td>
<td>2.33</td>
<td>38</td>
<td>61</td>
<td>66</td>
</tr>
<tr>
<td>AM1993</td>
<td>10</td>
<td>1.67</td>
<td>1.33</td>
<td>17</td>
<td>36</td>
<td>47</td>
</tr>
<tr>
<td>CDFR1996</td>
<td>10</td>
<td>0.44</td>
<td>0.78</td>
<td>52</td>
<td>60</td>
<td>57</td>
</tr>
</tbody>
</table>

First defection is set to Horizon + 1 if there are no defection.

Table: Cooperation rates and mean round to first defection
### Meta: Standard Analysis

<table>
<thead>
<tr>
<th>Experiment</th>
<th>H</th>
<th>g</th>
<th>l</th>
<th>Average</th>
<th>Cooperation rate (%)</th>
<th>Mean round to first def.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>Round 1</td>
<td>Last Round</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB2005</td>
<td>2</td>
<td>1.17</td>
<td>0.83</td>
<td>14</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>within sub.</td>
<td>2</td>
<td>0.83</td>
<td>1.17</td>
<td>25</td>
<td>32</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.17</td>
<td>0.83</td>
<td>33</td>
<td>44</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.83</td>
<td>1.17</td>
<td>31</td>
<td>37</td>
<td>20</td>
</tr>
<tr>
<td>FO212</td>
<td>8</td>
<td>4.00</td>
<td>4.00</td>
<td>33</td>
<td>43</td>
<td>23</td>
</tr>
<tr>
<td>within sub.</td>
<td>8</td>
<td>2.00</td>
<td>4.00</td>
<td>38</td>
<td>43</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.33</td>
<td>0.67</td>
<td>40</td>
<td>43</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.67</td>
<td>0.67</td>
<td>44</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>BMR2006</td>
<td>10</td>
<td>2.33</td>
<td>2.33</td>
<td>38</td>
<td>61</td>
<td>22</td>
</tr>
<tr>
<td>AM1993</td>
<td>10</td>
<td>1.67</td>
<td>1.33</td>
<td>17</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>CDFR1996</td>
<td>10</td>
<td>0.44</td>
<td>0.78</td>
<td>52</td>
<td>60</td>
<td>20</td>
</tr>
</tbody>
</table>

First defection is set to Horizon + 1 if there are no defection.

Table: Cooperation rates and mean round to first defection
**META: STANDARD ANALYSIS**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>H</th>
<th>g</th>
<th>ℓ</th>
<th>Cooperation rate (%)</th>
<th>Mean round to first def.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Average</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>DB2005</td>
<td>2</td>
<td>1.17</td>
<td>0.83</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>within sub.</td>
<td>2</td>
<td>0.83</td>
<td>1.17</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.17</td>
<td>0.83</td>
<td>33</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.83</td>
<td>1.17</td>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>FO212</td>
<td>8</td>
<td>4.00</td>
<td>4.00</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>within sub.</td>
<td>8</td>
<td>2.00</td>
<td>4.00</td>
<td>38</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.33</td>
<td>0.67</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.67</td>
<td>0.67</td>
<td>44</td>
<td>69</td>
</tr>
<tr>
<td>BMR2006</td>
<td>10</td>
<td>2.33</td>
<td>2.33</td>
<td>38</td>
<td>66</td>
</tr>
<tr>
<td>AM1993</td>
<td>10</td>
<td>1.67</td>
<td>1.33</td>
<td>17</td>
<td>47</td>
</tr>
<tr>
<td>CDFR1996</td>
<td>10</td>
<td>0.44</td>
<td>0.78</td>
<td>52</td>
<td>57</td>
</tr>
</tbody>
</table>

First defection is set to Horizon + 1 if there are no defection.

Table : Cooperation rates and mean round to first defection
## Meta: Standard Analysis

<table>
<thead>
<tr>
<th>Experiment</th>
<th>H</th>
<th>g</th>
<th>( \ell )</th>
<th>Average</th>
<th>Cooperate rate (%)</th>
<th>Mean round to first def.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>L</td>
<td>1</td>
</tr>
<tr>
<td>DB2005</td>
<td>2</td>
<td>1.17</td>
<td>0.83</td>
<td>14</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>within sub.</td>
<td>2</td>
<td>0.83</td>
<td>1.17</td>
<td>25</td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.17</td>
<td>0.83</td>
<td>33</td>
<td>20</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.83</td>
<td>1.17</td>
<td>31</td>
<td>22</td>
<td>37</td>
</tr>
<tr>
<td>FO212</td>
<td>8</td>
<td>4.00</td>
<td>4.00</td>
<td>33</td>
<td>33</td>
<td>43</td>
</tr>
<tr>
<td>within sub.</td>
<td>8</td>
<td>2.00</td>
<td>4.00</td>
<td>38</td>
<td>34</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1.33</td>
<td>0.67</td>
<td>40</td>
<td>48</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.67</td>
<td>0.67</td>
<td>44</td>
<td>69</td>
<td>50</td>
</tr>
<tr>
<td>BMR2006</td>
<td>10</td>
<td>2.33</td>
<td>2.33</td>
<td>38</td>
<td>66</td>
<td>61</td>
</tr>
<tr>
<td>AM1993</td>
<td>10</td>
<td>1.67</td>
<td>1.33</td>
<td>17</td>
<td>47</td>
<td>36</td>
</tr>
<tr>
<td>CDFR1996</td>
<td>10</td>
<td>0.44</td>
<td>0.78</td>
<td>52</td>
<td>57</td>
<td>60</td>
</tr>
</tbody>
</table>

First defection is set to Horizon + 1 if there are no defection.

