Zero-Determinant Strategies

Junhao Hua

2014.7.22

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

William H. Pressa, and Freeman J. Dyson

PANS | June 26, 2012

Abstract

The two-player Iterated Prisoner's Dilemma game is a model for both sentient and evolutionary behaviors, especially including the emergence of cooperation. It is generally assumed that there exists no simple ultimatum strategy whereby one player can enforce a unilateral claim to an unfair share of rewards. Here, we show that such strategies unexpectedly do exist. In particular, a player X who is witting of these strategies can (*i*) deterministically set her opponent Y's score, independently of his strategy or response, or (*ii*) enforce an extortionate linear relation between her and his scores. Against such a player, an evolutionary player's best response is to accede to the extortion. Only a player with a theory of mind about his opponent can do better, in which case Iterated Prisoner's Dilemma is an Ultimatum Game.

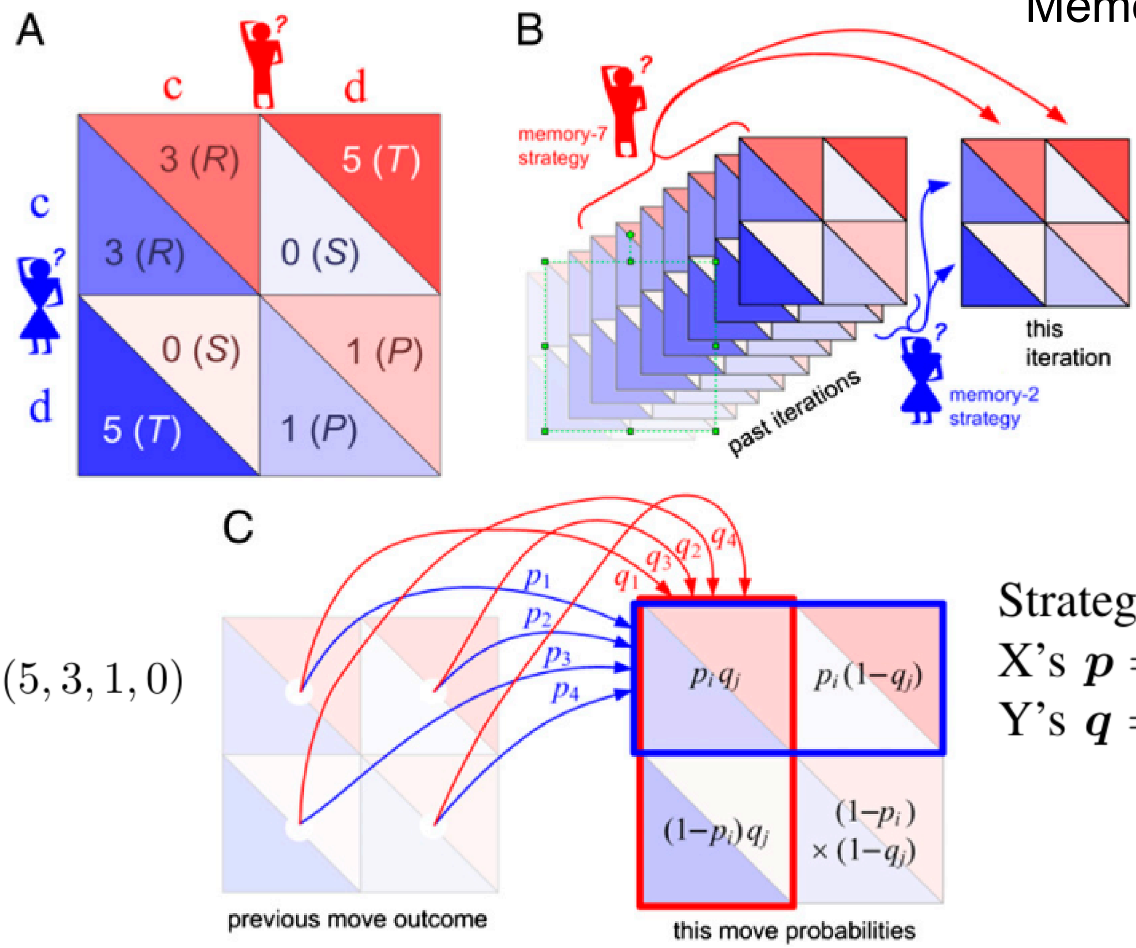
Motivation

- Iterated Prisoner's Dilemma(IPD)
  - have long been touchstone models for elucidating both sentient human behaviors, such as cartel pricing, and Darwinian phenomena, such as the evolution of cooperation.
  - further establish IPD as foundational lore in fields as diverse as political science and evolutionary biology.

This paper found a significant mathematical feature of IPD!

Problem

Memory-one game



$T > R > P > S$

$2R > T + S$

$e.g.(T, R, P, S) = (5, 3, 1, 0)$

Strategy (cc, cd, dc, dd)

X's $\boldsymbol{p} = (p_1, p_2, p_3, p_4)$

Y's $\boldsymbol{q} = (q_1, q_2, q_3, q_4)$

Fig. 1. (*A*) Single play of PD. Players X (blue) and Y (red) each choose to cooperate (c) or defect (d) with respective payoffs *R*, *T*, *S*, or *P* as shown (along with the most common numerical values). (*B*) IPD, where the same two players play arbitrarily many times; each has a strategy based on a finite memory of the previous plays. (*C*) Case of two memory-one players. Each player's strategy is a vector of four probabilities (of cooperation), conditioned on the four outcomes of the previous move.

