The Robustness of Extortion in Iterated Prisoner's Dilemma*

Jing Chen† Aleksey Zinger‡

June 4, 2013

Abstract

Press and Dyson (2012) discovered a special set of strategies, called *extortionate strategies*, in the two-player Iterated Prisoner's Dilemma. Surprisingly, a player using such strategies obtains an advantageous share of the total payoff score of the players. The other player maximizes his own score by fully cooperating, and when doing so he maximizes the score of the extortioner as well. When an extortionary player faces an evolutionary player, Press and Dyson conjecture that there always exist evolutionary paths leading to the maximum possible scores for both players.

In this work we confirm their conjecture in a very strong sense. We show that not only such evolutionary paths always exist, but that actually every evolutionary path leads to the maximum possible scores, although some paths may not lead to full cooperation of the evolutionary player. This is true even in the rare cases where the setup of Press and Dyson (2012) is not directly applicable. Our result shows that extortionate strategies are even more powerful than as pointed out by their discoverers. Given our result, the extortioner is assured to receive his desired score without knowing how the other player will evolve, making it more desirable to extort.

Keywords: Iterated Prisoner's Dilemma, Extortionate strategies, Evolutionary game theory

1 Introduction

The two-player Iterated Prisoner's Dilemma (IPD) is one of the standard models for studying the emergence of cooperative behavior among competitive players. It has long been investigated in economics, political science, evolutionary biology, and computer science.[1] Because IPD has been so widely studied, it was surprising when Press and Dyson (2012) discovered a completely new property of this game, namely, the existence of *Zero-Determinant* (*ZD*) strategies. Roughly speaking, such strategies allow one player to unilaterally set the payoff score of the other, or to enforce a linear relation between the two players' scores. Among such strategies, of particular interest is the class called *extortionate strategies*, which allow one player to ensure a disproportionate share of the total score, as opposite to the previously general belief that no ultimatum strategy can enforce one player an unfair share.

---

*The authors would like to thank Freeman J. Dyson for explaining their result, and some anonymous reviewers for their comments. The first author is supported by the Zurich Financial Services and NSF grant CCF-0832797. The second author is partially supported by the IAS Fund for Math.

†Institute for Advanced Study, Princeton and Department of Computer Science, Stony Brook University, `jingchen@math.ias.edu`.

‡Institute for Advanced Study, Princeton and Department of Mathematics, Stony Brook University, `azinger@math.sunysb.edu`

[1] See Dawkins (1976), Axelrod and Hamilton (1981), Axelrod (1984), Roberts (1985), Axelrod and Dion (1988), Rubinstein (1986), Poundstone (1992), Nowak and Sigmund (1993), Papadimitriou and Yannakakis (1994), Bhaskar and Obara (2002), Ely and Välimäki (2002), Nowak (2006), Kendall et al. (2007), and Bhaskar et al. (2008), as just a few examples.

Player Y

Player X	C	D
C	(R, R)	(S, T)
D	(T, S)	(P, P)

Figure 1: Scores for the players X and Y in each of the four outcomes in a single play of Prisoner's Dilemma

More precisely, the game under consideration is of discrete time and with infinitely many rounds. In each round, the same two players, X and Y, play the one-shot Prisoner's Dilemma (PD). As illustrated in Figure 1, each player can choose to cooperate (C) or to defect (D), without knowing the other's choice. If both cooperate, then each receives score R. If both defect, then each receives a smaller score P. If one cooperates and the other defects, then the defector rips off a score T larger than R, and the cooperator gets ripped off with a score of S smaller than P. The literature typically assumes $2R > T + S$, so that the total score of the players is maximized when both cooperate. For example, a conventional realization of the parameters has $T = 5$, $R = 3$, $P = 1$, and $S = 0$.

Press and Dyson (2012) assume that both players have short memory and what a player does in the current round only depends on the outcome of the previous round, rather than the whole history of the play or the number of rounds played.[2] Accordingly, a (mixed) strategy of a player consists of a mapping from the four possible outcomes of PD to the probabilities of cooperating, and a strategy profile together with a starting outcome determines a Markov chain. Given a strategy profile, the players' payoff scores, s_X and s_Y, are defined to be the expected scores they would receive under the stationary distribution of the Markov chain.

As is formally defined in Section 2, an extortionate strategy of player X is specified by an extortion factor $\chi \geq 1$ (and another parameter ϕ). When using such a strategy, X is guaranteed to have $s_X - P = \chi(s_Y - P)$, no matter what strategy Y uses (if Y always defects, then the scores are both exactly P). In particular, when Y adjusts his own strategy to increase his score, he increases X's score even more, and when he achieves his own maximum score by fully cooperating, X's score is also maximized.[3]

An extortionary player facing an evolutionary player. The results of Press and Dyson (2012) have led to completely new viewpoints on IPD; see the discussion by Stewart and Plotkin (2012). One question which is not completely answered by Press and Dyson (2012) is what a player X witting of extortionate strategies should do if she believes that her opponent Y is an evolutionary player, who tries to improve his own score following some unknown optimization scheme, but without explicitly considering or trying to alter the strategy of X. The answer to this question depends on how Y evolves. Although it is of Y's best interest to fully cooperate, being evolutionary, he may not realize this fact and may only make local movement to gradually improve his score. Since the direction of improvement is not unique, in principle Y might end up at a local optimum and leave X with a score much smaller than what she expects when Y fully cooperates.

[2]As shown in their paper, a player with longer memory will not have any advantage over his/her short-memory opponent. Because ZD strategies only depend on the outcome of the previous round, assuming short-memory players is without any loss of generality.

[3]As proved in the same paper, it does not even matter if Y tries to vary his strategy "faster than the Markov chain can stabilize". If X plays a fixed extortionate (or in general, ZD) strategy, then no matter how fast Y changes his own strategy, when averaged over time it seems like he is using a fixed strategy against X, and the constraint enforced by X still applies.

If this can happen, then X would use an extortionate strategy only if she believes that Y will take a desirable evolutionary path, and would otherwise continue monitoring the behavior of Y and change her strategy when necessary.

