Extortion and cooperation in the Prisoner's Dilemma

**Alexander J. Stewart and Joshua B. Plotkin[1]**
Department of Biology, University of Pennsylvania, Philadelphia, PA 19104

Self-serving, rational agents sometimes cooperate to their mutual benefit. However, when and why cooperation emerges is surprisingly hard to pin down. To address this question, scientists from diverse disciplines have used the Prisoner's Dilemma, a simple two-player game, as a model problem. In PNAS, Press and Dyson (1) dramatically expand our understanding of this classic game by uncovering strategies that provide a unilateral advantage to sentient players pitted against unwitting opponents. By exposing these results, Press and Dyson have fundamentally changed the viewpoint on the Prisoner's Dilemma, opening a range of new possibilities for the study of cooperation.

For nearly a century now game theory has influenced the way we think about the world. It has entered into the study of almost every type of human interaction, including economics, political science, war games, and evolutionary biology. This is because, at its core, game theory seeks to explain how rational players should behave to best serve their own interests. Game theory emerged as a field of active research with the work of John von Neumann on mathematical economics (2, 3). John Nash built on this foundation when he introduced the concept of a Nash Equilibrium (4), which occurs when no player can gain by unilaterally changing strategy. The impact of Nash's work, which began with economics, was soon felt in evolutionary biology, when John Maynard Smith and George Price introduced a variant form of equilibrium, the evolutionary stable strategy (5), which, when dominant in an evolving population, cannot be invaded and replaced by any other strategy.

The Prisoner's Dilemma itself is well established as a way to study the emergence of cooperative behavior. Each player is simultaneously offered two options: to cooperate or defect. If both players cooperate, they each receive the same payoff, R; if both defect, they each receive a lower payoff, P. However, if one player cooperates and the other defects, the defector receives the largest possible payoff, T, and the cooperator the lowest possible payoff, S.

If the Prisoner's Dilemma is played only once, it always pays to defect—even though both players would benefit by both cooperating. Thus, the game seems to offer a grim, and unfamiliar, view of social interactions. If the game is played more
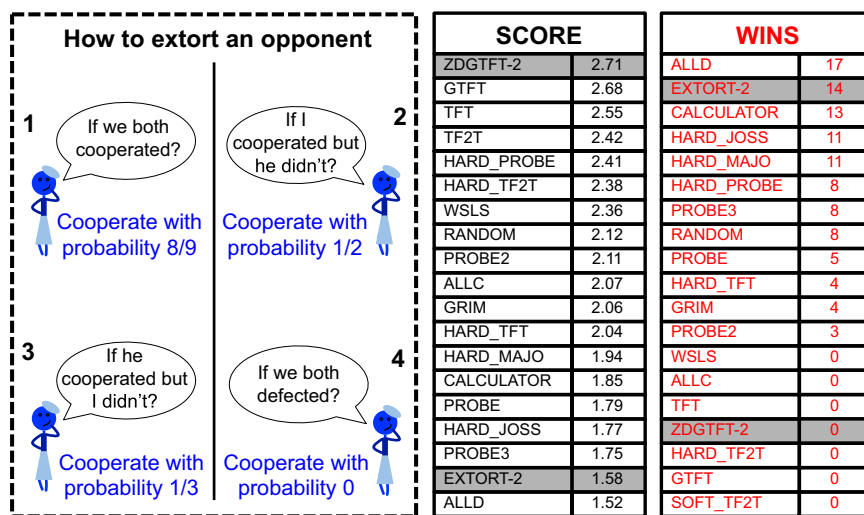
Fig. 1. How to extort your opponent, and what you stand to gain by extortion. A ZD strategy is specified in terms of four probabilities: the chance that a player will cooperate, given the four possibilities for both players' actions in the previous round. *Left*: A specific example, called Extort-2, forces the relationship $S_X - P = 2(S_Y - P)$ between the two players' scores. Extort-2 guarantees player X twice the share of payoffs above P, compared with those received by her opponent Y. A related ZD strategy that we call ZDGTFT-2 forces the relationship $S_X - R = 2(S_Y - R)$ between the players' scores. ZDGTFT-2 is more generous than Extort-2, offering Y a higher portion of payoffs above P. *Right*: We simulated (13) these two strategies in a tournament similar to that of Axelrod (6, 7). ZDGTFT-2 received the highest total payoff, higher even than Tit-For-Tat and Generous-Tit-For-Tat, the traditional winners. Extort-2 received a lower total payoff (because its opponents are not evolving), but it won more head-to-head matches than any other strategy, save Always Defect.