Methods

- Markov transition matrix M(p,q)

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

Because $\mathbf{M}$ has a unit eigenvalue, the matrix $\mathbf{M}' \equiv \mathbf{M} - \mathbf{I}$ is singular, with thus zero determinant. The stationary vector $\mathbf{v}$ of the Markov matrix, or any vector proportional to it, satisfies

$$\mathbf{v}^T\mathbf{M} = \mathbf{v}^T, \quad \text{or} \quad \mathbf{v}^T\mathbf{M}' = \mathbf{0}. \qquad \textbf{[1]}$$

Cramer's rule, applied to the matrix $\mathbf{M}'$, is

$$\boxed{\mathrm{Adj}(\mathbf{M}')}\mathbf{M}' = \det(\mathbf{M}')\mathbf{I} = 0, \qquad \textbf{[2]}$$

where $\mathrm{Adj}(\mathbf{M}')$ is the adjugate matrix (also known as the classical adjoint or, as in high-school algebra, the "matrix of minors"). Eq. **2** implies that every row of $\mathrm{Adj}(\mathbf{M}')$ is proportional to $\mathbf{v}$. Choosing the fourth row, we see that the components of $\mathbf{v}$ are (up to a sign) the determinants of the 3×3 matrices formed from the first three columns of $\mathbf{M}'$, leaving out each one of the four rows in turn. These determinants are unchanged if we add the first column of $\mathbf{M}'$ into the second and third columns.

Methods (cont.)

Adjugate matrix : $Adj(M') = C^T$
where $C_{ij} = (-1)^{i+j} B_{ij}$ is (i,j) entry of cofactor matrix, with minor B_{ij}.
Using laplace expansion

$$det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} B_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} B_{ij}$$

we have $v \cdot f = \sum_{i=1}^{4} (-1)^{i+1} B_{ij} f_i$

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

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

$$= \det \begin{bmatrix} -1 + p_1 q_1 & -1 + p_1 & -1 + q_1 & f_1 \\ p_2 q_3 & -1 + p_2 & q_3 & f_2 \\ p_3 q_2 & p_3 & -1 + q_2 & f_3 \\ p_4 q_4 & p_4 & q_4 & f_4 \end{bmatrix}$$

$$\begin{array}{cc} \equiv \widetilde{\mathbf{p}} & \equiv \widetilde{\mathbf{q}} \end{array}$$

