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

William H. Press[a,1] and Freeman J. Dyson[b]

[a]Department of Computer Science and School of Biological Sciences, University of Texas at Austin, Austin, TX 78712; and [b]School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540

Contributed by William H. Press, April 19, 2012 (sent for review March 14, 2012)

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.

evolution of cooperation | game theory | tit for tat

Iterated 2×2 games, with Iterated Prisoner's Dilemma (IPD) as the notable example, have long been touchstone models for elucidating both sentient human behaviors, such as cartel pricing, and Darwinian phenomena, such as the evolution of cooperation (1–6). Well-known popular treatments (7–9) have further established IPD as foundational lore in fields as diverse as political science and evolutionary biology. It would be surprising if any significant mathematical feature of IPD has remained undescribed, but that appears to be the case, as we show in this paper.

Fig. 1*A* shows the setup for a single play of Prisoner's Dilemma (PD). If X and Y cooperate (c), then each earns a reward R. If one defects (d), the defector gets an even larger payment T, and the naive cooperator gets S, usually zero. However, if both defect, then both get a meager payment P. To be interesting, the game must satisfy two inequalities: $T > R > P > S$ guarantees that the Nash equilibrium of the game is mutual defection, whereas $2R > T + S$ makes mutual cooperation the globally best outcome. The "conventional values" $(T, R, P, S) = (5, 3, 1, 0)$ occur most often in the literature. We derive most results in the general case, and indicate when there is a specialization to the conventional values.

Fig. 1*B* shows an iterated IPD game consisting of multiple, successive plays by the same opponents. Opponents may now condition their play on their opponent's strategy insofar as each can deduce it from the previous play. However, we give each player only a finite memory of previous play (10). One might have thought that a player with longer memory always has the advantage over a more forgetful player. In the game of bridge, for example, a player who remembers all of the cards played has the advantage over a player who remembers only the last trick; however, that is not the case when the same game (same allowed moves and same payoff matrices) is indefinitely repeated. In fact, it is easy to prove (*Appendix A*) that, for any strategy of the longer-memory player Y, shorter-memory X's score is exactly the same as if Y had played a certain shorter-memory strategy (roughly, the marginalization of Y's long-memory strategy: its average over states remembered by Y but not by X), disregarding any history in excess of that shared with X. This fact is important. We derive strategies for X assuming that both players have memory of only a single previous move, and the above theorem shows that this involves no loss of generality. Longer memory will not give Y any advantage.

Fig. 1*C*, then, shows the most general memory-one game. The four outcomes of the previous move are labeled $1, \ldots, 4$ for the respective outcomes $xy \in (cc, cd, dc, dd)$, where c and d denote cooperation and defection. X's strategy is $\mathbf{p} = (p_1, p_2, p_3, p_4)$, her probabilities for cooperating under each of the previous outcomes. Y's strategy is analogously $\mathbf{q} = (q_1, q_2, q_3, q_4)$ for outcomes seen from his perspective, that is, in the order of $yx \in (cc, cd, dc, dd)$. The outcome of this play is determined by a product of probabilities, as shown in Fig. 1.

Methods and Results

Zero-Determinant Strategies. As is well understood (10), it is not necessary to simulate the play of strategies $\mathbf{p}$ against $\mathbf{q}$ move by move. Rather, $\mathbf{p}$ and $\mathbf{q}$ imply a Markov matrix whose stationary vector $\mathbf{v}$, combined with the respective payoff matrices, yields an expected outcome for each player. (We discuss the possibility of nonstationary play later in the paper.) With rows and columns of the matrix in X's order, the Markov transition matrix $\mathbf{M}(\mathbf{p}, \mathbf{q})$ from one move to the next is shown in Fig. 2*A*.

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

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

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

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

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.

The result of these manipulations is a formula for the dot product of an arbitrary four-vector $\mathbf{f}$ with the stationary vector $\mathbf{v}$ of the Markov matrix, $\mathbf{v} \cdot \mathbf{f} \equiv D(\mathbf{p}, \mathbf{q}, \mathbf{f})$, where D is the 4×4 determinant shown explicitly in Fig. 2*B*. This result follows from expanding the determinant by minors on its fourth column and noting that the 3×3 determinants multiplying each f_i are just the ones described above. What is noteworthy about this formula for $\mathbf{v} \cdot \mathbf{f}$ is that it is a determinant whose second column,

$$\tilde{\mathbf{p}} \equiv (-1 + p_1, -1 + p_2, p_3, p_4), \qquad [3]$$

is solely under the control of X; whose third column,

Author contributions: W.H.P. and F.J.D. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

See Commentary on page 10134.

