An Evolutionary Resolution to the Finitely Repeated Prisoner's Dilemma Paradox

Daniel B. Neill

July 2003
CMU-CS-03-155

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Argument by backward induction forces us to conclude that two "rational" players will defect on every turn of the Finitely Repeated Prisoner's Dilemma (FRPD) game, thus performing significantly worse than agents with imperfect rationality. When this game is treated from an evolutionary perspective, using the standard evolutionary model, we encounter a similar paradox: a population which cooperates through turn k can be invaded by a strategy which cooperates through turn $k-1$, and this process continues until the population is dominated by defectors. However, though the strategy of continual defection is evolutionarily stable, it is inferior to nearly all other FRPD strategies: a bistable equilibrium occurs, in which a very small proportion of the other strategy can take over the population. Thus we propose and defend an alternative evolutionary model, a random invasion model in which "evolutionary dominance" is used instead of Maynard Smith's invasion criteria. This model combines the Lamarckian spread of ideas through a population with Darwinian natural selection of the organisms adopting those ideas, and thus is a more reasonable model of *communicating* populations. When the new evolutionary model is applied to the Finitely Repeated Prisoner's Dilemma, we find that defectors dominate the population for very short FRPD games, but as game length increases, it becomes more and more certain that successful strategies will cooperate until near the end of the game. Defining rationality based on evolutionary fitness (or fictitious evolutionary play) using this model, we achieve a resolution to the Finitely Repeated Prisoner's Dilemma paradox. Additionally, the model can be generalized and applied to many other decision situations, and thus it serves as a possible standard for rational decision-making under uncertainty.

Keywords: game theory, evolutionary games, Finitely Repeated Prisoner's Dilemma

1 Introduction

The Prisoner's Dilemma is a widely used game-theoretic model of interactions between individuals with partially conflicting goals. Each player must choose between two options: *cooperation* for mutual benefit, and *defection* for individual benefit (harming the other player). The payoffs to each player are determined by the choices of both players; these possibilities are given in Table 1.

If both players cooperate, each receives a high payoff R as a *reward* for mutual cooperation. If both defect, each receives a low payoff P as a *punishment* for mutual defection. If one player defects while the other cooperates, the defector receives a very high payoff T as a *temptation* to defect, and the cooperator receives a very low *sucker* payoff S. The Prisoner's Dilemma is defined by $T > R > P > S$ and $2R > T + S$. Despite the fact that mutual cooperation is preferred to mutual defection, each player scores higher if he defects regardless of the opponent's choice, and hence mutual defection is the only "rational" outcome of the single-shot Prisoner's Dilemma game.

In an "Iterated Prisoner's Dilemma" (IPD), a Prisoner's Dilemma interaction is repeated between the same two players over a number of rounds. Based on past results, reciprocal altruism can develop, enabling mutual cooperation to become a rational option. The IPD game has sparked various theories of the evolution of cooperation based on reciprocity (Axelrod, 1984), and these models have been applied to fields ranging from economics to biology (Maynard Smith, 1982; Axelrod, 1984; Milinski, 1987).

1.1 The Finitely Repeated Prisoner's Dilemma paradox

The Finitely Repeated Prisoner's Dilemma (FRPD) is a variant of the Iterated Prisoner's Dilemma in which the number of rounds N is given and known in advance by both players. We would expect that two rational strategies would be able to achieve cooperation in the FRPD; however, under the assumption of common knowledge of rationality, we can use an argument of "backward induction" to show that each player defects continually. The argument proceeds as follows: on round N, both players will defect, as in the one-shot Prisoner's Dilemma. Then on round $N - 1$, both players know that, whatever they do now, both will defect on the next round in any case. Since the player's move on round $N - 1$ will not influence the opponent to reciprocate cooperation on future moves, it is rational to defect on round $N - 1$. This argument can be applied to round $N - 2$ and each preceding move, in turn, and we reach the conclusion that two rational players will defect on every round. But human players tend to treat the FRPD as a Prisoner's Dilemma interaction of indeterminate length until nearly the end of the game: mutual cooperation is achieved until the final few rounds. Thus we have a "paradox of rational choice" in which perfectly rational agents perform significantly worse than agents with imperfect rationality. In other words, in the FRPD, it is not necessarily rational to be "rational."