X's payoff matrix is $\mathbf{S}_X = (R, S, T, P)$, whereas Y's is $\mathbf{S}_Y = (R, T, S, P)$. In the stationary state, their respective scores are then

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

Because the scores s in Eq. **5** depend linearly on their corresponding payoff matrices $\mathbf{S}$, the same is true for any linear combination of scores, giving

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

It is Eq. **6** that now allows much mischief, because both X and Y have the possibility of choosing unilateral strategies that will make the determinant in the numerator vanish. That is, if X chooses a strategy that satisfies $\tilde{\mathbf{p}} = \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1}$, or if Y chooses a strategy with $\tilde{\mathbf{q}} = \alpha \mathbf{S}_X + \beta \mathbf{S}_Y + \gamma \mathbf{1}$, then the determinant vanishes and a linear relation between the two scores,

$$\alpha s_X + \beta s_Y + \gamma = 0 \qquad \textbf{[7]}$$

will be enforced. We call these zero-determinant (ZD) strategies.

Results 1

- X Unilaterally Sets Y's Score. $\tilde{\mathbf{p}} = \beta \mathbf{S}_Y + \gamma \mathbf{1}$

$$p_2 = \frac{p_1(T-P)-(1+p_4)(T-R)}{R-P}$$

$$p_3 = \frac{(1-p_1)(P-S)+p_4(R-S)}{R-P}.$$

$$s_Y = \frac{(1-p_1)P+p_4R}{(1-p_1)+p_4}. \qquad \textbf{[9]}$$

All PD games satisfy $T>R>P>S$. By inspection, Eq. **8** then has feasible solutions whenever p_1 is close to (but $\leq$) 1 and p_4 is close to (but $\geq$) 0. In that case, p_2 is close to (but $\leq$) 1 and p_3 is close to (but $\geq$) zero. Now also by inspection of Eq. **9**, a weighted average of P and R with weights $(1-p_1)$ and p_4, we see that all scores $\boxed{P \leq s_Y \leq R}$ (and no others) can be forced by X. That is, X can set Y's score to any value in the range from the mutual non-cooperation score to the mutual cooperation score.

Results 2

- X Tries to Set Her Own Score. $\tilde{\mathbf{p}} = \alpha \mathbf{S}_X + \gamma \mathbf{1}$

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

This strategy has only one feasible point, the singular strategy $\mathbf{p} = (1, 1, 0, 0)$, "always cooperate or never cooperate." Thus, X cannot unilaterally set her own score in IPD.

Results 3

- X Demands and Gets an Extortionate Share.

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

where $\chi \geq 1$ is the extortion factor. Solving these four equations for the p's gives

$$p_1 = 1 - \phi(\chi - 1)\frac{R-P}{P-S}$$

$$p_2 = 1 - \phi\left(1 + \chi\frac{T-P}{P-S}\right)$$

$$p_3 = \phi\left(\chi + \frac{T-P}{P-S}\right)$$

$$p_4 = 0$$

if $\phi = 0$, produces only the singular strategy (1,1,0,0)

Evidently, feasible strategies exist for any χ and sufficiently small ϕ. It is easy to check that the allowed range of ϕ is

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

Under the extortionate strategy, X's score depends on Y's strategy **q**, and both are maximized when Y fully cooperates, with $\mathbf{q} = (1, 1, 1, 1)$. If Y decides (or evolves) to maximize his score by cooperating fully, then X's score under this strategy is

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

The above discussion can be made more concrete by specializing to the conventional IPD values $(5, 3, 1, 0)$; then, Eq. **12** becomes

