Cooperation in Social Dilemmas

Christoph Hauert
Department of Mathematics, UBC Vancouver, Canada

Sanctioning institutions for governing the commons

www.evoludo.org
Public goods games

Games and population dynamics

- Large population, players interact in randomly formed groups of size \( N \).
- Two strategic types
  - cooperators \( x \) - contribute to common pool at cost \( c \).
  - defectors \( y \) - contribute nothing
- Total contributions are multiplied by \( r > 1 \) and equally split among all other participants (irrespective of their type):
  \[
  P_{y} = \frac{r c}{N - 1} x (N - 1) = r c \ x \\
  P_{x} = P_{y} - c = (r x - 1) c
  \]
- Payoffs translate into reproductive fitness.
  \[
  \dot{x} = x (P_{x} - \bar{P}) \\
  = x (1 - x) (P_{x} - P_{y})
  \]
- Cooperators go extinct.
Punishment

Promoting cooperation - part I

- Punishment is costly - punisher pays $\gamma$, punishment fine $\beta$.
- Three strategic types
  - cooperators $x$ - contribute to public goods, do not punish
  - defectors $y$ - do not contribute, do not punish
  - peer punishers $w$ - contribute and punish those that did not

- Payoffs
  
  \[
  P_x = B - c \\
  P_y = B - (N - 1)w \beta \\
  P_w = B - c - (N - 1)y \gamma
  \]

  \[
  B = rc(x + w) \quad \text{benefits from public good}
  \]
Punishment

Effects of punishment

- Defection is the only stable state.
- No selection in populations of cooperators and peer punishers (line of fixed points).
- Cooperators pave the way for defectors.
- How can punishment gain a foothold in the population?

*Sigmund, Hauert & Nowak (2001) PNAS 98 10757.*
Volunteering

Promoting cooperation - part II

- Participation in public goods interactions is voluntary.
- Joint effort is risky - potential for high costs and large benefits.
- Risk averse individuals obtain small but fixed payoff $\sigma$
  
  $$0 < \sigma < (r - 1)c,$$
  better than mutual defection but worse than mutual cooperation).
- Three strategic types
  - cooperators $x$ - contribute to public goods
  - defectors $y$ - do not contribute
  - loners $z$ - refuse to participate
- Single participant receives $\sigma$. 
Volunteering

Theory

- Payoffs

\[
\begin{align*}
 P_y &= \sigma z^{N-1} + (1 - z^{N-1}) r c \frac{x}{1-z} \\
 P_x &= P_y - (1 - z^{N-1}) c \\
 P_z &= \sigma 
\end{align*}
\]

- Rock-Scissors-Paper type cyclic dominance along boundary of \(S_3\).

- Loners provide an escape hatch out of states of mutual defection - but this is a fleeting state.

Hauert, De Monte, Hofbauer & Sigmund (2002)
Science 296 1129.
Volunteering & Punishment

Promoting cooperation - part III

- Four strategic types
  - cooperators $x$: contribute to public goods, do not punish
  - defectors $y$: participate but do not contribute, do not punish
  - loners $z$: do not participate
  - peer punishers $w$: contribute and punish

- Allow for second order punishment - punish those that failed to punish ($\alpha$ controls strength).

- Payoffs

  \[
  P_x' = \alpha \beta w (N - 1) (1 - (1 - y)^{N-2})
  \]

  \[
  P_y' = \beta w (N - 1)
  \]

  \[
  P_z = \sigma
  \]

  \[
  P_w' = \alpha \gamma x (N - 1) (1 - (1 - y)^{N-2})
  \]

  \[
  \gamma y (N - 1)
  \]

Voluntary public goods

Punishment
Volunteering & Punishment

Population dynamics

- Replicator dynamics exhibits two basins of attraction:
  - neutral mixtures of punishers and cooperators (line of fixed points).
  - loners only.

- Fails to explain the evolution of punishment.

- Second order punishment barely affects the dynamics.

- Degenerate dynamics - long term outcome unclear.

- Stochastic model.

*Brandt, Hauert, Sigmund (2006) PNAS 103 495.*
Finite Populations

Genetic reproduction or social imitation

- Interaction:
  - Random sampling of interaction group without replacement.

- Evolution:
  - Randomly choose focal individual $i$.
  - Randomly choose model individual $j$.
  - Focal individual adopts strategy of model with probability proportional to payoff difference $P_j - P_i$:
    \[
    \frac{1}{1 + \exp[-s(P_j - P_i)]}
    \]
  - $s \geq 0$: strength of selection
  - $s \to 0$: random selection
  - $s \to +\infty$: deterministic selection

\[s \geq 0\] strength of selection
\[s \to 0\] random selection
\[s \to +\infty\] deterministic selection

Saturday, November 6, 2010
Finite populations

Stochastic dynamics

- Rare mutations $\mu$:
  - population is homogeneous most of the time.
  - occasionally a single mutant strategy occurs.
  - mutant disappears or takes over the entire population before next mutation occurs.

- Stochastic dynamics along edges of simplex $S_n$.

- Probability that type $i$ increases competing against type $j$, $T_{ij}^+$:
  \[
  T_{ij}^+ = \frac{X_i M - X_i}{M} \frac{1}{1 + \exp \left[ -s(P_i - P_j) \right]}
  \]
  $M$: population size, $X_i$: number of type $i$ individuals, $M-X_i$: number of $j$ types.