Press and Dyson (2012) conjecture that in all cases, that is, with different parameters (R,T,S,P), different starting points of IPD, and different original strategies of Y, there *exist* evolutionary paths that lead to the globally maximum scores. However, the existence of desirable evolutionary paths is not sufficient for one to conclude that X should extort Y. If there are other evolutionary paths where Y ends up at local optimum, it is unclear what X should do, as discussed above. In the numerical experiment of Press and Dyson (2012) for the conventional parameters (that is, $T=5$, $R=3$, $P=1$, and $S=0$) the evolutionary paths examined do not end up at local optimum, but formal analysis of the general case is missing.

Our contribution. We prove the conjecture of Press and Dyson (2012) in a very strong form and analytically justify the use of extortionate strategies against evolutionary players. We show that in all cases, *all* evolutionary paths of Y lead to the maximum scores, although the strategy of Y may not end up at full cooperation. This holds even in some degenerate cases where the analysis of Press and Dyson (2012) does not apply. Accordingly, as long as Y does not stop at a locally sub-optimal strategy and does not evolve at a speed that goes to 0 as time goes to infinity, the dynamics will always end up at the maximum scores. Therefore, it is always "safe" for X to use extortionate strategies, and she will receive her desired score in a very robust way, without knowing which evolutionary path Y will follow.

As an easy consequence of our main result, if X does not want to extort Y, but instead is benevolent and wishes to maximize the total score of the two players, that is, to enforce scores (R,R), then she is able to do so in all cases, via a "fair" extortionate strategy where $\chi=1$. This is true even when Y only evolves selfishly and does not care about the total score at all.

The remaining part of the paper is organized as follows. We recall the results of Press and Dyson (2012) in Section 2. We develop and state our main theorem in Section 3. Section 4 concludes the proof of this theorem by analyzing several degenerate cases. We prove some technical propositions in Section 5 and conclude with a few remarks in Section 6.

2 Review of Zero-Determinant Strategies

Following Press and Dyson (2012), we denote an outcome of a one-shot PD game by $xy \in \{CC,CD,DC,DD\}$, with the first letter being the choice of X. As mentioned in Section 1, for the study of ZD strategies it is sufficient to consider strategies depending only on the outcome of the last round. Such a strategy for X is described by a tuple $\mathbf{p}=(p_1,p_2,p_3,p_4)$, corresponding to the probabilities that X cooperates in the current move when the previous outcome is CC, CD, DC, and DD, respectively. Symmetrically, a strategy of Y is described by a tuple $\mathbf{q}=(q_1,q_2,q_3,q_4)$, corresponding to the probabilities that Y cooperates when the previous outcome is CC, DC, CD, and DD, respectively. Notice that p_2 and q_2 correspond to different outcomes, representing the different views of X and Y: if we switch the letters of an outcome and let the first letter be the choice of Y, then $\mathbf{q}$ represents the probabilities that Y cooperates when the previous outcome is CC, CD, DC, and DD, respectively.

Each pair of strategies $(\mathbf{p},\mathbf{q})$ induces a Markov chain with transition matrix

$$\mathbf{M}(\mathbf{p},\mathbf{q})=\begin{pmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1)\\ p_2q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3)\\ p_3q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2)\\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4)\end{pmatrix}, \tag{1}$$

where rows and columns are indexed from the view of X (that is, by the vector (CC, CD, DC, DD)) and $\mathbf{M}(\mathbf{p},\mathbf{q})_{xy,x'y'}$ represents the probability of seeing outcome $x'y'$ when the previous outcome is xy. The players' scores under $(\mathbf{p},\mathbf{q})$, denoted by $s_X(\mathbf{p},\mathbf{q})$ and $s_Y(\mathbf{p},\mathbf{q})$, are defined to be their expected scores under the stationary distribution $\mathbf{v}(\mathbf{p},\mathbf{q})$, which is treated as a row vector and is multiplied by $\mathbf{M}(\mathbf{p},\mathbf{q})$ on the right.[4]

The fundamental observation of Press and Dyson (2012) is that the stabilized scores $s_X(\mathbf{p},\mathbf{q})$ and $s_Y(\mathbf{p},\mathbf{q})$, in nearly all cases, are given by

$$s_X(\mathbf{p},\mathbf{q}) = \frac{D(\mathbf{p},\mathbf{q},\mathbf{S}_X)}{D(\mathbf{p},\mathbf{q},\mathbf{1})}, \quad s_Y(\mathbf{p},\mathbf{q}) = \frac{D(\mathbf{p},\mathbf{q},\mathbf{S}_Y)}{D(\mathbf{p},\mathbf{q},\mathbf{1})}, \tag{2}$$

where $\mathbf{S}_X = (R, S, T, P)$, $\mathbf{S}_Y = (R, T, S, P)$, $\mathbf{1} = (1, 1, 1, 1)$, and

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

for any $\mathbf{f} = (f_1, f_2, f_3, f_4)$. Equation 2 is valid as long as $D(\mathbf{p},\mathbf{q},\mathbf{1}) \neq 0$.

Equation 2 allows player X to choose, for any fixed $\chi \geq 1$, an *extortionate strategy* $\mathbf{p}$ so that $s_X - P = \chi(s_Y - P)$ for any strategy $\mathbf{q}$ of Y. The value χ is called the *extortion factor*. When $\chi = 1$, strategy $\mathbf{p}$ becomes *fair*.

For computational purposes, we find it convenient to measure all scores relative to P. Thus, we shift down the values of R, T, S, P, s_X, s_Y in Press and Dyson (2012) by P, so that

$$S < P = 0 < R < T \qquad \text{and} \qquad 0 < S + T < 2R. \tag{3}$$

Under this convention, the extortionate strategy $\mathbf{p}$ (specified by equation 12 in Press and Dyson (2012)) becomes

$$\mathbf{p} = \big(1 - \phi(\chi-1)R, 1 - \phi(\chi T - S), \phi(T - \chi S), 0\big), \tag{4}$$

where ϕ can be any value such that

$$0 < \phi(\chi T - S) \leq 1\,, \tag{5}$$

as specified by equation 13 in Press and Dyson (2012).[5] The players' scores under such an extortionate strategy satisfy $s_X = \chi s_Y$ if $D(\mathbf{p},\mathbf{q},\mathbf{1}) \neq 0$.