[1]To whom correspondence should be addressed. E-mail: wpress@cs.utexas.edu.

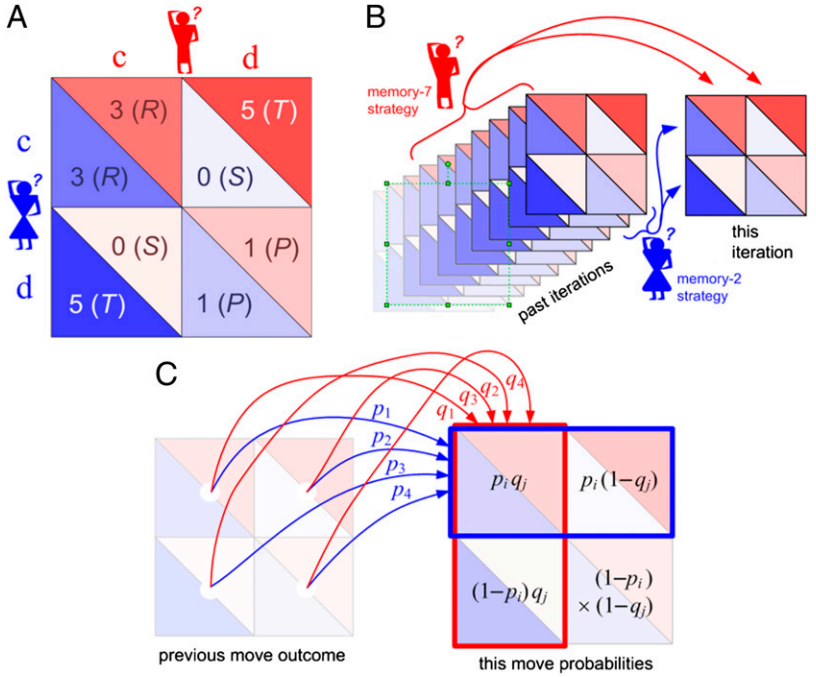

Fig. 1. (*A*) Single play of PD. Players X (blue) and Y (red) each choose to cooperate (c) or defect (d) with respective payoffs *R*, *T*, *S*, or *P* as shown (along with the most common numerical values). (*B*) IPD, where the same two players play arbitrarily many times; each has a strategy based on 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.

$$\tilde{\mathbf{q}} \equiv (-1+q_1,\, q_3,\, -1+q_2,\, q_4), \qquad [4]$$

is solely under the control of Y; and whose fourth column is simply **f**.

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

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

where **1** is the vector with all components 1. The denominators are needed because **v** has not previously been normalized to have its components sum to 1 (as required for a stationary probability vector).

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

Fig. 2. (*A*) Markov matrix for the memory-one game shown in Fig. 1*C*. (*B*) The dot product of any vector **f** with the Markov matrix stationary vector **v** can be calculated as a determinant in which, notably, a column depends only on one player's strategy.

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

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

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,

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

will be enforced. We call these zero-determinant (ZD) strategies. We are not aware of any previous recognition of these strategies in the literature; they exist algebraically not only in IPD but in all iterated 2×2 games. However, not all ZD strategies are feasible, with probabilities **p** all in the range $[0,1]$. Whether they are feasible in any particular instance depends on the particulars of the application, as we now see.

X Unilaterally Sets Y's Score. One specialization of ZD strategies allows X to unilaterally set Y's score. From the above, X need only play a fixed strategy satisfying $\tilde{\mathbf{p}} = \beta\mathbf{S}_Y + \gamma\mathbf{1}$ (i.e., set $\alpha = 0$ in Eq. **7**), four equations that we can solve for p_2 and p_3 in terms of p_1 and p_4, that is, eliminating the nuisance parameters β and γ. The result, for general R,S,T,P (not necessarily a PD game), is

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

With this substitution, Y's score (Eq. **5**) becomes

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

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 $P \leq s_Y \leq R$ (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 noncooperation score to the mutual cooperation score.