Thus the Finitely Repeated Prisoner's Dilemma game presents a serious challenge to the notion of game-theoretic equilibrium solutions as a theory of rationality in repeated games. Various researchers have attempted to resolve this paradox, but none of the proposed solutions are widely accepted. These include weakening the assumption of "common knowledge

		P2	
		C	D
P1	C	R/R	S/T
	D	T/S	P/P

Table 1: Prisoner's Dilemma Payoffs to P1/P2

of rationality" to "common belief of rationality" (Kreps et al, 1982; Binmore, 1987). In this interpretation, the equilibrium solution is defined as "rational play," but a player's certainty of his opponent's rationality is weakened to "belief with probability 1" (Binmore) or "belief with probability $1-\epsilon$" (Kreps et al). Though Kreps et al have shown that cooperation can result in the FRPD if both players assume some non-zero probability of non-equilibrium play, this assumption of "uncertain belief" is not a reasonable definition of "common knowledge" and so does not resolve the original paradox. Moreover, defining rationality as the equilibrium solution, but allowing moves out of equilibrium, leads to a "self-defeating" concept of rationality in which it is better to make the irrational move; thus the common belief in "rational play" under this definition would not be a reasonable assumption.

1.2 The evolutionary FRPD paradox

A more promising solution to the Finitely Repeated Prisoner's Dilemma paradox lies in the concept of "evolutionary games." In these models, evolution is driven by natural selection: strategies which earn higher average payoffs have higher "evolutionary fitness," and are able to survive and reproduce, while less successful strategies die off. If we assume that the main goal of a decision-making agent is to achieve long-term survival in the evolutionary sense, we can consider "rational" behavior to be following the strategy which achieves the greatest evolutionary success.

However, the use of evolutionary games does not immediately resolve the FRPD paradox: using standard evolutionary models, we find that continual defection is the most evolutionarily successful strategy. In fact, we observe an evolutionary equivalent of the backward induction argument. First assume that we have a population of continual cooperators; this strategy can be invaded, and taken over, by a strategy that cooperates until move $N-1$ and defects on move N. The strategy that cooperates until move $N-1$, in turn, can be taken over by a strategy that cooperates until move $N-2$, and defects thereafter. This process continues, with each strategy being invaded by a strategy that starts defecting one round earlier, until the population consists entirely of continual defectors. Continual defection is an *evolutionarily stable strategy* (ESS), a strategy which cannot be invaded by any other strategy, and thus the evolutionary process ends with a population of defectors. In order to resolve the paradox, then, we must examine the FRPD game, and our evolutionary model, in greater detail.

2 The Finitely Repeated Prisoner's Dilemma game

A Finitely Repeated Prisoner's Dilemma game is defined by the 4-tuple of payoffs (T,R,P,S) and the number of rounds N. We will consider payoff tables with $P = 0$, $S = -1$, and $T = 1 + R$. These can be thought of as describing the following situation: on each round, each player independently decides whether to give the other player a "gift" costing him one point and awarding the other player T points, $T > 1$. We will initially assume $T = 3$, but will also examine other values of T.[1]

A *strategy* for the FRPD is a method of deciding whether to cooperate or defect on any given turn, based on the results of previous turns that game. To simplify our calculations, and present the paradox in its most common and essential form, we assume that once cooperation has failed (i.e. either player has defected), it cannot be resumed, and the rest of the game will consist of mutual defection. Since we have assumed $P = 0$, this is equivalent to an FPRD game that stops upon either player's defection, similar to the "centipede" game. Alternatively, we can think of this as a standard no-noise FRPD, with all strategies chosen from the strategy space C_i ($0 \leq i \leq N$). A strategy C_i will cooperate on turn k if and only if both of the following are true: $k \leq i$, and its opponent has cooperated on turns 1 through $k - 1$. Thus C_0 always defects, and C_N always reciprocates cooperation with cooperation.

We now define $W(X|Y)$ as the total payoff to strategy X in an N-round FRPD game against strategy Y: $W(X|Y) = \sum_{i=1}^{N} p(i)$, where $p(i)$ is the payoff received by X after turn i. Then we can compute $W(C_i|C_j)$ for all i and j. There are three cases. First, if $i = j$, both strategies will cooperate until round i and defect thereafter, receiving total payoff iR. Second, if $i > j$, strategy C_i receives j reward payoffs, followed by a sucker payoff. Thus its total payoff is $jR + S$. Third, if $i < j$, strategy C_i receives i reward payoffs, followed by a temptation payoff. Thus its total payoff is $iR + T$. We can summarize this as $W(C_i|C_j) = R \min(i, j) + T\delta(i < j) + S\delta(i > j)$, where $\delta = 1$ if the expression is true and $\delta = 0$ if false.