For each $\chi \geq 1$, the maximum possible scores, taken from equation 14 of Press and Dyson (2012) and rewritten under our convention, are

$$s^*_X(\chi) = R + (\chi-1)\frac{(T-R)R}{T-R+\chi(R-S)}, \quad s^*_Y(\chi) = R - (\chi-1)\frac{(R-S)R}{T-R+\chi(R-S)}. \tag{6}$$

Notice that they do not depend on ϕ. Those scores are achieved at $\mathbf{q} = (1, 1, 1, 1)$, that is, when Y is fully cooperative. A priori, this does not mean that there are evolutionary paths for Y that eventually lead to these scores, as the function $s_Y(\mathbf{q})$ could have local peaks. Press and Dyson (2012) conjecture that,

(a) the directional derivatives of s_Y with respect to q_1, q_2, q_3, q_4 are everywhere positive, and thus

[4] In the degenerate cases, when $\mathbf{M}(\mathbf{p},\mathbf{q})$ has more than one stationary distributions, the scores also depend on the starting outcome.

[5] There is a typo in equation 12 in Press and Dyson (2012): ϕ should be replaced by $(P-S)\phi$ to be consistent with their equation 11. This typo carries over to their equation 13.

(b) Y would evolve to a strategy that achieves the maximum possible scores.

They provide numerical evidence for the latter, for the conventional values of R, T, S, P and the fully non-cooperative initial strategy $(0,0,0,0)$ as the initial strategy of Y. We prove that

(i) the former is true for all values of R, T, S, P satisfying inequalities 3 and for all strategies $\mathbf{q}$, except for some rare degenerate ones, and

(ii) all evolutionary paths for Y lead to the maximum possible scores given by equation 6, but not necessarily to the fully cooperative strategy of Y, in a finite time, even though some evolutionary paths may pass through degenerate points where the score $s_Y(\mathbf{q})$ is not continuous in $\mathbf{q}$.

We further discuss the degenerate cases in Section 4 and show that even in these cases the game will evolve to the maximum possible scores.

3 Main Result

In order to understand the evolutionary paths of Y, we characterize the partial derivatives of s_Y with respect to $\mathbf{q}$. By the second part of equation 2, these partials are well defined as long as $D(\mathbf{p},\mathbf{q},\mathbf{1}) \neq 0$.

With $\mathbf{p}$ given by equation 4, $D(\mathbf{p},\mathbf{q},\mathbf{1})$ can be rewritten as the determinant of a 3×3 matrix:

$$D(\mathbf{p},\mathbf{q},\mathbf{1}) = \phi \det \begin{pmatrix} -1+p_1q_1 & (1-\chi)R & -1+q_1-q_4 \\ p_2q_3 & S-\chi T & q_3-q_4 \\ p_3q_2 & T-\chi S & -1+q_2-q_4 \end{pmatrix}. \tag{7}$$

An important feature of this function is that it is linear in each variable q_1, q_2, q_3, q_4 separately. Thus, in order to check whether $D(\mathbf{p},\mathbf{q},\mathbf{1}) = 0$ for any of the possible values of $\mathbf{q}$, it is sufficient to consider the values of $D(\mathbf{p},\mathbf{q},\mathbf{1})$ only for the extremal values of q, that is, for $q_1, q_2, q_3, q_4 = 0, 1$, which is simple to do.

For example, setting $q_3, q_4 = 0$, we find that $D(\mathbf{p},\mathbf{q},\mathbf{1}) \leq 0$ whenever $q_1, q_2 \in [0,1]$ and the equality holds if and only if either $q_1, q_2 = 1$ or $\chi, q_1 = 1$. Setting $q_3 = 0$ and $q_4 = 1$, we find that $D(\mathbf{p},\mathbf{q},\mathbf{1}) \leq 0$ whenever $q_1, q_2 \in [0,1]$ and the equality holds if and only if $\chi, q_1 = 1$ and $q_2 = 0$. If $\chi > 1$, this implies that $D(\mathbf{p},\mathbf{q},\mathbf{1}) \leq 0$ if $q_3 = 0$, $q_1, q_2, q_4 \in [0,1]$ and the equality holds if and only if $(q_1, q_2, q_4) = (1,1,0)$.

We obtain similar conclusions for all extremal values and thus for all admissible values of q_1, q_2, q_3, q_4, as summarized by the following proposition and proved in Section 5.

Proposition 1. *For every $\mathbf{q} \in [0,1]^4$, $D(\mathbf{p},\mathbf{q},\mathbf{1}) \leq 0$. Furthermore, $D(\mathbf{p},\mathbf{q},\mathbf{1}) = 0$ if and only if*

- *$(q_1, q_2, q_4) = (1,1,0)$, or*
- *$\chi = 1$ and either $(q_1, q_4) = (1,0)$, or $(q_2, q_3, q_4) = (1,1,0)$, or $(q_1, q_2, q_3) = (1,0,0)$, or*
- *$\chi = 1$, $p_2 = 0$, and either $(q_1, q_2) = (1,0)$ or $(q_3, q_4) = (1,0)$.*

Proposition 1 qualitatively describes the denominator of s_Y in equation 2. The enumerator $D(\mathbf{p},\mathbf{q},\mathbf{S}_Y)$ can also be rewritten as the determinant of a 3×3 matrix:

$$D(\mathbf{p},\mathbf{q},\mathbf{S}_Y) = -\phi q_4 \det \begin{pmatrix} -1+p_1q_1 & (1-\chi)R & R \\ p_2q_3 & S-\chi T & T \\ p_3q_2 & T-\chi S & S \end{pmatrix}. \tag{8}$$

Thus, the partial derivatives are described by

$$D(\mathbf{p},\mathbf{q},\mathbf{1})^2 \cdot \frac{\partial s_Y}{\partial q_i} = \frac{\partial D(\mathbf{p},\mathbf{q},\mathbf{S}_Y)}{\partial q_i} D(\mathbf{p},\mathbf{q},\mathbf{1}) - D(\mathbf{p},\mathbf{q},\mathbf{S}_Y)\frac{\partial D(\mathbf{p},\mathbf{q},\mathbf{1})}{\partial q_i}, \tag{9}$$

whenever $D(\mathbf{p},\mathbf{q},\mathbf{1}) \neq 0$. By (7) and (8), the difference in (9) splits as a product of two factors, each of which is linear in q_1, q_2, q_3, q_4 separately. We can thus again determine whether each of the four partials of $s_Y(\mathbf{q})$ is always positive for the admissible values of $\mathbf{q}$ by evaluating each of the two corresponding factors at the extremal values of q_1, q_2, q_3, q_4. From this, we obtain the following proposition, which is also proved in Section 5.