- Fixation probability of single $i$ mutant in $j$ population.
  \[
  \rho_{ij} = \frac{1}{\sum_{k=0}^{M-1} \prod_{X_i=1}^k T_{ij}^-} = \frac{1}{\sum_{k=0}^{M-1} \exp \left[ s \sum_{X_i=1}^k (P_j - P_i) \right]}
  \]

- Embedded Markov chain for transitions between homogenous states.
Finite populations

**Strong imitation $s \to \infty$**

- Strong imitation significantly simplifies Markov chain.
  
  Ex. voluntary public goods games: $\rho_{XY} = \rho_{YZ} = 1$ and $\rho_{ZX} = 1/2$ all other $\rho_{ij} = 0$.

\[
\begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{4} & 0 & \frac{3}{4}
\end{pmatrix}
\to
\begin{pmatrix}
\frac{1}{4} \\
\frac{1}{4} \\
\frac{1}{2}
\end{pmatrix}
\]

- The result becomes independent of the parameters! (as long as the cyclic dominance $X \to Y \to Z \to X$ persists).

- $\rho_{ZX} = 1/2$ because two cooperators are required to invade a loner population (non-hyperbolic fixed point in replicator equation).

- Neutral evolution (no fitness differences) yields fixation probability of $1/M$ where $M$ denotes the population size.
Volunteering & Punishment

Results

Compulsory public goods games with peer punishment: defectors rule.

Voluntary public goods games: cyclic dominance.

Voluntary public goods games with peer punishment: punishers reign ($M=92$).

In cultural evolution ‘mutation’ rates may not be small - individuals randomly experiment with different strategies.

For smaller $\mu$, punishment prevails.

The stochastic dynamics in finite populations can resolve the problem of establishing altruistic punishment.

Cooperators prevail for large $\mu$.
(Note that here the contributors also get a return on their own investment - otherwise loners dominate.)

Loners are no longer crucial.

Punishers are pivotal for the success of mild cooperators (second order free riders).

Sanctioning Institutions

Promoting cooperation - part IV

- Second order free riders (contribute but do not punish) undermine punishing efforts.
- Even $\alpha=1$ cannot prevent this because cannot identify among contributors.
- Establish punishment pool where individual contribute before engaging in the public goods interaction.
- Precursor to institutionalized punishment.
- Easy identification of free riders.

- Five strategic types
  - cooperators $x$ - contribute to public goods, do not punish
  - defectors $y$ - participate but do not contribute, do not punish
  - loners $z$ - do not participate
  - peer punishers $w$ - contribute and punish those that did not contribute
  - pool punishers $v$ - contribute to public goods and to punishment pool

- Pool punishers pay an additional amount $G > 0$ into the punishment pool and free riders are fined proportional to the number of pool punishers $N_v: N_v G$. 
Peer versus pool punishment

No second order punishment, $\alpha=0$

- **Payoffs** (infinite populations)
  
  \[
  P_x = P'_x - \alpha G v(N - 1)
  \]
  
  \[
  P_y = P'_y - G(v(N - 1))
  \]
  
  \[
  P_z = \sigma
  \]
  
  \[
  P_w = P'_w - \alpha G v(N - 1)
  \]
  
  \[
  P_v = P'_v - G
  \]

- Punishment pool only used to punish non-contributors (defectors).
  
  - For small $\mu$ peer-punishment dominates.
  
  - For large $\mu$ the public good collapses.
  
  - Pool punishment ineffective.

- For compulsory interactions (no loners) defectors dominate.
Peer versus pool punishment

With second order punishment, \( \alpha=1 \)

- Punishment pool used to punish non-contributors (defectors) as well as those that do not commit to pool-punishment (cooperators and peer-punishers).
- Pool punishers prevail for most \( \mu \).
- For very large \( \mu \) the public good collapses.
- For compulsory interactions (no loners) defectors again dominate.
- Punishment often fails in compulsory public goods.
- Preservation of global resources (climate, air, water, fish...).
- “Mutual coercion mutually [and voluntarily] agreed upon”. Hardin, 1968

Conclusions

Sanctioning Institutions

- Peer punishment relates to the instinct for revenge.
- Pool punishment as a step towards establishing sanctioning institutions: commit resources to prepare for punishing free-riders.
- Pool punishment is based on foresight rather than anger.
- Populations of peer punishers are better off than pool punishers.
- The upkeep of the punishment pool incurs costs.
- With second order punishment, pool punishers prevail.
- Higher efficiency of peer punishment is traded for greater stability of pool punishment.
Acknowledgements

**Karl Sigmund**, University of Vienna, Austria.

**Josef Hofbauer**, University of Vienna, Austria.

**Hannelore De Silva (née Brandt)**, Vienna University of Economics, Austria.

**Martin Nowak**, Program for Evolutionary Dynamics, Harvard University.

**Arne Traulsen**, Max-Planck Institute for Evolutionary Biology, Plön, Germany.

**Silvia De Monte**, École Normale Superieure, Paris, France.

**Tutorials:**
http://www.univie.ac.at/virtuallabs
and soon http://www.evoludo.org

VirtualLabs installation on 3D NOVA display
Location: main train station Zürich, Switzerland
Media artist: Chandrasekhar Ramakrishnan