What is surprising is not that Y can, with X's connivance, achieve scores in this range, but that X can force any particular score by a fixed strategy **p**, independent of Y's strategy **q**. In other words, there is no need for X to react to Y, except on a timescale of her own choosing. A consequence is that X can simulate or "spoof" any desired fitness landscape for Y that she wants, thereby guiding his evolutionary path. For example, X might condition Y's score on some arbitrary property of his last 1,000 moves, and thus present him with a simulated fitness landscape that rewards that arbitrary property. (We discuss the issue of timescales further, below.)

X Tries to Set Her Own Score. What if X tries to set her own score? The analogous calculation with $\tilde{\mathbf{p}} = \alpha \mathbf{S}_X + \gamma \mathbf{1}$ yields

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

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.

X Demands and Gets an Extortionate Share. Next, what if X attempts to enforce an extortionate share of payoffs larger than the mutual noncooperation value P? She can do this by choosing

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

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

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

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

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

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

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

The coefficients in the numerator and denominator are all positive as a consequence of $T > R > P > S$. The case $\phi = 0$ is formally allowed, but produces only the singular strategy $(1,1,0,0)$ mentioned above.

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

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

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

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

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.

Extortionate Strategy Against an Evolutionary Player. We can say, loosely, that Y is an evolutionary player if he adjusts his strategy **q** according to some optimization scheme designed to maximize his score s_Y, but does not otherwise explicitly consider X's score or her own strategy. In the alternative case that Y imputes to X an independent strategy, and the ability to alter it in response to his actions, we can say that Y has a theory of mind about X (11–13).

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

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

The fourth component is positive for the conventional values $(T,R,P,S) = (5,3,1,0)$, but we see that it can become negative as P approaches R, because we have $2R > S + T$. With the conventional values, however, evolution away from the origin yields positive gradients for the other three components.

We have not proved analytically that there exist in all cases evolutionary paths for Y that lead to the maximum possible scores (Eq. **16**) and that have positive directional derivatives everywhere along them. However, this assertion seems likely from numerical evidence, at least for the conventional values. Fig. 3 shows a typical numerical experiment in which X plays an extortionate strategy (here, $\chi = 5$, with maximum scores $s_X = 3.94$ and $s_Y = 1.59$), and Y takes small steps that locally increase his score. Y has no unique gradient direction because the mapping from the score gradient (a covariant vector) to the step direction (a contravariant vector) involves an arbitrary metric, signifying how easily Y can evolve in each direction. Fig. 3 shows 10 arbitrary choices for this metric. In no case does Y's evolution get hung up at a local maximum. That is, all of the evolutions shown (and all of a much larger number tried) reach the value of Eq. **14**.

Discussion

We have several times alluded to issues of timescale. The ZD strategies are derived mathematically under the assumption that the players' expected scores are generated by a Markov stationary state defined by their respective strategies **p** and **q**. However, we have also suggested situations in which X may vary her ZD strategy so as to spoof Y with a fictitious fitness landscape. The question also arises whether Y can somehow vary his strategy on timescales faster than that for Markov equilibrium to be established. Perhaps by playing "inside the equilibration timescale" he can evade the linear constraint on scores (Eq. **7**) imposed by X.

Interestingly, it is easy to prove that this latter situation cannot occur (*Appendix B*). If X plays a constant ZD strategy, then any strategy of Y's, rapidly varying or not, turns out to be equivalent (from X's perspective) to a fixed strategy against which X's imposition of a constraint is effective.

In the former situation, where it is X whose strategies are changing (e.g., among ZD strategies that set Y's score), things are not as crisp. Because X must be basing her decisions on Y's behavior, which only becomes evident with averaging over time, the possibility of a race condition between X's and Y's responses is present with or without Markov equilibration. This reason is sufficient for X to vary her strategy only slowly. If X chooses components of **p** in Eqs. **9** and **8** that are bounded away from the extreme values 0 and 1, then the Markov equilibration time will not be long and thus not a consideration. In short, a deliberate X still has the upper hand.

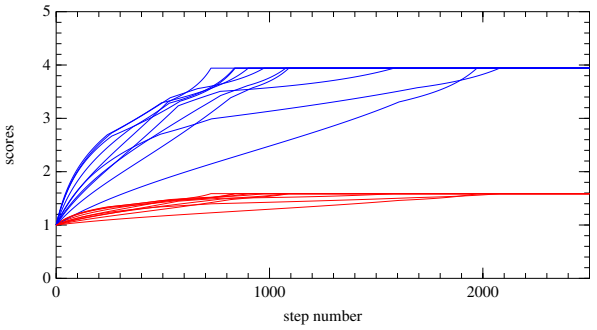

Fig. 3. Evolution of X's score (blue) and Y's score (red) in 10 instances. X plays a fixed extortionate strategy with extortion factor $\chi = 5$. Y evolves by making small steps in a gradient direction that increases his score. The 10 instances show different choices for the weights that Y assigns to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

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.

