Iterated Prisoner’s Dilemma contains strategies that dominate any evolutionary opponent

By William H. Press and Freeman J. Dyson

Presented by Damion Terrell, CS523

- Abstract - IPD & Paper hypothesis
- IPD discussion
- Sentient strategies vs Evolved Strategies
- Calculating optimal strategy
- Discussion

Iterated Prisoner's' Dilemma

- Two Suspects detained for a crime,
  - Interrogated in separate rooms.
- Does either Prisoner Defect on the other?
  - Or Cooperate by staying quiet?

Hyp: ∃ Strategy to Dominate

Assumption: No simple ultimatum strategy,
But can X

1. deterministically set Y’s score; and
2. enforce linear relation between X & Y’s score

Deep Dive into the IPD

- Two Player Game
  - Two Options for each player
  - Four possible ways to score

Two Options

- Cooperate with partner
- Defect/Turn on partner

Scoring:

- Turn and your partner cooperates
- Reciprocal cooperation
- Prisoners both turn
- Sucker you, cooperates while your partner turns

Scoring:

- $T \sim 5$
- $R \sim 3$
- $P \sim 1$
- $S \sim 0$

$$T > R > P > S$$

$$2R > T + S$$

Spotting an Evolutionary Player

Y adjusts its strategy, q, by an optimization scheme:

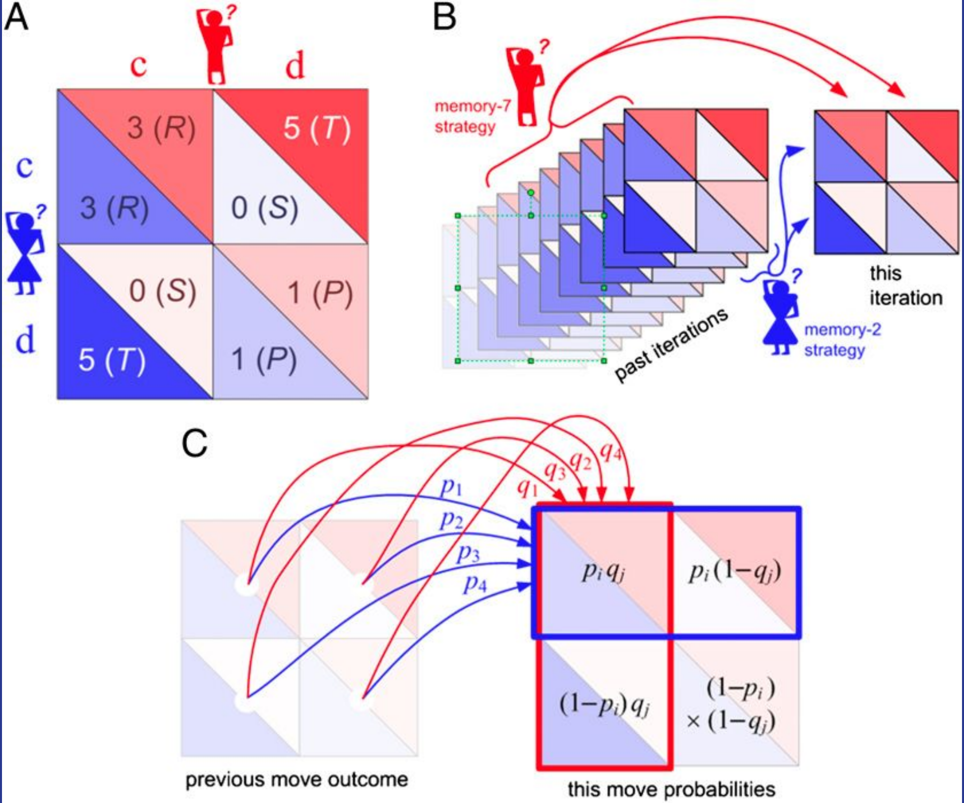
- In order to maximize its own score, s_Y
- Does not explicitly consider opponent's score or strategy.