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

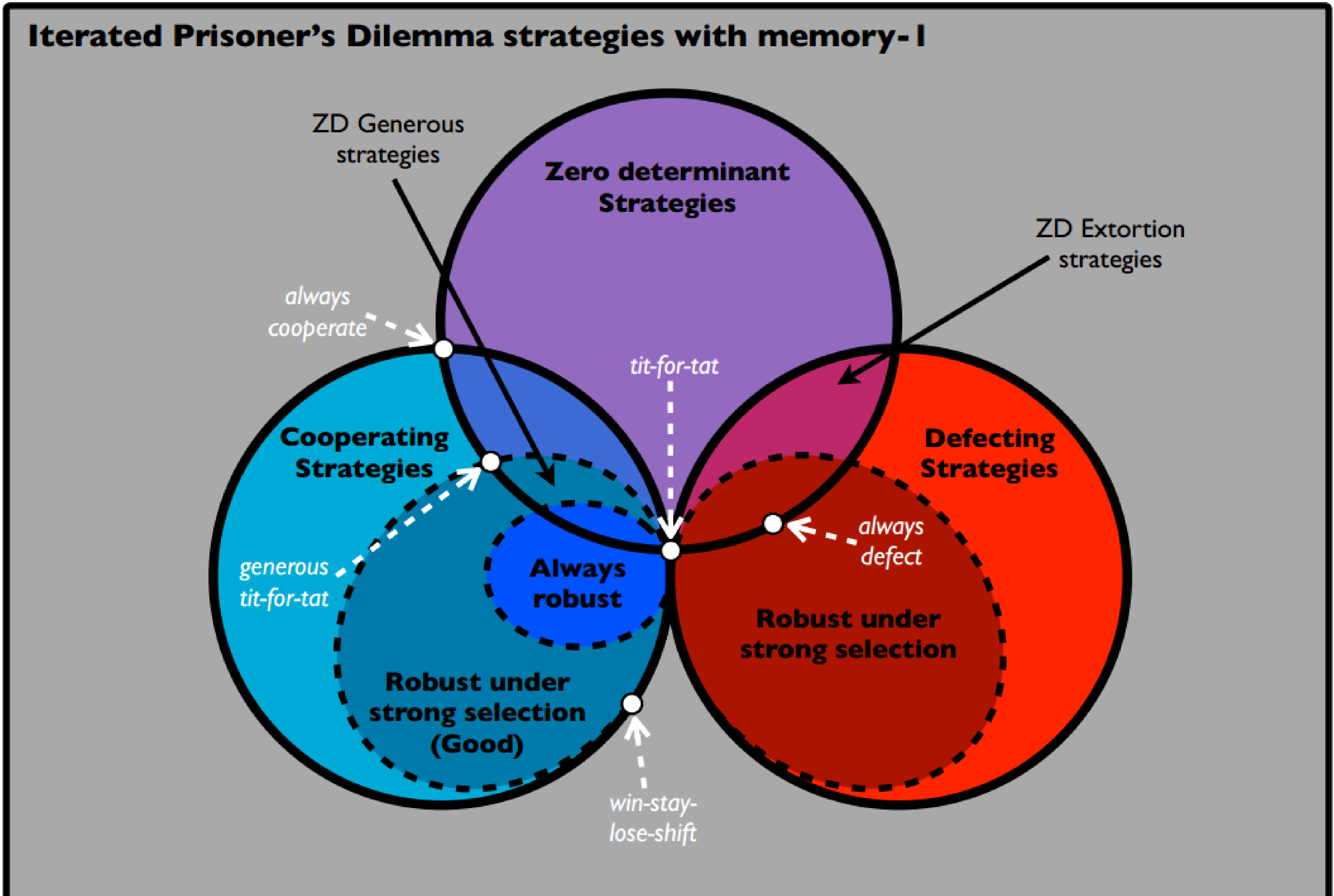
a solution that is both feasible and extortionate for $0 < \phi \le (4\chi + 1)^{-1}$. X's and Y's best respective scores are

