Social preferences, positive expectations, and trust based cooperation

Ryan O. Murphy and Kurt Ackermann

Chair of Decision Theory and Behavioral Game Theory
ETH Zürich

ICSD June 24th, 2015
Overview

- Started with a discussion and a question:
  - How is SVO related to trust?
Overview

• Started with a discussion and a question:
  • How is SVO related to trust?
  • SVO is a necessary, but not sufficient, precursor to trust based cooperation.
Overview

I. PD game in a generalized sense
II. SVO- Social preferences
III. Beliefs- Positive expectations
IV. Trust based cooperation
### PD game

<table>
<thead>
<tr>
<th></th>
<th>Player 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C</strong></td>
<td><strong>R, R</strong></td>
<td><strong>S, T</strong></td>
</tr>
<tr>
<td><strong>D</strong></td>
<td><strong>T, S</strong></td>
<td><strong>P, P</strong></td>
</tr>
</tbody>
</table>

- \( T > R > P > S \)
- \( 2R > (T + S) \)

Defining the interaction
$T > R > P > S$

$2R > (T + S)$

Defining the interaction
But many people choose to cooperate—Even if the game is one shot, fully anonymous, and incentive compatible interaction. Why?

Why would someone choose to cooperate?

\[ T > R > P > S \]

\[ 2R > (T + S) \]
Why would someone choose to cooperate?

Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another.

Rousseau et al., 1998

\[ T > R > P > S \]

\[ 2R > (T + S) \]
<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>R, R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S, T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T, S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P, P</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Player 1:
- C
- D

Player 2:
- R, R
- S, T
- T, S
- P, P

$T > R > P > S$

$2R > (T + S)$

Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another.

Rousseau et al., 1998

...with the intention of improving collective outcomes.
Generalizing the PD game

\[
\begin{array}{c|c|c|c}
\toprule
 & \text{C} & \text{D} \\
\hline
\text{C} & (R, R) & (S, T) \\
\hline
\text{D} & (T, S) & (P, P) \\
\bottomrule
\end{array}
\]

\[T > R > P > S\]
\[2R > (T + S)\]
\[T = 1 \quad S = 0\]
\[
\begin{array}{c|c}
\text{Player 2} & C & D \\
\hline
C & (R, R) & (S, T) \\
D & (T, S) & (P, P) \\
\end{array}
\]

- \( T > R > P > S \)
- \( 2R > (T + S) \)
- \( T = 1 \quad S = 0 \)
- \( R, P \in (0, 1) \)

Generalizing the PD game
Player 1

Player 2

C    D

C    R, R    S, T

D    T, S    P, P

$T > R > P > S$

$2R > (T + S)$

$T = 1$    $S = 0$

$R, P \in (0, 1)$
<table>
<thead>
<tr>
<th>PD game</th>
<th>T</th>
<th>R</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.6</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.7</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.8</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.9</td>
<td>0.8</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.6</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.7</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.8</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0.9</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0.6</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0.7</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0.8</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0.9</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>0.6</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0.7</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0.8</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>0.9</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0.6</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>0.7</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>0.8</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0.9</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>0.7</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>0.8</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>0.9</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>0.8</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>0.9</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>0.9</td>
<td>0.1</td>
<td>0</td>
</tr>
</tbody>
</table>
$K = \frac{(R - P)}{(T - S)}$
\[ K = \frac{(R - P)}{(T - S)} \]
\[ K = \frac{(R - P)}{(T - S)} \]
Social preferences

\[ u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o \]

$u(\pi_S, \pi_O) = \pi_S + \alpha \cdot \pi_O$
Social value orientation results
Beliefs

• How likely does a DM believe it is that the other player will choose to cooperate.
  • Zero if the DM is certain the other will defect
  • One if the DM is certain the other will cooperate
  • In-between values correspond to different degrees of belief about what the other player will choose

\[ \beta \in [0, 1] \]
Integration

- PD game in a generalized sense
- SVO- Social preferences
- Beliefs- Positive expectations
- Trust based cooperation

Bring this all together in a coherent and well integrated way

Build a framework that can rationalize trust based cooperation given heterogeneity in peoples’ preferences and beliefs
Positive expectation (Beta)

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>.8, .8</td>
</tr>
<tr>
<td>D</td>
<td>1, 0</td>
</tr>
</tbody>
</table>

T = 1      R = 0.8      P = 0.6      S = 0

Defect

Cooperate

1, 0

.6, .6

C

D

Player 1

Player 2

.8, .8

0, 1

Prosocial

Individualistic

SVO (Alpha)

Homo economicus

0º

45º
<table>
<thead>
<tr>
<th></th>
<th>Player 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.8, .8</td>
<td>0, 1</td>
</tr>
<tr>
<td>D</td>
<td>1, 0</td>
<td>.6, .6</td>
</tr>
</tbody>
</table>

- **Positive expectation (Beta)**
- **SVO (Alpha)**
- **Defect**
- **Cooperate**

Another DM

- **Prosocial**
- **Individualistic**
- **Pessimistic**
- **Optimistic**

- **Homo economicus**
The figure presents a payoff matrix and a corresponding graph. The matrix shows the payoffs for two players, Player 1 and Player 2, with strategies C (Cooperate) and D (Defect). The payoffs are as follows:

- Player 1:
  - C (Cooperate) for Player 2:
    - C: 0.8, 0.8
    - D: 1, 0

- Player 2:
  - C (Cooperate) for Player 1:
    - C: 0.8, 0.8
    - D: 1, 0.6

The graph illustrates the relationship between the SVO (Alpha) and Positive expectation (Beta) with T = 1, R = 0.8, P = 0.6, and S = 0. The graph's shading indicates regions labeled as 'Cooperate' and 'Defect.' The blue dot represents a cooperative strategy, while the red dot represents a defective strategy.
\[ u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o \]

\[ u(C) = [\beta \cdot (R + \alpha \cdot R)] + [(1 - \beta) \cdot (S + \alpha \cdot T)] \]

\[ u(D) = [\beta \cdot (T + \alpha \cdot S)] + [(1 - \beta) \cdot (P + \alpha \cdot P)] \]
\[ u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o \]

\[ u(C) = [\beta \cdot (R + \alpha \cdot R)] + [(1 - \beta) \cdot (S + \alpha \cdot T)] \]

\[ u(D) = [\beta \cdot (T + \alpha \cdot S)] + [(1 - \beta) \cdot (P + \alpha \cdot P)] \]
\[
\begin{align*}
\text{Social preferences} & \\
\text{Preferences and beliefs} & \\

u(\pi_S, \pi_o) &= \pi_S + \alpha \cdot \pi_o \\

u(C) &= [\beta \cdot (R + \alpha \cdot R)] + [(1 - \beta) \cdot (S + \alpha \cdot T)] \\

u(D) &= [\beta \cdot (T + \alpha \cdot S)] + [(1 - \beta) \cdot (P + \alpha \cdot P)] \\

\Rightarrow \quad \beta_{\text{crit}} &= \frac{P - S + \alpha P - \alpha T}{P + R - S - T + \alpha P + \alpha R - \alpha S - \alpha T}
\end{align*}
\]
\[ u(\pi_s, \pi_o) = \pi_s + \alpha \cdot \pi_o \]

\[ u(C) = [\beta \cdot (R + \alpha \cdot R)] + [(1 - \beta) \cdot (S + \alpha \cdot T)] \]

\[ u(D) = [\beta \cdot (T + \alpha \cdot S)] + [(1 - \beta) \cdot (P + \alpha \cdot P)] \]

\[ \beta_{\text{crit}} = \frac{P - S + \alpha P - \alpha T}{P + R - S - T + \alpha P + \alpha R - \alpha S - \alpha T} \]
Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another. 

Rousseau et al., 1998

...with the intention of improving collective outcomes.

Expectations alone are insufficient to justify trust based cooperation. Social preferences are necessary also.
Normalized PD game space

K-index

P

R
Mechanisms for nudging people to act more cooperatively

Interventions to change people’s preferences and/or mechanisms to change their beliefs

Conclusions

• This started with a question:
  • How is SVO related to trust?
  • SVO is a necessary, but not sufficient, precursor to trust based cooperation.
  • If a DM is sufficiently prosocial, and has sufficient expectations that the other player will cooperate, then we expect trust based cooperation to emerge, provided the dilemma is not too severe.
Conclusions

• Develop a psychologically grounded model of trust based cooperation that is consistent with the principles of rationality and is measurable
• Integrate SVO (preferences), expectations (beliefs), trust based cooperation among interdependent players
• Derive precise trust thresholds over combinations of social preferences and beliefs
• Rapoport's K-index is isomorphic to the minimum level of SVO required by a DM to justify cooperation given a uniform prior (beliefs about the other player)
• Different PD games can have radically different properties


• See also [http://vlab.ethz.ch/svo](http://vlab.ethz.ch/svo) for code and details about the measures and models.
Normalizing the payoffs to be between 0 and 1

\[
\begin{align*}
T_n &= \frac{T - S}{T - S} = 1, \\
R_n &= \frac{R - S}{T - S}, \\
P_n &= \frac{P - S}{T - S}, \\
S_n &= \frac{S - S}{T - S} = 0
\end{align*}
\]
• Rapoport's K-index is isomorphic of the minimum level of SVO required by a DM to justify cooperation given a uniform prior (beliefs about the other player)

• For any PD game, assuming a uniform prior, the critical level of SVO can be found with:

\[ \alpha_{crit} = - \frac{P - R - S + T}{P - R + S - T} \]
Player 1

Player 2

C  D

C  R, R  S, T

D  T, S  P, P

$T > R > P > S$

$2R > (T + S)$

Trust

Player:

1  2

Terminal Node:

Payoff to Player A:
P=5  S=0

Payoff to Player B:
P=5  T=15
\[ \pi_s + \alpha \cdot \pi_o \]

Simple joint utility function

\[ \pi_s + 1 \cdot \alpha \pi_o \]

Contingent function

\[ (\pi_s^{1-\alpha}) \cdot (\pi_o^\alpha) \]

Cobb-Douglas function