Proposition 2. *For every $\mathbf{q} \in [0,1]^4$ such that $D(\mathbf{p},\mathbf{q},\mathbf{1}) \neq 0$,*

$$\frac{\partial s_Y}{\partial q_1}, \frac{\partial s_Y}{\partial q_2}, \frac{\partial s_Y}{\partial q_3}, \frac{\partial s_Y}{\partial q_4} \geq 0\,. \tag{10}$$

Moreover,

- $\frac{\partial s_Y}{\partial q_1} = 0$ *if and only if* $q_4 = 0$*, or* $(q_2, q_3) = (0,0)$*, or* $(p_2, q_2) = (0,0)$*,*
- $\frac{\partial s_Y}{\partial q_2} = 0$ *if and only if* $q_4 = 0$ *or* $(\chi, q_1) = (1,1)$*,*
- $\frac{\partial s_Y}{\partial q_3} = 0$ *if and only if* $q_4 = 0$*, or* $(q_1, q_2) = (1,1)$*, or* $(\chi, q_1) = (1,1)$*,*
- $\frac{\partial s_Y}{\partial q_4} = 0$ *if and only if* $(q_1, q_2) = (1,1)$*, or* $(\chi, q_1) = (1,1)$*, or* $(\chi, q_2, q_3) = (1,1,1)$*, or* $(\chi, p_2, q_3) = (1,0,1)$*.*

If $D(\mathbf{p},\mathbf{q},\mathbf{1}) \neq 0$, then an *evolutionary path* of Y is a gradient trajectory of s_Y under a particular metric on $[0,1]^4$. For example, under the standard metric on $[0,1]^4$, the gradient vector field $(\frac{\partial s_Y}{\partial q_1}, \ldots, \frac{\partial s_Y}{\partial q_4})$ determines the optimal direction and the speed of how the strategy of Y evolves at each point, if Y treats all directions equally. However, an arbitrary evolutionary Y may have different preferences about directions and speeds at different points, and may not follow the optimal path under the standard metric. Such a Y corresponds to a different metric on $[0,1]^4$, which rescales the components of the standard gradient vector field by positive continuous functions that are bounded away from 0. Since the rescalings are by positive continuous functions bounded away from 0, it is sufficient to focus on the behavior of the partial derivatives of s_Y, and the choice of the metric does not enter the discussion below.

By Proposition 1 and the last statement of Proposition 2, the value of q_4 must increase along any gradient trajectory until the latter reaches either

- $q_4 = 1$ or $(q_1, q_2) = (1,1)$, or
- $q_1 = 1$ or $(q_2, q_3) = (1,1)$, provided $\chi = 1$, or
- $q_3 = 1$, provided $\chi = 1$ and $p_2 = 0$;

see the first two rows in Figure 2. In order to see why such points are reached in finite time, notice that $\frac{\partial s_Y}{\partial q_4} > 0$ outside of such points, and thus q_4 must become and stay positive along a gradient trajectory not starting at such a point. Unless $(\chi, q_1) = (1,1)$, q_2 then must also become and stay positive, making the gradient bounded away from 0 for the remainder of the flow; the case $(\chi, q_1) = (1,1)$ is already on the above list. Thus, a point $\mathbf{q}$ in the above list must be reached in a finite time.

The evolution may not stop at the points above. If a gradient trajectory reaches (or starts at) $q_4 = 1$, the value of q_2 must increase along the trajectory until the latter reaches either

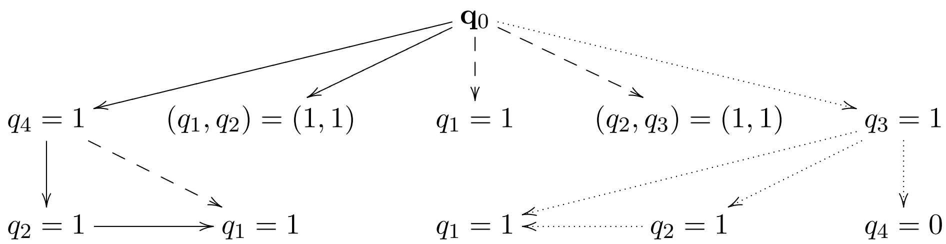


Figure 2: The special points of the possible evolutionary trajectories for Y starting from an initial strategy $\mathbf{q}_0$. The solid arrows apply in all cases. The dashed arrows apply only if $\chi = 1$. The dotted arrows apply only if $\chi = 1$ and $p_2 = 0$.

- $q_2 = 1$, or
- $q_1 = 1$, provided $\chi = 1$,

see the left side of the last two rows in Figure 2. Such a point $\mathbf{q}$ must again be reached in a finite time. By Propositions 1 and 2, a gradient trajectory that reaches $q_2, q_4 = 1$ must also reach $q_1 = 1$.

If $\chi = 1$ and $p_2 = 0$, a gradient trajectory that reaches $q_3 = 1$ and $q_4 > 0$ must then reach either $q_2 = 1$ or $q_1 = 1$; see the right side of the last two rows in Figure 2. In the former case, it eventually reaches $q_1 = 1$. The case $(q_3, q_4) = (1, 0)$ is listed and discussed separately below.

In summary, Propositions 1 and 2 imply that every s_Y-gradient trajectory either starts from or arrives at in a finite amount of time, a limiting strategy $\mathbf{q}$ of the form

- $(1, 1, q_3, q_4)$ if $\chi \neq 1$,
- $(1, q_2, q_3, q_4)$ or $(q_1, 1, 1, 0)$ if $\chi = 1$ and $p_2 \neq 0$,
- $(1, q_2, q_3, q_4)$ or $(q_1, q_2, 1, 0)$ if $\chi = 1$ and $p_2 = 0$.