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

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

With $\chi > 1$, X's score is always greater than the mutual cooperation value of 3, and Y's is always less. X's limiting score as $\chi \to \infty$ is 13/3. However, in that limit, Y's score is always 1, so there is no incentive for him to cooperate. X's greed is thus limited by the necessity of providing some incentive to Y. The value of ϕ is irrelevant, except that singular cases (where strategies result in infinitely long "duels") are more likely at its extreme values. By way of concreteness, the strategy for X that enforces an extortion factor 3 and sets ϕ at its midpoint value is $\mathbf{p} = \left(\frac{11}{13}, \frac{1}{2}, \frac{7}{26}, 0\right)$, with best scores about $s_X = 3.73$ and $s_Y = 1.91$.

In the special case $\chi = 1$, implying fairness, and $\phi = 1/5$ (one of its limit values), Eq. **15** reduces to the strategy $(1, 0, 1, 0)$, which is the well-known tit-for-tat (TFT) strategy (7). Knowing only TFT among ZD strategies, one might have thought that strategies where X links her score deterministically to Y must always be symmetric, hence fair, with X and Y rewarded equally. The existence of the general ZD strategy shows this not to be the case.



Iterated Prisoner's Dilemma strategies with memory-1
ZD Generous strategies
Zero determinant Strategies
ZD Extortion strategies
always cooperate
tit-for-tat
Cooperating Strategies
Defecting Strategies
always defect
generous tit-for-tat
Always robust
Robust under strong selection
Robust under strong selection (Good)
win-stay-lose-shift

Results 4

- Extortionate Strategy Against an Evolutionary Player.
  - Y is an evolutionary player: adjusts his strategy q to maximize his score S_Y.
  - Y has a theory of mind about X: Y imputes to X an independent strategy, and the ability to alter it in response to his actions

Against X's fixed extortionate ZD strategy, a particularly simple evolutionary strategy for Y, close to if not exactly Darwinian, is for him to make successive small adjustments in **q** and thus climb the gradient in s_Y. [We note that true Darwinian evolution of a trait with multiple loci is, in a population, not strictly "evolutionary" in our loose sense (14)].

Because Y may start out with a fully noncooperative strategy $\mathbf{q}_0 = (0,0,0,0)$, it is in X's interest that her extortionate strategy yield a positive gradient for Y's cooperation at this value of **q**. That gradient is readily calculated as