Spotting the Mindful Player

Y has a theory of mind about X if Y

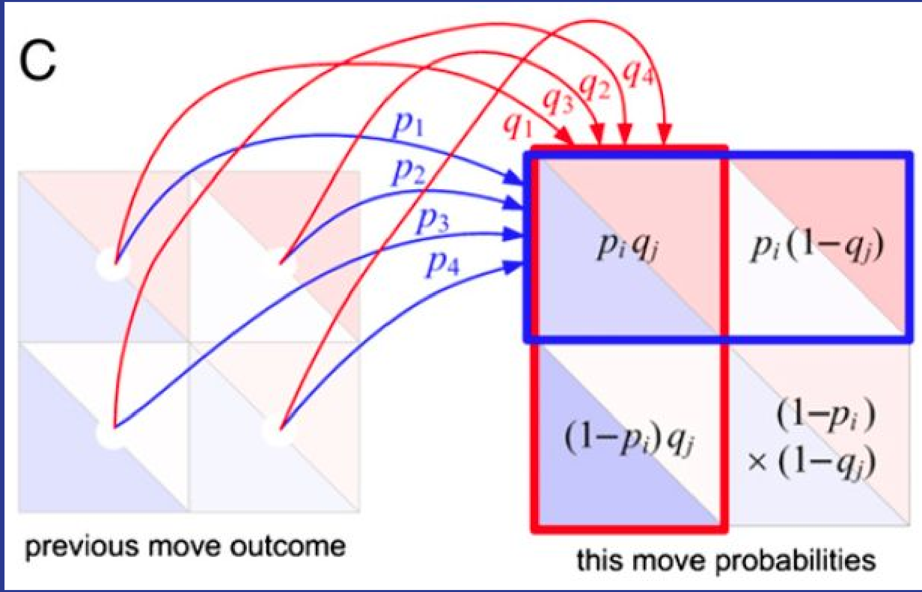
- Imputes to X, an independent strategy; and
- Has ability to alter its response to its opponents actions.

Calculating the Optimal Strategy in IPD:



If options for xy & yx $\in$ (cc,cd,dc,dd); then

- X's strategy is $p = (p_1,p_2,p_3,p_4)$
- Y's strategy is $q = (q_1,q_2,q_3,q_4)$



Zero-Determinant Strategies

i) Markov transition matrix $M(p,q)$
with stationary vector v.

ii) Singular matrix $M' \equiv M - I$ is zero determinant

iii) Stationary vector v (or any proportional)

Zero-Determinant Strategies

$v^T M = v^T$, or $v^T M' = 0$ Eq. 1

Cramer's rule, applied to the matrix M′

$\mathrm{Adj}(M')M' = \det(M')\, I = 0$ Eq. 2

Result is dot product

$v \cdot f \equiv D(p,q,f)$

Zero-Determinant Strategies

A

$$\begin{bmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3) \\ p_3q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2) \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{bmatrix}$$

B

$$\mathbf{v}\cdot\mathbf{f} \equiv D(\mathbf{p},\mathbf{q},\mathbf{f}) = \det\begin{bmatrix} -1+p_1q_1 & -1+p_1 & -1+q_1 & f_1 \\ p_2q_3 & -1+p_2 & q_3 & f_2 \\ p_3q_2 & p_3 & -1+q_2 & f_3 \\ p_4q_4 & p_4 & q_4 & f_4 \end{bmatrix}$$

$\equiv \tilde{\mathbf{p}}$ $\equiv \tilde{\mathbf{q}}$

Zero-Determinant Strategies

Second Column:

$$\tilde{p} \equiv (-1 + p_1, -1 + p_2, p_3, p_4) \quad \text{Eq. 3}$$

Third Column:

$$\tilde{q} \equiv (-1 + q_1, q_3, -1 + q_2, q_4) \quad \text{Eq. 4}$$

Fourth column is simply: f

Payoff Matrix

X score, S_X = (R,S,T,P),
Y score, S_Y = (R,T,S,P)

$$s_X = \frac{\mathbf{v} \cdot \mathbf{S}_X}{\mathbf{v} \cdot \mathbf{1}} = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{S}_X)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}$$

$$s_Y = \frac{\mathbf{v} \cdot \mathbf{S}_Y}{\mathbf{v} \cdot \mathbf{1}} = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{S}_Y)}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}, \quad [5]$$

Linear Mischief and Unilateral Strategies

$$\alpha s_X + \beta s_Y + \gamma = \frac{D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}. \qquad [6]$$