SCORE	
ZDGTFT-2	2.71
GTFT	2.68
TFT	2.55
TF2T	2.42
HARD_PROBE	2.41
HARD_TF2T	2.38
WSLS	2.36
RANDOM	2.12
PROBE2	2.11
ALLC	2.07
GRIM	2.06
HARD_TFT	2.04
HARD_MAJO	1.94
CALCULATOR	1.85
PROBE	1.79
HARD_JOSS	1.77
PROBE3	1.75
EXTORT-2	1.58
ALLD	1.52

WINS	
ALLD	17
EXTORT-2	14
CALCULATOR	13
HARD_JOSS	11
HARD_MAJO	11
HARD_PROBE	8
PROBE3	8
RANDOM	8
PROBE	5
HARD_TFT	4
GRIM	4
PROBE2	3
WSLS	0
ALLC	0
TFT	0
ZDGTFT-2	0
HARD_TF2T	0
GTFT	0
SOFT_TF2T	0

than once, however, other strategies, that reward cooperation and punish defection, can dominate the defectors, especially when played in a spatial context or for an indeterminate number of rounds (6–8). Thus, the Iterated Prisoner's Dilemma (IPD) offers a more hopeful, and more recognizable, view of human behavior. What strategy is best in the IPD, however, is not straightforward. Beginning in the 1950s with the RAND Corporation, right up to today, people have looked to real-life experiments to understand how various strategies perform. The most influential experiments were performed by Robert Axelrod in the early 1980s with his IPD tournaments (6, 7). These established the simple Tit-For-Tat strategy as an extraordinarily successful way to foster cooperation and accumulate a large payoff: a player should cooperate if her opponent cooperated on the last round, and otherwise defect. Nonetheless, no strategy is universally best in such a tournament (8, 9), because a player's performance depends on the strategies of her opponents. This is true even in the simple, one-shot Prisoner's Dilemma, a fact that is now commonly known and

exploited, to entertaining effect, by the British game show Golden Balls (10).

Press and Dyson (1) now illustrate like never before the power granted to a player with a theory of mind (i.e., a player who realizes that her behavior can influence her opponents' strategies). Their article contains a number of remarkable, unique results on the IPD, and indeed, all iterated two-player games. First, they prove that any "long-memory" strategy is equivalent to some "short-memory" strategy, from the perspective of a short-memory player. This means that an opponent who decides his next move by analyzing a long sequence of past encounters might as well play a much simpler strategy that considers only the immediately previous encounter, when playing against a short-memory player. Thus, the possible outcomes of the IPD can be understood by analyzing strategies that remember only the previous round.

Press and Dyson derive a simple formula for the long-term scores of two IPD players, in terms of their one-memory strategies. Their formula naturally suggests a special class of strategies, called zero determinant (ZD) strategies, that enforce a linear relationship between the two players' scores. The existence of such strategies has far-reaching consequences. For example, if a player X is aware of ZD strategies, then she can choose a strategy that determines her opponent Y's long term score, regardless of how Y plays. There is nothing Y can do to improve his score, although his choices may affect X's score. This is similar to the "equalizer" strategies, discussed in the context of folk theorems (11). Such a strategy is essentially just mischievous, because it does not guarantee that X will receive a high payoff, or even that she will outperform Y.

However, ZD strategies offer more than just mischief. Suppose once again that X is aware of ZD strategies, but that Y is an "evolutionary player," who possesses no theory of mind and instead simply seeks to adjust his strategy to maximize his own score in response to whatever X is doing, without trying to alter X's behavior. X can now choose to extort Y. Extortion strategies, whose existence Press and Dyson report, grant a disproportionate number of high payoffs to X at Y's expense (example in Fig. 1). It is in Y's best interest to cooperate with X, because Y is able to increase his score by doing so. However, in so doing, he ends up increasing X's score even more than his own. He will never catch up to her, and he will accede to her extortion because it pays him to do so.