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

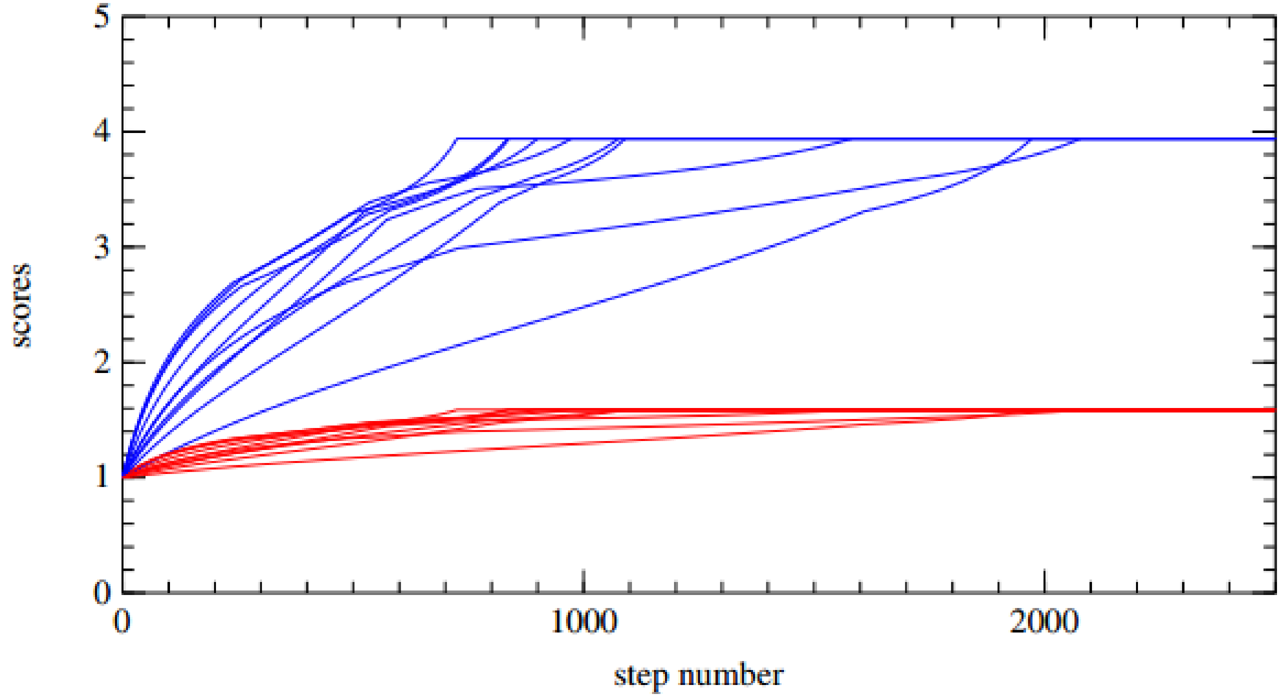


Fig. 3. Evolution of X's score (blue) and Y's score (red) in 10 instances. X plays a fixed extortionate strategy with extortion factor $\chi = 5$. Y evolves by making small steps in a gradient direction that increases his score. The 10 instances show different choices for the weights that Y assigns to different components of the gradient, i.e., how easily he can evolve along each. In all cases, X achieves her maximum possible (extortionate) score.

Discussion

The extortionate ZD strategies have the peculiar property of sharply distinguishing between "sentient" players, who have a theory of mind about their opponents, and "evolutionary" players, who may be arbitrarily good at exploring a fitness landscape (either locally or globally), but who have no theory of mind. The distinction does not depend on the details of any particular theory of mind, but only on Y's ability to impute to X an ability to alter her strategy.

If X alone is witting of ZD strategies, then IPD reduces to one of two cases, depending on whether Y has a theory of mind. If Y has a theory of mind, then IPD is simply an ultimatum game (15, 16), where X proposes an unfair division and Y can either accept or reject the proposal. If he does not (or if, equivalently, X has fixed her strategy and then gone to lunch), then the game is dilemma-free for Y. He can maximize his own score only by giving X even more; there is no benefit to him in defecting.

If X and Y are both witting of ZD, then they may choose to negotiate to each set the other's score to the maximum cooperative value. Unlike naive PD, there is no advantage in defection, because neither can affect his or her own score and each can punish any irrational defection by the other. Nor is this equivalent to the classical TFT strategy (7), which produces indeterminate scores if played by both players.

Discussion(cont.)

To summarize, player X, witting of ZD strategies, sees IPD as a very different game from how it is conventionally viewed. She chooses an extortion factor χ, say 3, and commences play. Now, if she thinks that Y has no theory of mind about her (13) (e.g., he is an evolutionary player), then she should go to lunch leaving her fixed strategy mindlessly in place. Y's evolution will bestow a disproportionate reward on her. However, if she imputes to Y a theory of mind about herself, then she should remain engaged and watch for evidence of Y's refusing the ultimatum (e.g., lack of evolution favorable to both). If she finds such evidence, then her options are those of the ultimatum game (16). For example, she may reduce the value of χ, perhaps to its "fair" value of 1.

Now consider Y's perspective, if he has a theory of mind about X. His only alternative to accepting positive, but meager, rewards is to refuse them, hurting both himself and X. He does this in the hope that X will eventually reduce her extortion factor. However, if she has gone to lunch, then his resistance is futile.

It is worth contemplating that, though an evolutionary player Y is so easily beaten within the confines of the IPD game, it is exactly evolution, on the hugely larger canvas of DNA-based life, that ultimately has produced X, the player with the mind.