### Table: Cooperation rates and mean round to first defection
**VALUE OF COOPERATION VS. DEFECTION**

**Basins of Attraction**

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1, 1</td>
<td>$-\ell, 1+g$</td>
</tr>
<tr>
<td>D</td>
<td>$1+g, -\ell$</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

Normalized PD
# Value of Cooperation vs. Defection

## Basins of Attraction

<table>
<thead>
<tr>
<th>Normalized PD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C</strong></td>
</tr>
<tr>
<td><strong>D</strong></td>
</tr>
<tr>
<td><strong>C</strong></td>
</tr>
<tr>
<td><strong>D</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reduced Game</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grim</strong></td>
</tr>
<tr>
<td><strong>AD</strong></td>
</tr>
<tr>
<td><strong>Grim</strong></td>
</tr>
<tr>
<td><strong>AD</strong></td>
</tr>
</tbody>
</table>
VALUE OF COOPERATION VS. DEFECTION

Basins of Attraction

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1, 1</td>
<td>-(\ell), 1 + g</td>
</tr>
<tr>
<td>D</td>
<td>1 + g, -(\ell)</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

Reduced Game

<table>
<thead>
<tr>
<th></th>
<th>Grim</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grim</td>
<td>(H, H)</td>
<td>-(\ell), 1 + g</td>
</tr>
<tr>
<td>AD</td>
<td>1 + g, -(\ell)</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

\[\text{sizeBAD} = \frac{\ell}{(H-1)+\ell-g}\]
### META: DETERMINANTS OF COOPERATION

Table: Probit regression (Marginal Effects): round 1.

<table>
<thead>
<tr>
<th></th>
<th>Column (1)</th>
<th>Column (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>basin of att.</td>
<td>-0.063*** (0.020)</td>
<td>-0.359*** (0.093)</td>
</tr>
<tr>
<td>g</td>
<td>-0.006 (0.023)</td>
<td>-0.034* (0.018)</td>
</tr>
<tr>
<td>l</td>
<td>0.034*** (0.009)</td>
<td>0.003 (0.014)</td>
</tr>
<tr>
<td>horizon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td>5398</td>
<td>5398</td>
</tr>
</tbody>
</table>

Clustered (session level) standard errors in parentheses.

***1%, **5%, *10% significance.

Note: Regressors not reported here include cooperation in round 1 of first match, opponent cooperation in round 1 of previous match, and number of matches played.
DESIGN

2 × 2 factorial design (between subjects)

Summary of treatments

<table>
<thead>
<tr>
<th>PD/Horizon</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>E4</td>
<td>E8</td>
</tr>
<tr>
<td>Difficult</td>
<td>D4</td>
<td>D8</td>
</tr>
</tbody>
</table>

- Horizon: 4 or 8 rounds
- Payoffs of the PD: Easy vs. Difficult
- In each session subjects played 20-30 finitely repeated PDs

Easy PD

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>51, 51</td>
<td>22, 63</td>
</tr>
<tr>
<td>D</td>
<td>63, 22</td>
<td>39, 39</td>
</tr>
</tbody>
</table>

Difficult PD

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>51, 51</td>
<td>5, 81</td>
</tr>
<tr>
<td>D</td>
<td>81, 5</td>
<td>39, 39</td>
</tr>
</tbody>
</table>
DESIGN

BASIN OF ATTRACTION OF GRIM

Potential Gains to Cooperation

Baseline of Attraction

FIRST DEFECTION

Mean time to first defection

![Graph showing mean time to first defection for different rounds and matches.](chart.png)
COOPERATION BY ROUND

Cooperation Rate
First round

Match

E4

D4

D8

E8

Cooperation Rate

First round

Match

22 / 26
UNRAVELLING OF COOPERATION

Mean cooperation rate by round
Treatment: E8
BREAKDOWN OF COOPERATION

Breakdown of cooperation

Round

Probability

D8

E8

Matches 1–10
Matches 11–20
Matches 21–30

Matches 1–10
Matches 11–20
Matches 21–30
FIRST DEFECTION AND LAST COOPERATION

Evolution of first defection
Treatment: E8

- Last cooperation
- First defection
CONCLUSION

- Parameters of the game have a significant impact on round 1 cooperation rates.
  - The majority of this effect can be captured by the basin of attraction.
- Despite differences in round 1 behavior, subjects behavior is consistent with the logic of backward induction.