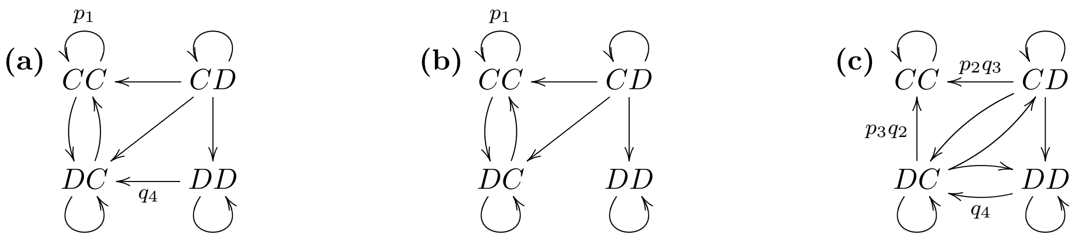
If $\chi > 1$ and $\mathbf{q} = (1, 1, q_3, q_4)$, or if $\chi = 1$ and $\mathbf{q} = (1, q_2, q_3, q_4)$, then there is no incentive for Y to further improve his score, and his strategy will stay there. If $\chi = 1$ and $\mathbf{q} = (q_1, q_2, 1, 0)$, then Y can still improve his score. We show in the last two paragraphs of Section 4 that this case further evolve into a strategy $(1, q_2', 1, q_4')$, but due to a different reasoning. This finally establishes the statement of the theorem below describing the terminal strategy of Y.

In order to compute the final scores, notice that $D(\mathbf{p}, \mathbf{q}, \mathbf{1}) \neq 0$ for most of the above limiting strategies $\mathbf{q}$. Equation 2 applies in these cases and yields the maximal scores in equation 6. These scores can also be computed directly by first determining the stationary distributions first and then combining with the payoff vectors of the PD game; this is done in Section 4. For the remaining strategies, $D(\mathbf{p}, \mathbf{q}, \mathbf{1}) = 0$ and there are two or three different stationary distributions; separate analysis is thus required. This analysis is carried out in Section 4. Below we state the overall conclusion.

Theorem. *Given $\chi \geq 1$ and ϕ satisfying (5), let $s_Y(\mathbf{q})$ be Y's score function corresponding to the extortionate strategy $\mathbf{p}$ of X defined in (4). For every metric on $[0, 1]^4$ and every starting point of Y's strategy, the gradient trajectory of $s_Y(\mathbf{q})$ arrives at a strategy of the form*

- $\mathbf{q} = (1, 1, q_3, q_4)$ *if* $\chi > 1$,
- $\mathbf{q} = (1, q_2, q_3, q_4)$ *if* $\chi = 1$,

in a finite time. Furthermore, the scores reach the maximum scores $s_X^*(\chi)$ and $s_Y^*(\chi)$.



$\chi > 1,\ \mathbf{q} = (1, 1, q_3, q_4 \neq 0)$ $\qquad$ $\chi \geq 1,\ \mathbf{q} = (1, 1, q_3, 0)$ $\qquad$ $\chi = 1,\ \mathbf{q} = (1, q_2, q_3, q_4)$

Figure 3: **(a)** possible transitions of IPD for a typical limiting strategy $\mathbf{q}$ of Y for $\chi > 1$, with the game stabilizing at the CC-DC cycle. **(b)** transitions for a limiting strategy of Y to which the analysis of Press-Dyson does not apply directly. **(c)** possible transitions of IPD for a typical limiting strategy $\mathbf{q}$ of Y for $\chi = 1$, with the game stabilizing at the mutual cooperation outcome.

4 Degenerate Cases and Final Scores

For every strategy $\mathbf{q} = (1, 1, q_3, q_4)$ with $q_4 \neq 0$, it is easy to verify that there is still a unique stationary distribution for the matrix defined in equation (1); it is given by

$$\mathbf{v} = \frac{(p_3, 0, 1-p_1, 0)}{1 - p_1 + p_3}$$

and yields the maximum scores $s^*_X(\chi)$ and $s^*_Y(\chi)$. As s_Y does not depend on (q_3, q_4) in this case, Y has no incentive to adjust q_3 and q_4. Thus, Y's strategy does not necessarily reach $(1, 1, 1, 1)$, but the scores do reach the maximum scores, as conjectured in Press and Dyson (2012). The transition diagram of the Markov chain in this case is depicted in Figure 3(a).

Thus, for $\chi \neq 1$, the only degenerate cases are the strategies $\mathbf{q} = (1, 1, q_3, 0)$, depicted in Figure 3(b). In this case, $D(\mathbf{p}, \mathbf{q}, 1) = 0$ and separate analysis is needed. There are two stationary distributions, $\mathbf{v}$ above and $\mathbf{v}' = (0, 0, 0, 1)$. If the initial state is CC or DC, the IPD cycles between these two states and the stationary distribution is $\mathbf{v}$ with the desired scores as before. If the initial state is CD, the IPD arrives at $\mathbf{v}'$ and the corresponding zero scores with probability

$$(1-p_2)(1-q_3) \sum_{r=0}^{\infty} \big(p_2(1-q_3)\big)^r = 1 - \frac{q_3}{1 - p_2 + p_2 q_3}\,.$$

If $q_3 = 1$, this probability is 0 and the stationary scores are those of $\mathbf{v}$ as desired. If $q_3 < 1$, this probability is positive. However, once $\mathbf{v}'$ is reached, Y's stationary score as a function of q_4 becomes 0 for $q_4 = 0$ and $s^*_Y(\chi)$ for $q_4 > 0$, i.e. Y would benefit from increasing q_4. The game would then eventually enter the CC-DC cycle with its expected scores. The same conclusion holds when the initial state is DD.

If $\chi = 1$ (and thus $p_1 = 1$, $p_2 + p_3 = 1$), then $\mathbf{q} = (1, q_2, q_3, q_4)$ is a possible limiting strategy of the s_Y-gradient trajectory; the corresponding transition diagram is depicted in Figure 3(c). If $q_4 \neq 0$, $(q_2, q_3) \neq (0, 0)$, and $(p_2, q_2) \neq (0, 0)$, then

$$\mathbf{v} = (1, 0, 0, 0)$$

with its scores of $s^*_X(\chi) = s^*_Y(\chi) = R$ is the only stationary distribution (recall that by equations (4) and (5), $p_3 > 0$). As s_Y does not depend on (q_2, q_3, q_4) in this case, Y has no incentive to adjust

q_2, q_3, q_4 to the maximal value 1. If the initial state is CC, then the same conclusion holds even if $q_4 = 0$, or $(q_2, q_3) = (0,0)$, or $(p_2, q_2) = (0,0)$.