Another possibility is that X alone is aware of ZD strategies, but Y does at least have a theory of mind. X can once again decide to extort Y. However, Y will eventually notice that something is amiss: whenever he adjusts his strategy to improve his own score, he improves X's score even more. With a theory of mind he may then decide to sabotage both his own score and X's score, by defection, in the hopes of altering X's behavior. The IPD has thus reduced to the ultimatum game (12), with

Press and Dyson have fundamentally changed the viewpoint on the Prisoner's Dilemma.

X proposing an unfair ultimatum and Y responding either by acceding or by sabotaging the payoffs for both players.

Finally, if both players are sentient and witting of ZD strategies, then each will initially try to extort the other, resulting in a low payoff for both. The rational thing to do, in this situation, is to negotiate a fair cooperation strategy. Players ignorant of ZD strategies might eventually adopt something like the Tit-For-Tat strategy, which offers each player a high score for cooperation but punishes defection. However, knowledge of ZD strategies offers sentient players an even better option: both can agree to unilaterally set the other's score to an agreed value (presumably the maximum possible). Neither player can then improve his or her score by violating this treaty, and each is punished for any purely malicious violation.

Extortion strategies work best when other players do not realize they are being extorted. Press and Dyson discuss how extortion strategies allow a sentient player to dominate an evolutionary player, who mindlessly updates his strategy to increase his payoff. However, this is not the only use of ZD strategies. Had Press and Dyson kept their results to themselves, they may have enjoyed an advantage in tournaments like those set up by Axelrod (6, 7), in which a variety of fixed strategies compete. To test whether this is true, we reran Axelrod's original tournament, but with the addition of some ZD strategies (Fig. 1). We found ZD strategies that foster cooperation and receive the highest total payoff in the tournament—higher even than Tit-For-Tat's payoff. In addition, we found extortion strategies that win the largest number of the head-to-head competitions in the tournament.

Following Press and Dyson, future research on the IPD will surely be framed in terms of ZD strategies. How do such strategies fare in iterated games with finite but undetermined time horizons, or in the presence of noise, or in a spatial context, etc.? Additionally, what does the existence of ZD strategies mean for evolutionary game theory: can such strategies naturally arise by mutation, invade, and remain dominant in evolving populations?

What is immediately clear is that, by publishing their results, Press and Dyson have changed the game. Readers of PNAS, some of whom have a theory of mind, will now be tempted to use an extortion strategy when faced with an IPD opponent. Yet even in a game as simple as the Prisoner's Dilemma, even knowing that a long-term memory does not provide an advantage and that ZD strategies exist, a sentient player after Press and Dyson is still faced with a kind of Turing test each time she meets an opponent: she must determine what her opponent knows about the game, what he knows about her, and what he might be able to learn, and only then, in the face of these tentative assumptions, can she apply her own understanding to devise a strategy that best serves her interest.

1. Press WH, Dyson FJ (2012) Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proc Natl Acad Sci USA*, 10.1073/pnas.1206569109.
2. von Neumann J (1928) Zur theorie der gesellschaftsspiele. *Mathematische Annalen* 100:295–320.
3. von Neumann J, Morgenstern O (1944) *Theory of Games and Economic Behavior* (Princeton Univ Press, Princeton).
4. Nash JF (1950) Equilibrium points in n-person games. *Proc Natl Acad Sci USA* 36:48–49.
5. Maynard Smith J, Price GR (1973) The logic of animal conflict. *Nature* 246:158.
6. Axelrod R, Hamilton WD (1981) The evolution of cooperation. *Science* 211:1390–1396.
7. Axelrod R (1984) *The Evolution of Cooperation* (Basic Books, New York).
8. Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314:1560–1563.
9. Fudenberg D, Maskin E (1990) Evolution and cooperation in noisy repeated games. *Am Econ Rev* 80:274–279.
10. YouTube.com. Golden balls: the weirdest split or steal ever! Available at: http://www.youtube.com/watch?v=S0qjK3TWZE8. Accessed May 23, 2012.
11. Sigmund K (2010) The Calculus of Selfishness (Princeton University Press, Princeton).
12. Nowak MA, Page KM, Sigmund K (2000) Fairness versus reason in the ultimatum game. *Science* 289:1773–1775.
13. Beaufils B, Dalahey JP, Mathie P (2001) Adaptive behaviour in the classical Iterated Prisoner's Dilemma. *Proceedings of the Artificial Intelligence and the Simulation of Behaviour Symposium on Adaptive Agents and Multi-agent Systems*, eds Kudenko D, Alonso E (The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, York, UK), pp 65–72.