Ways to zero the determinant:

- X chooses $p = \alpha S_X + \beta S_Y + \gamma 1$; or
- Y chooses $q = \alpha S_X + \beta S_Y + \gamma 1$

Zero-determinant (ZD) strategies

Eq. 7

$$\alpha s_X + \beta s_Y + \gamma = 0$$

Not all ZD Strategies are feasible

- Probability p in range [0,1]

X unilaterally sets Y's score

If X sets α = 0 from previous equation

$P = \beta S_Y + \gamma 1$

Eq. 8 for solving p2 & p3 for p1 & p4 in terms R,S,T,P

$$p_2 = \frac{p_1(T-P)-(1+p_4)(T-R)}{R-P}$$
$$p_3 = \frac{(1-p_1)(P-S)+p_4(R-S)}{R-P}. \quad [8]$$

X unilaterally sets Y’s score

Using weights $(1 - p_1)$ with substitution,
Y’s score from Eq. 5 becomes

$$s_Y = \frac{(1-p_1)P + p_4 R}{(1-p_1) + p_4}. \qquad [9]$$

Therefore:

X can only force Y’s score $P \le S_Y \le R$

What if X tries to set its own score?

The analogous calculation with
$\tilde{p} = \alpha S_X + \gamma \mathbf{1}$ yields

$$p_2 = \frac{(1+p_4)(R-S)-p_1(P-S)}{R-P} \geq 1$$
$$p_3 = \frac{-(1-p_1)(T-P)-p_4(T-R)}{R-P} \leq 0. \qquad \textbf{[10]}$$

What if X extorts payoffs larger than mutual noncooperation value of P?

If X chooses strategy ~p

$$\tilde{\mathbf{p}} = \phi[(\mathbf{S}_X - P\mathbf{1}) - \chi(\mathbf{S}_Y - P\mathbf{1})], \quad [11]$$

x ≥1 is the extortion factor

Solving for X's strategy p[1:4] gives:

$$
\begin{aligned}
p_1 &= 1 - \phi(\chi - 1)\frac{R-P}{P-S} \\
p_2 &= 1 - \phi\left(1 + \chi\frac{T-P}{P-S}\right) \\
p_3 &= \phi\left(\chi + \frac{T-P}{P-S}\right) \\
p_4 &= 0
\end{aligned} \qquad \textbf{[12]}
$$

X’s score depends on Y’s strategy

Feasible strategies exist for any x and sufficiently small ϕ, thus the allowed range of φ

$$0 < \phi \leq \frac{(P-S)}{(P-S)+\chi(T-P)}. \quad [13]$$

X’s score depends on Y’s strategy

If Y chooses q = (1,1,1,1), both X & Y are maximized when Y fully cooperates with

$$s_X = \frac{P(T-R) + \chi[R(T-S) - P(T-R)]}{(T-R) + \chi(R-S)}. \quad \textbf{[14]}$$

What if we reinforced this by the std IPD values?

Reinforced by std IPD values (T=5,R=3,P=1,S=0)

Eq. 12 becomes:

$$\mathbf{p} = [1 - 2\phi(\chi - 1), 1 - \phi(4\chi + 1), \phi(\chi + 4), 0], \qquad \mathbf{[15]}$$

Range: $0 < \Phi < (4\,x + 1\,)^{-1}$

If $\Phi = 1/26$ and $x = 3$

Then Y’s strategy becomes:

$$p = (11/13, ½, 7/26, 0)$$

X extorts more than its fair share

If p = (11/13,½,7/26,0), then

$$s_X = \frac{2+13\chi}{2+3\chi}, \quad s_Y = \frac{12+3\chi}{2+3\chi}. \qquad [16]$$

Best scores ~

S_X = 3.73 and S_Y = 1.91

Extortion against Evolutionary Player

The gradient is readily calculated as the derivative of Y's score and Y's strategy

$$\left.\frac{\partial s_Y}{\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}_0} = \left(0, 0, 0, \frac{(T-S)(S+T-2P)}{(P-S)+\chi(T-P)}\right). \qquad \textbf{[17]}$$

The 4th component is positive for values

(T,R,P,S) = (5,3,1,0)

Discussion

Press and Dyson did not prove analytically

- $\forall$ cases $\exists$ evolutionary paths for Y that yield a maximum score (Eq. 16); nor
- That these paths have positive directional derivatives everywhere along them.