If $\chi = 1$, $\mathbf{q} = (1, q_2, q_3, q_4)$, $q_3 \neq 1$, and either $q_4 = 0$, or $(q_2, q_3) = (0,0)$, or $(p_2, q_2) = (0,0)$, then there are two stationary distributions, $\mathbf{v}$ above and

$$\mathbf{v}' = \frac{(0, q_4, q_4, 1-q_3)}{2q_4 + 1-q_3}.$$

At $\mathbf{v}'$, $s_X = s_Y = \alpha(S+T)$ for some $\alpha \in [0, 1/2]$. If q_2 or q_4 (whichever is 0) is increased from 0, Y's stationary score changes to R, the score for the now unique stationary distribution $\mathbf{v}$ above. Since $R > (S+T)/2$, Y would benefit from increasing q_2 or q_4, even by the slightest amount. The game would then eventually enter the cooperative outcome CC with its scores (R, R), as desired. The same reasoning works if $q_3 = 1$, $q_4 \neq 0$, and $(p_2, q_2) = (0,0)$.

If $\chi = 1$, $\mathbf{q} = (1, q_2, 1, 0)$, and $(p_2, q_2) \neq (0,0)$, there are two stationary distributions:

$$\mathbf{v} = (1,0,0,0), \qquad \mathbf{v}' = (0,0,0,1).$$

If the game starts at CD or DC, it eventually arrives at either CC, where it stays, or at DD. In the latter case, increasing q_4 by the slightest amount would eliminate the Nash equilibrium DD as a stationary state and eventually push the game to CC with its score $R > 0$ for Y; so, Y would benefit from increasing q_4 from 0. Thus, Y would eventually modify his strategy, so that the game eventually enters the cooperative state CC.

If $\chi = 1$, $p_2 = 0$, and $\mathbf{q} = (1, 0, 1, 0)$, there are three stationary distributions:

$$\mathbf{v} = (1,0,0,0), \quad \mathbf{v}' = (0, \frac{1}{2}, \frac{1}{2}, 0), \quad \mathbf{v}'' = (0,0,0,1).$$

If the game has entered the CD-DC cycle, increasing q_2 by the slightest amount would eliminate this cycle as a stationary distribution and eventually push the game to CC with its score $R > (S+T)/2$ for Y (as well as X); so, Y would benefit from increasing q_2 from 0. If the game is at the Nash equilibrium DD, increasing q_4 from 0 even by the slightest amount would increase the score to $(S+T)/2 > 0$ for Y; so, Y would benefit from increasing q_4 from 0. Thus, Y would eventually modify his strategy, possibly in two distinct stages, so that the game eventually enters the cooperative state CC.

Separate analysis is also needed for the initial strategies $\mathbf{q} = (q_1, 1, 1, 0)$ if $\chi = 1$ and $p_2 \neq 0$, depicted in Figure 4(a), since $D(\mathbf{p}, \mathbf{q}, \mathbf{1}) = 0$ in these cases. There are two stationary distributions now,

$$\mathbf{v} = \frac{(p_3, 1-q_1, 1-q_1, 0)}{p_3 + 2(1-q_1)}, \qquad \mathbf{v}' = (0,0,0,1),$$

with the scores $s_X = s_Y$ equal

$$\frac{p_3 R + (1-q_1)(S+T)}{p_3 + 2(1-q_1)} > 0$$

and 0, respectively. If the game enters the first stationary distribution, Y would benefit from increasing q_1, since $R > (S+T)/2$; so, his strategy will reach $(1,1,1,0)$ with stationary scores (R, R). If the game is at the second stationary distribution, increasing q_4 from 0 even by the slightest amount would make the score positive for each player; thus Y would benefit from doing so. Thus, Y would eventually modify his strategy, possibly in two distinct stages, and the game enters the cooperative outcome CC in the end.

Finally, separate analysis is again needed for the initial strategies $\mathbf{q} = (q_1, q_2, 1, 0)$ if $\chi = 1$ and $p_2 = 0$, depicted in Figure 4(b), since $D(\mathbf{p}, \mathbf{q}, \mathbf{1}) = 0$ again. The case $q_1 = 1$ is discussed above. If

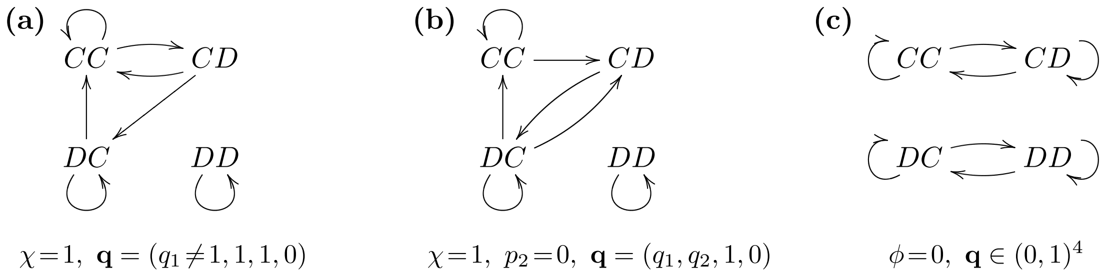


Figure 4: **(a),(b)** possible transitions of IPD for intermediate strategies $\mathbf{q}$ of Y to which the gradient trajectory reasoning does not apply, but the outcome still eventually evolves to mutual cooperation. **(c)** possible transitions of IPD in the $\phi=0$ case, which does not fit with the Press-Dyson scheme and in which an evolutionary player Y evolves his strategy to the constant strategy D.

$q_1<1$, there are two stationary distributions,

$$\mathbf{v} = \frac{(q_2, 1-q_1, 1-q_1, 0)}{q_2 + 2(1-q_1)}, \qquad \mathbf{v}' = (0,0,0,1),$$

with the scores $s_X = s_Y$ equal

$$\frac{q_2 R + (1-q_1)(S+T)}{q_2 + 2(1-q_1)} > 0$$

and 0, respectively. If the game enters the first stationary distribution, Y would benefit from increasing q_1 and q_2, since $R > (S+Y)/2$; so his strategy will reach $(1, q_2, 1, 0)$ with stationary scores (R, R). If the game arrives at the second stationary distribution, increasing q_4 from 0 even by the slightest amount would make the score positive for each player; thus Y would benefit from doing so. Thus, Y would eventually modify his strategy, possibly in two distinct stages, and the game enters the cooperative outcome in the end.