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.

Appendix A: Shortest-Memory Player Sets the Rules of the Game. In iterated play of a fixed game, one might have thought that a player Y with longer memory of past outcomes has the advantage over a more forgetful player X. For example, one might have thought that player Y could devise an intricate strategy that uses X's last 1,000 plays as input data in a decision algorithm, and that can then beat X's strategy, conditioned on only the last one iteration. However, that is not the case when the same game (same allowed moves and same payoff matrices) is indefinitely repeated. In fact, for any strategy of the longer-memory player Y, X's score is exactly the same as if Y had played a certain shorter-memory strategy (roughly, the marginalization of Y's long-memory strategy), disregarding any history in excess of that shared with X.

Let X and Y be random variables with values x and y that are the players' respective moves on a given iteration. Because their scores depend only on (x,y) separately at each time, a sufficient statistic is the expectation of the joint probability of (X,Y) over past histories H (of course in their proportion seen). Let $H = [H_0, H_1]$, where H_0 is the recent history shared by both X and Y, and H_1 is the older history seen only by Y. Then a straightforward calculation is,

$$\begin{aligned}\langle P(x,y|H_0,H_1)\rangle_{H_0,H_1} &= \sum_{H_0,H_1} P(x,y|H_0,H_1)P(H_0,H_1)\\ &= \sum_{H_0,H_1} P(x|H_0)P(y|H_0,H_1)P(H_0,H_1)\\ &= \sum_{H_0} P(x|H_0)\left[\sum_{H_1} P(y|H_0,H_1)P(H_1|H_0)P(H_0)\right]\\ &= \sum_{H_0} P(x|H_0)\left[\sum_{H_1} P(y,H_1|H_0)\right]P(H_0)\\ &= \sum_{H_0} P(x|H_0)P(y|H_0)P(H_0)\\ &= \langle P(x,y|H_0)\rangle_{H_0}\end{aligned} \quad [18]$$

Here, the first line makes explicit the expectation, and the second line expresses conditional independence.

Thus, the result is a game conditioned only on H_0, where Y plays the marginalized strategy

$$P(y|H_0) \equiv \sum_{H_1} P(y,H_1|H_0). \quad [19]$$

Because this strategy depends on H_0 only, it is a short-memory strategy that produces exactly the same game results as Y's original long-memory strategy.

Note that if Y actually wants to compute the short-memory strategy equivalent to his long-memory strategy, he has to play or simulate the game long enough to compute the above expectations over the histories that would have occurred for his long-memory strategy. Then, knowing these expectations, he can, if he wants, switch to the equivalent short-memory strategy.

To understand this result intuitively, we can view the game from the forgetful player X's perspective: If X thinks that Y's memory is the same as her own, she imputes to Y a vector of probabilities the same length as his own. Because the score for the play at time t depends only on expectations over the players' conditionally independent moves at that time, Y's use of a longer history, from X's perspective, is merely a peculiar kind of random number generator whose use does not affect either player. So Y's switching between a long- and short-memory strategy is completely undetectable (and irrelevant) to X.

The importance of this result is that the player with the shortest memory in effect sets the rules of the game. A player with a good memory-one strategy can force the game to be played, effectively, as memory-one. She cannot be undone by another player's longer-memory strategy.

Appendix B: ZD Strategies Succeed Without Markov Equilibrium. We here prove that Y cannot evade X's ZD strategy by changing his own strategy on a short timescale—even arbitrarily on every move of the game. The point is that Y cannot usefully "keep the game out of Markov equilibrium" or play "inside the Markov equilibration time scale."

