# Culture, Cooperation, and Repeated Games*

E. Maskin**
Institute for Advanced Study

October 2010

* Prepared for a volume of papers from the Samuel P. Huntington Symposium, Higher School of Economics, Moscow, May 2010.
** I thank NSF for research support.

A *social norm* can be thought of as a behavioral rule together with the adverse consequences that individuals face if they don't adhere to that rule. Such consequences are sometimes external to individuals (e.g., ostracism) and sometimes internal (e.g., a stricken conscience). From this perspective, a society's *culture* can be viewed as the collection of social norms in force in that society.

One behavioral rule of critical importance to successful cultures is the precept that individuals should behave *cooperatively*, i.e., that they should exert effort on behalf of some collective good. But cooperative behavior is often tricky to sustain through social norms. This is because cooperation by others presents an individual with the incentive to "free ride": if everyone else is exerting effort, I may well be better off sitting back and enjoying the collective good they produce without lifting a finger myself.

The tension between cooperation and free-riding is encapsulated by recasting the Prisoners' Dilemma, the famous parable from game theory. Instead of "confessing" or "not confessing"—the two prisoners' options in the classic formulation—let us suppose the two individuals can each choose to "cooperate" (i.e., to exert effort) at a personal cost of 4, or to defect (i.e., to sit back), which costs nothing. If both individuals (whom I will call "players") cooperate, then each gets a reward of 6, which given the cost of effort leads to a net payoff of 2 each. However, if just one player cooperates, then each player gets a gross payoff of only 3. Thus the cooperator's net payoff in that case is $3-4=-1$, whereas the defector enjoys a payoff of $3-0=3$. If neither player cooperates, they each get a payoff of 0. The game is summarized in Table 1, where one player (player 1) chooses between rows, the other player (player 2) chooses between columns, and the numbers in the entries are the payoffs corresponding to their choices (the first number in each entry is player 1's net payoff, and the second is 2's payoff).

Notice that if player 2 cooperates (chooses *C*), then player 1 does better by defecting (a payoff of 3 rather than 2). Moreover, if player 2 defects (chooses *D*), then again player 1 does better by defecting (a payoff of 0 rather than $-1$). That is, defecting is a *dominant strategy* for player 1, as it is for player 2 (since the game is symmetric). The Prisoners' Dilemma thus illustrates the conundrum that cooperation presents: the players are *individually* induced to defect and get a payoff of 0 each, but *jointly* they would be better off cooperating and getting a payoff of 2 each.

| | *C* | *D* |
|---|---|---|
| **Cooperate** | 2, 2 | −1, 3 |
| **Defect** | 3, −1 | 0, 0 |

Table 1

Repeating the Prisoners' Dilemma can solve this conundrum. Let us imagine that the game is played *many* times, instead of just once, and that each player is interested in maximizing her long-run average payoff per iteration. Suppose that player 2 adheres to the behavioral rule in which he plays *C* until either player plays *D*, and thereafter plays *D;* I will call this rule "conditionally cooperate." Then, it is in player 1's interest to conditionally cooperate too: if she does so, then players will, in fact, end up cooperating in every iteration and her long-run average payoff will be 2 per iteration. If, by contrast, she defects at some point, then she'll get a payoff of 3 that time (given that 2 plays *C*). But in all subsequent iterations, player 2 will defect, and so player 1 can then get at most 0 per iteration herself. In other words, the payoff she gets per iteration after a single defection will be approximately 0 (since the average of 3 followed by an indefinite string of 0s is just slightly more than 0).

This argument shows that repetition makes conditionally cooperative behavior (CC) a sustainable social norm: as long as there have been no previous defections, each player will cooperate for fear that if she does not, the other player won't continue to cooperate in future iterations.

Unfortunately, CC does not constitute the *only* sustainable social norm in the repeated Prisoners' Dilemma. "Always play *D*" (that is, "play *D* in every iteration regardless of what happened up to that point"), for example, is also a possibility: if player 2 is never going to cooperate, then player 1 does best to refrain from cooperating too, and vice versa. In fact, the Folk Theorem for Repeated Games (see Fudenberg and Maskin 1986) shows that essentially all outcomes between total cooperation (achievable through CC) and total defection (achievable through "Always play *D*") are sustainable through social norms. Thus, the theory makes no sharp predictions about how much cooperation we can expect to see in a repeated game.

Given this frustrating result, one might hope that evolution could come to the rescue, i.e., that evolutionary forces (biological or cultural) might "root out" uncooperative behavior. The rough idea behind this hope is that a flexible strategy, in which one cooperates when playing against a cooperative player but defects when playing against an uncooperative player seems, on the face of it, to be a more profitable strategy than unremittingly uncooperative behavior. If so, then in an evolutionary setting, the flexible strategy should "reproduce," and uncooperative behavior should be driven out. And, in the long run, cooperation would predominate.