5 Proofs of Propositions 1 and 2

Recall that $\phi>0$ following inequality 5. Thus, $p_1 > 0$, $p_1 > p_2$, $p_3 < 1$, and $p_1<1$ unless $\chi=1$. Moreover,

$$\big(1 - p_2 - p_3\big)R = (1-p_1)(T+S), \qquad p_3(\chi-1)R = (1-p_1)(T-\chi S),$$
$$p_3(\chi T - S) - (p_1-p_2)(T-\chi S) = (\chi-1)R(T-\chi S)\phi.$$

These observations are used below.

Proof of Proposition 1. We first use equation 7 to confirm Proposition 1. Since this function is linear in each q_1, q_2, q_3, q_4 separately, it is sufficient to consider $D(\mathbf{p}, \mathbf{q}, \mathbf{1})$ for the extremal values of q_1, q_2, q_3, q_4. For $(q_3, q_4) = (0, 0)$,

$$-\phi^{-1} D(\mathbf{p}, \mathbf{q}, \mathbf{1}) = \big((1-q_2)(1-p_1 q_1) + p_3 q_2(1-q_1)\big)(\chi T - S) \geq 0;$$

the equality holds if and only if either $q_1, q_2 = 1$ or $\chi, q_1 = 1$. For $(q_3, q_4) = (1, 0)$,

$$-\phi^{-1} D(\mathbf{p}, \mathbf{q}, \mathbf{1}) = (1-q_2)\big((1-p_1 q_1)(\chi-1)(T+S) + p_2(\chi-1)R + (1-q_1)p_2(T-\chi S)\big)$$
$$+ q_2\big((1-q_1)(p_3(\chi T - S) - (p_1-p_2)(T-\chi S)) + p_3(\chi-1)R - (1-p_1)(T-\chi S)\big) \geq 0;$$

the equality holds if and only if either $q_1, q_2 = 1$ or $\chi = 1$ along with one of $q_1 = 1$, $q_2 = 1$, or $p_2 = 0$. For $(q_3, q_4) = (0, 1)$,

$$
\begin{aligned}
-\phi^{-1} D(\mathbf{p}, \mathbf{q}, \mathbf{1}) &= (1-p_1q_1)(2-q_2)(\chi T - S) + \big((1-q_2)(1-p_1q_1) + p_1q_2(1-q_1)\big)(T - \chi S) \\
&\quad + (2-q_1)p_3q_2(\chi T - S) + q_2\big((1-p_1)(T-\chi S) - p_3(\chi-1)R\big) \geq 0;
\end{aligned}
$$

the equality holds if and only if $\chi, q_1 = 1$ and $q_2 = 0$. For $(q_3, q_4) = (1, 1)$,

$$
\begin{aligned}
-\phi^{-1} D(\mathbf{p}, \mathbf{q}, \mathbf{1}) &= \big((1-p_1q_1 + (2-q_1)p_3)q_2 + 2(1-q_2)(1-p_1q_1)\big)(\chi T - S) \\
&\quad + p_2\big((2-q_2)(\chi-1)R + (2-q_1)(T-\chi S)\big) \geq 0;
\end{aligned}
$$

the equality holds if and only if $\chi, q_1 = 1$ and $p_2, q_2 = 0$. □

Proof of Proposition 2. We next use a similar approach to confirm Proposition 2. By equations 2, 7, and 8,

$$
D(\mathbf{p}, \mathbf{q}, \mathbf{1})^2 \frac{\partial s_Y}{\partial q_4} = \begin{pmatrix} -1 + p_1q_1 & R & R \\ p_2q_3 & S & T \\ p_3q_2 & T & S \end{pmatrix} \begin{pmatrix} -1 + p_1q_1 & (1-\chi)R & 1 - q_1 \\ p_2q_3 & S - \chi T & -q_3 \\ p_3q_2 & T - \chi S & 1 - q_2 \end{pmatrix}.
$$

The first determinant above is

$$
(1-p_1q_1)(T^2 - S^2) + (p_2q_3 + p_3q_2)R(T-S) \geq 0;
$$

the equality holds if and only if $\chi = 1$, $q_1 = 1$, $q_2 = 1$, and either $q_3 = 0$ or $p_2 = 0$. The second determinant above is linear in q_1, q_2, and q_3. Thus, it is sufficient to check that it is nonnegative at the eight extremal values $q_1, q_2, q_3 = 0, 1$. For $q_3 = 0$, this determinant is

$$
\big((1-q_2)(1-p_1q_1) + (1-q_1)p_3q_2\big)(\chi T - S) \geq 0;
$$

the equality holds if and only if either $q_1, q_2 = 1$ or $\chi, q_1 = 1$. For $q_3 = 1$, we obtain

$$
\begin{aligned}
&q_2 = 0\colon (\chi-1)\big(p_2R + (1-p_1q_1)(T+S)\big) + (1-q_1)p_2(T-\chi S) \geq 0, \\
&q_2 = 1\colon \phi(1-q_1)(\chi-1)R(T-\chi S) \geq 0.
\end{aligned}
$$

The equality holds if and only if either $q_1, q_2 = 1$, or $\chi, q_2 = 1$, $(\chi, q_1, q_2) = (1, 1, 0)$, or $(\chi, p_2, q_2) = (1, 0, 0)$.