For arbitrary $\mathbf{p}$ and $\mathbf{q}$, the Markov matrix is as shown in Fig. 2*A*. We suppose that X is playing a ZD strategy with some fixed $\mathbf{p}_z$ and write $\mathbf{M}(\mathbf{q}) \equiv \mathbf{M}(\mathbf{p}_z,\mathbf{q})$. The key point is that each row of $\mathbf{M}(\mathbf{q})$ is linear in exactly one component of $\mathbf{q}$. Thus, the average of any number of different $\mathbf{M}(\mathbf{q}_i)$'s satisfies

$$\langle \mathbf{M}(\mathbf{q}_i)\rangle_i = \mathbf{M}\big(\langle \mathbf{q}_i\rangle_i\big). \quad [20]$$

Now consider the result of N consecutive plays, $i = 1,2,\ldots,N$, where N is a large number. The game goes through N states α_i, with $\alpha \in \{cc,cd,dc,dd\}$. Comparing times i and $i+1$, the game goes from state α_i to state α_{i+1} by a draw from the four probabilities $M_{\alpha_i\alpha_{i+1}}(q_{i\alpha_i})$, $\alpha_{i+1} \in \{cc,cd,dc,dd\}$, where $q_{i\alpha_i}$ is the α_i th component of $\mathbf{q}_i$ (at time i). So the expected number of times that the game is in state β is

$$\begin{aligned}\langle N_\beta\rangle &= \sum_{i=1}^{N} M_{\alpha_i\beta}(q_{i\alpha_i})\\ &= \sum_{\alpha}\sum_{i|\alpha} M_{\alpha\beta}(q_{i\alpha})\\ &= \sum_{\alpha} N_\alpha \langle M_{\alpha\beta}(q_{i\alpha})\rangle_{i|\alpha}\\ &= \sum_{\alpha} N_\alpha M_{\alpha\beta}\big(\langle q_{i\alpha}\rangle_{i|\alpha}\big)\end{aligned} \quad [21]$$

Here the notation $i|\alpha$ is to be read as "for values of i such that $\alpha_i = \alpha$."

Now taking the (ensemble) expectation of the right-hand side and defining probabilities

$$P_\alpha = \frac{1}{N}\langle N_\alpha\rangle, \quad [22]$$

Eq. **21** becomes

$$P_\beta = \sum_{\alpha} P_\alpha M_{\alpha\beta}\big(\langle q_{i\alpha}\rangle_{i|\alpha}\big). \quad [23]$$

This result shows that a distribution of states identical to those actually observed would be the stationary distribution of Y's playing the fixed strategy $q_\alpha = \langle q_{i\alpha}\rangle_{i|\alpha}$. Because X's ZD strategy is independent of any fixed strategy of Y's, we have shown that, for large N, X's strategy is not spoiled by Y's move-to-move strategy changes.

That the proofs in *Appendix A* and *Appendix B* have a similar flavor is not coincidental; both exemplify situations where Y devises a supposedly intricate strategy that an oblivious X automatically marginalizes over.

ACKNOWLEDGMENTS. We thank Michael Brenner, Joshua Plotkin, Drew Fudenberg, Jeff Hussmann, and Richard Rapp for helpful comments and discussion.

1. Axelrod R, Hamilton WD (1981) The evolution of cooperation. *Science* 211:1390–1396.
2. Roberts K (1985) Cartel behavior and adverse selection. *J Industr Econ* 33:401–413.
3. Axelrod R, Dion D (1988) The further evolution of cooperation. *Science* 242:1385–1390.
4. Nowak M, Sigmund K (1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* 364:56–58.
5. Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314:1560–1563.
6. Kendall G, Yao X, Chong SY (2007) *The Iterated Prisoners' Dilemma 20 Years On* (World Scientific, Singapore).
7. Axelrod R (1984) *The Evolution of Cooperation* (Basic Books, New York).
8. Dawkins R (1988) *The Selfish Gene* (Oxford Univ Press, New York).
9. Poundstone W (1992) *Prisoner's Dilemma* (Doubleday, New York).
10. Hauert Ch, Schuster HG (1997) Effects of increasing the number of players and memory steps in the Iterated Prisoner's Dilemma, a numerical approach. *Proc Biol Sci* 264:513–519.
11. Premack DG, Woodruff G (1978) Does the chimpanzee have a theory of mind? *Behav Brain Sci* 1:515–526.
12. Saxe R, Baron-Cohen S, eds (2007) *Theory of Mind: A Special Issue of Social Neuroscience* (Psychology Press, London).
13. Lurz RW (2011) *Mindreading Animals: The Debate over What Animals Know about Other Minds* (MIT Press, Cambridge, MA).
14. Ewens WJ (1989) An interpretation and proof of the fundamental theorem of natural selection. *Theor Popul Biol* 36:167–180.
15. Güth W, Schmittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. *J Econ Behav Organ* 3:367–388.
16. Nowak MA, Page KM, Sigmund K (2000) Fairness versus reason in the ultimatum game. *Science* 289:1773–1775.