Indeed, Axelrod (1984) developed this notion in a popular book over a quarter of a century ago (see also Axelrod and Hamilton 1981). Here, in part, is how the argument works. Imagine that we start with a population of players who use "Always play *D*" (AD), that is, they play *D* in every iteration. Let us introduce into that population a small group of "mutants" that play CC. Now, choose pairs of players at random from the population to play the repeated Prisoners' Dilemma, and adjust the population over time so that a strategy that does well on average in the repeated game grows as a proportion of the population, and a strategy that has not done so well shrinks proportionately. Notice that CC gets the same long-run average payoff per iteration (namely, 0) against AD as AD gets against AD. Furthermore, CC gets a payoff of 2 against itself, while AD gets only 0 against CC. So, CC performs better in expectation than AD against the population consisting of a large group of ADs and a small group of mutant CCs. Thus, we can expect that the cooperative strategy CC will *grow*, i.e., replicate itself and that the uncooperative strategy AD will *die out*, establishing—as the rough idea suggests—that uncooperative behavior is indeed weeded out by evolution.

This looks encouraging. But consider the strategy ALT, defined so that it starts with *C* and then alternates between *D* and *C* unless someone breaks the alternating pattern, in which case it thereafter plays *D*. This strategy against itself gets an long-run average payoff of 1 per iteration (in half the iterations players both select *C*, with a payoff of 2; in the other half, they select *D*, with a payoff of 0). Thus, ALT is quite uncooperative. Nevertheless, a population consisting mostly of ALT cannot be weeded out by a mutant strategy in the way AD was weeded out by CC. To see this, observe that if a mutant fails at some point to alternate between *C* and *D* the way ALT does, it will be severely punished by ALT (i.e., ALT will play *D* forever after). Thus, because ALT predominates in the population, such a mutant will not perform well in expectation against the overall population. Only a strategy that alternates like ALT can possibly do well against that population. But then such a strategy must essentially *be* ALT itself. In other words, evolution does not root ALT out.

Nevertheless, there is an important sense in which ALT is an overly brittle strategy: it relies critically on adhering perfectly to the alternating pattern; any failure to go back and forth between *C* and *D* as prescribed is punished harshly. This suggests that ALT may not survive as a successful strategy in a richer setting where strategies are sometimes misperceived by others.

To be more precise, suppose that we now introduce a small but positive probability *p* that, at some point in a repeated game, a player will misperceive his opponent's action and think that she has played *D* when she has really played *C* or vice versa. In this more elaborate setting, I claim that ALT can be weeded out by a mutant strategy $s'$ that is identical to ALT except, that if the alternating pattern is broken, $s'$ does not simply play *D* thereafter. Instead, after such a break, $s'$ plays *C* in the next iteration (in effect, signaling its "willingness" to cooperate). If the other player plays *C* in this iteration too, then $s'$ continues to play *C* thereafter. Otherwise, $s'$ plays *D* forever after.

Now, $s'$ does as well against ALT (payoff of 1 before the alternating pattern is broken; payoff of 0 afterwards) as ALT does against ALT. Moreover, $s'$ gets a payoff of 2 against $s'$ (after a break in the alternating pattern), whereas ALT gets only 0 against $s'$ after such a break. Furthermore, because *p* is positive, an alternating-pattern break will happen with positive probability. Thus, $s'$ earns a strictly higher expected payoff per iteration than ALT does against a population consisting of a large group of ALTs and a small group of $s'$s. So $s'$ will indeed weed out ALT.

This analysis generalizes to other noncooperative strategies besides ALT (see Fudenberg and Maskin 1990) and indicates that there are circumstances in which evolution indeed pushes a culture toward cooperation. Interestingly, however, the argument relies not just on mutations in behavior but on small but positive probabilities of misperceived behavior.

This suggests the intriguing possibility that a society may be more likely to evolve cooperation when it is open to at least some immigration. From the standpoint of theory, immigrants provide not only a ready source of "mutations," but also generate the misperceptions needed to permit a strategy like $s'$ to gain an advantage over the predominant behavior in the population. It would be interesting to consider ways in which this possibility might be tested empirically.

*References*


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

Axelrod, R. and W. Hamilton (1981), “The Evolution of Cooperation,” *Science*, 211: 1390-96.

Fudenberg, D. and E. Maskin (1986), “The Folk Theorem in Repeated Games with Discounting or Incomplete Information,” *Econometrica*, 54, 533-554.

Fudenberg, D. and E. Maskin (1990), “Evolution and Cooperation in Noisy Repeated Games,” *American Economic Review*, 80, 274-270.