Discussion

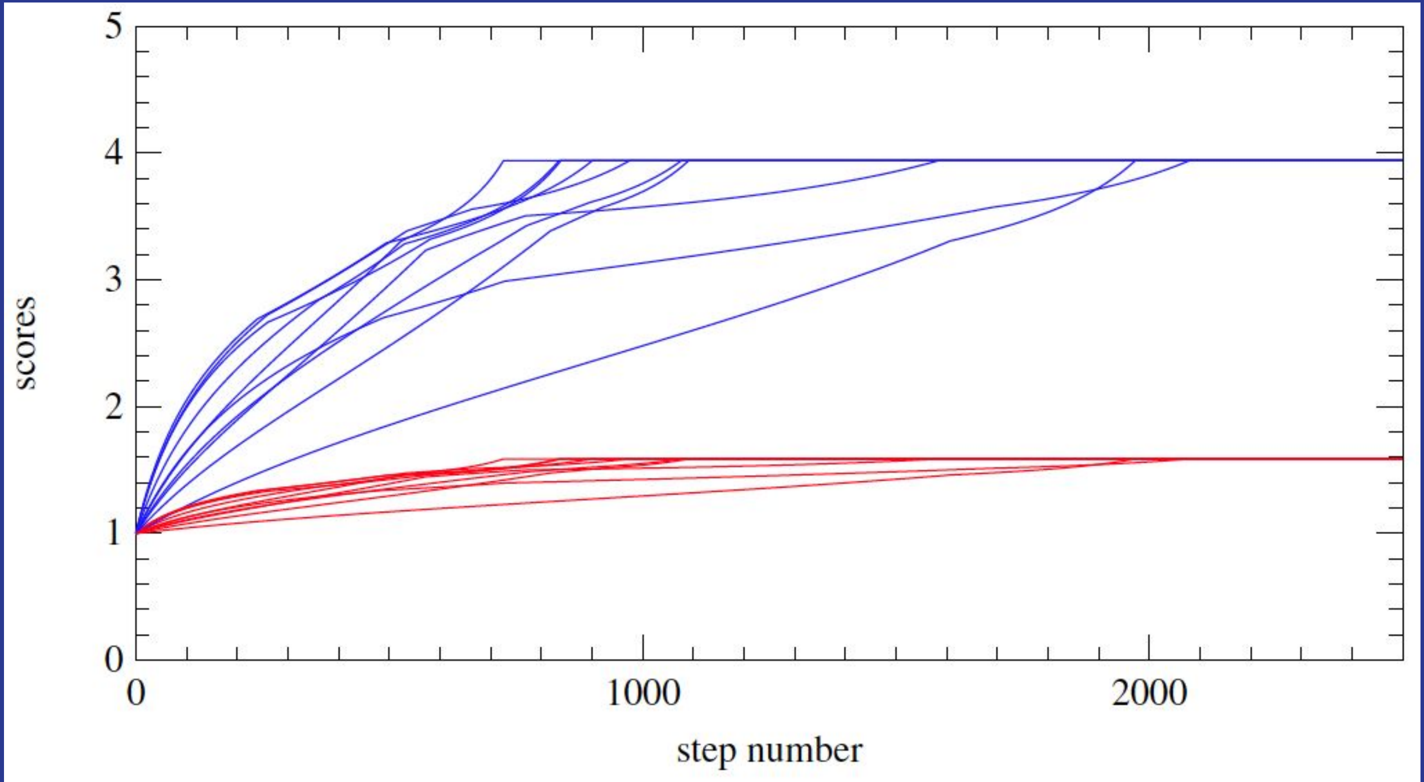
However, X can play an extortionate strategy such that

$x = 5$, with maximum scores

$s_X = 3.94$ and $s_Y = 1.59$

Y can take small steps to locally increase its score

Evolution of X’s (blue) and Y’s (red) scores:



Conclusion

The extort. ZD strategies property to distinguish

- “sentient” and “evolutionary”
- Good at exploring a fitness landscape; but
- Have no theory of mind.

Distinction is only on Y’s ability to impute to X’s ability to alter its strategy, leaving X to alter the extortion factor, x