By equations 2, 7, and 8, we find that

$$
\begin{aligned}
D(\mathbf{p}, \mathbf{q}, \mathbf{1})^2 \frac{\partial s_Y}{\partial q_1} &= q_4(T-S)\big(p_3q_2(\chi T - S) + p_2q_3(T - \chi S)\big) \\
&\quad \times \Big(\big((1-q_2+2q_4)p_1 - (p_1-p_2)q_3 + p_3q_2\big)R + (1-p_1-p_1q_4)(T+S)\Big).
\end{aligned}
$$

The factors on the first line are nonnegative and vanish if and only if $q_4 = 0$, or $(q_2, q_3) = 0$, or $(p_2, q_2) = (0, 0)$. The last factor above equals

$$
p_1q_4(2R - (T+S)) + \big((1-p_3)(1-q_2) + (1-p_1)q_2 + (p_1-p_2)(1-q_3)\big)R > 0.
$$

Similarly,

$$
\begin{aligned}
D(\mathbf{p}, \mathbf{q}, \mathbf{1})^2 \frac{\partial s_Y}{\partial q_2} &= q_4(T-S)\big(p_2q_3(\chi-1)R + (1-p_1q_1)(\chi T - S)\big) \\
&\times \Big(p_3q_4(2R - (T+S)) + \big((1-q_3)p_3 + p_2q_3\big)R + \big((1-p_3)(1-q_1) + (1-p_1)q_1\big)(T+S)\Big) \geq 0;
\end{aligned}
$$

the equality holds if and only if either $q_4 = 0$ or $(\chi, q_1) = (1, 1)$. Finally,

$$
\begin{aligned}
D(\mathbf{p}, \mathbf{q}, \mathbf{1})^2 \frac{\partial s_Y}{\partial q_3} &= q_4(T-S)\Big(\big((1-q_2)p_3(\chi-1) - p_3(\chi-1)\big)R + \big((1-q_1)p_1 + (1-p_1)\big)(T - \chi S)\Big) \\
&\times \Big(p_2q_4(2R - (T+S)) + \big((1-q_2)p_2 + p_3q_2\big)R + \big((1-p_2)(1-q_1) + (1-p_1)q_1\big)(T+S)\Big) \geq 0;
\end{aligned}
$$

the equality holds if and only if either $q_4 = 0$, or $q_1, q_2 = 1$, or $\chi, q_1 = 1$. □

6 Concluding Remarks

In all the discussion above we have $\phi > 0$ following inequality 5. The case $\phi = 0$ is formally allowed in Press and Dyson (2012), but produces only the singular strategy $\mathbf{p} = (1, 1, 0, 0)$ according to equation 4. For this strategy for X and most strategies $\mathbf{q}$ for Y, there are two stationary distributions,

$$\mathbf{v}_C = \frac{(q_3, 1-q_1, 0, 0)}{1 - q_1 + q_3}, \qquad \mathbf{v}_D = \frac{(0, 0, q_4, 1-q_2)}{1 - q_2 + q_4},$$

as depicted in Figure 4(c). Under either distribution, Y benefits from evolving to the fully non-cooperative strategy. The same holds for degenerate values of $\mathbf{q}$ as well. Thus, this strategy $\mathbf{p}$ does not work out well for X. This is not too surprising, as this is not an extortionate strategy in the sense of Press and Dyson (2012). In fact, when $\phi = 0$, equation 11 in Press and Dyson (2012) imposes no (extortionate or not) condition on the scores and the analysis of Press and Dyson (2012), including equation 14 in Press and Dyson (2012), does not apply.

We have thus formally shown that the extortionate strategies discovered by Press and Dyson (2012) are extremely robust at achieving high scores for a sentient player against an evolutionary opponent. Indeed, a sentient X can force an evolutionary Y into effectively full cooperation in all cases: Y's strategy may not evolve to $(1, 1, 1, 1)$, but in the long run Y will always cooperate, because the outcomes at which Y may not cooperate (i.e., $q_i < 1$) never occur when Y stops evolving and the Markov equilibrium is established. So, the score of X ends up being the same as if Y is fully cooperative, and always achieves the maximum possible score for a given extortion factor χ.

Finally, it is immediate from equation 6 that the score for the first player increases with χ, approaching $R(T-S)/(R-S)$ as χ approaches infinity. However, the total score,

$$s_X^*(\chi) + s_Y^*(\chi) = 2R - (\chi-1)\frac{(2R-(T+S))R}{T-R+\chi(R-S)},$$

decreases with χ, and also approaches $R(T-S)/(R-S)$ as χ approaches infinity. It is maximized at $\chi = 1$, with each player receiving the mutual cooperation reward R. In particular, the $\chi = 1$ outcome is the most desirable from the point of view of social welfare (maximizing the total benefit) and least desirable from the point of view of the sentient player: there is always some social welfare "burnt" when one player tries to extortionate the other.

References

Axelrod, Robert (1984), *The Evolution of Cooperation*. Basic Books.

Axelrod, Robert and Douglas Dion (1988), "The further evolution of cooperation." *Science*, 242, 1385–1390.

Axelrod, Robert and William D. Hamilton (1981), "The evolution of cooperation." *Science*, 211, 1390–1396.

Bhaskar, V., George J. Mailath, and Stephen Morris (2008), "Purification in the infinitely repeated prisoners dilemma." *Review of Economic Dynamics*, 11, 515528.

Bhaskar, V. and Ichiro Obara (2002), "Belief-based equilibria in the repeated prisoners dilemma with private monitoring." *Journal of Economic Theory*, 102, 4069.

Dawkins, Richard (1976), *The Selfish Gene*. Oxford University Press.

Ely, Jeffrey C. and Juuso Välimäki (2002), "A robust folk theorem for the prisoners dilemma." *Journal of Economic Theory*, 102, 84105.

Kendall, Graham, Xin Yao, and Siang Yew Chong (2007), *The Iterated Prisoners' Dilemma: 20 Years On*. World Scientific.

Nowak, Martin A. (2006), "Five rules for the evolution of cooperation." *Science*, 314, 1560–1563.

Nowak, Martin A. and Karl Sigmund (1993), "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game." *Nature*, 364, 56–58.

Papadimitriou, Christos H. and Mihalis Yannakakis (1994), "On complexity as bounded rationality." *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, 726–733.

Poundstone, William (1992), *Prisoner's Dilemma*. Doubleday.

Press, William H. and Freeman J. Dyson (2012), "Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent." *Proceedings of the National Academy of Sciences*, 109, 10409–10413.

Roberts, Kevin (1985), "Cartel behavior and adverse selection." *Journal of Industrial Economics*, 33, 401–413.

Rubinstein, Ariel (1986), "Finite automata play the repeate prisoners dilemma." *Journal of Economic Theory*, 39, 83–96.

Stewart, Alexander J. and Joshua B. Plotkin (2012), "Extortion and cooperation in the prisoner's dilemma." *Proceedings of the National Academy of Sciences*, 109, 10134–10135.