3 The standard evolutionary model

We begin by considering one well-known model for evolutionary games: the invasion model proposed by Nowak & Sigmund (1992, 1993). This model assumes that a large homogeneous population of a given strategy X is visited periodically by mutant strains. We first select the initial strategy X randomly (assuming some, possibly uniform, prior distribution) from a given space of strategies. Then we repeat the following steps: a mutant strain Y is selected randomly (using the prior distribution) from our space of strategies, and the mutant strain either takes over the population (replacing strategy X) or dies off. To compute which of these outcomes will occur, we use the invasion criteria of Maynard Smith (1982): strategy Y *invades* strategy X if $W(Y|X) > W(X|X)$, or $W(Y|X) = W(X|X)$ and $W(Y|Y) >$

[1] This statement of the problem makes it clear why defection is the only rational (and reasonable) alternative in a one-shot Prisoner's Dilemma game. Both players choose independently whether or not to give a gift, and giving a gift always harms the gift-giver, so it makes sense for a self-interested individual to both behave rationally (not give a gift) and hope that the opponent behaves irrationally.

$W(X|Y)$. If Y invades X according to the Maynard Smith criteria, then Y will take over the population; otherwise the initial population of X will continue.

To apply this to the Finitely Repeated Prisoner's Dilemma, we can use the equation for $W(C_i|C_j)$ derived in the previous section to determine when strategy C_i can be invaded by C_j, $i \neq j$. C_j can invade if $W(C_j|C_i) > W(C_i|C_i)$, or $W(C_j|C_i) = W(C_i|C_i)$ and $W(C_j|C_j) > W(C_i|C_j)$. First, we know $W(C_i|C_i) = iR$. Then if $j > i$, we know $W(C_j|C_i) = iR+S < W(C_i|C_i)$, so C_j cannot invade. If $j < i$, we know $W(C_j|C_i) = jR+T$. In this case, C_j can invade if $jR+T \geq iR$; solving for j, we obtain $j \geq i - \frac{T}{R}$. Since $\frac{T}{R} > 1$, we know that C_i can always be invaded by C_{i-1}. It may also be possible for C_i to be invaded by strategies that defect even earlier, that is, by C_{i-k} for $k > 1$. Assuming $T = R + 1$ as stated above, this will occur for low values of the reward payoff R: C_i can be invaded by C_{i-k} if $R \leq \frac{1}{k-1}$. In any case, the result of this model is clear: the number of rounds of cooperation decreases monotonically, until C_0 (the strategy that defects continually) dominates the population. Since no other strategy can invade C_0, the evolutionary process halts at this point.

In fact, this result can be extended to the space of all Finitely Repeated Prisoner's Dilemma strategies. For any strategy $S = S_N$, let S_i denote the strategy which follows strategy S through turn i, and defects for the remainder of the game. Let k be the last turn on which S_i has some non-zero probability of cooperating against its clone; it is clear that $k \leq i$. Then if $k > 0$, we know $W(S_{k-1}|S_i) > W(S_i|S_i)$, so S_{k-1} invades S_i. This process continues, with each strategy being taken over by a strategy that starts defecting at least one round earlier, until the population is taken over by a strategy with $k = 0$: at this point the population consists entirely of continual defectors. Since the strategy of continual defection is evolutionarily stable, the evolutionary process halts at this point, leaving us with a population of defectors.

4 Improving the evolutionary model

Maynard Smith's invasion criteria, and the related notion of evolutionary stability, serve as the foundations for much of the field of evolutionary game theory. However, these criteria make several assumptions which may be unreasonable when modeling some types of real-world interactions. First, the Maynard Smith criteria only deal with a pair of strategies at a time, and assume that one of the two strategies "wins" (and takes over the population) before any more invasion occurs. In real interactions, however, multiple strategies can coexist in a stable equilibrium in which neither strategy dominates. Even if stable equilibria do not occur, more complicated evolutionary interactions can result when multiple strategies attempt to invade simultaneously: for example, Boyd & Lorberbaum (1987) have shown that an evolutionarily stable strategy (though impervious to invasion by any single strategy) can be taken over by the combined action of multiple strategies. Thus the Maynard Smith criteria are an oversimplification of evolutionary interactions; however, this oversimplification is necessary because the complexity of the model is dramatically increased by allowing multi-strategy interactions. In some cases, these interactions may lead to cyclical population patterns, and the result of the model may vary significantly depending on the amount of time between successive invasions. Since we are attempting to approximate the evolutionary success of strategies (in order to determine which strategy is "best" and hence rational to play), rather than precisely forecasting how strategies will evolve, it seems that the extra realism of multiple strategy interaction is not worth the resulting increase in model complexity.

A more soluble difficulty, and one more immediately relevant to the Finitely Repeated Prisoner's Dilemma, is that Maynard Smith's invasion criteria assume that the population of the invading strategy is vanishingly small compared to the population being invaded. This assumption would be accurate if the population was invaded by only a single invader at a time (resulting either from migration or mutation), and this invader had no others of its kind to interact with. However, we can think of many situations where this assumption does not hold, for example migration by multiple invaders, or multiple simultaneous mutations. This assumption would also be broken occasionally by the essential randomness of the evolutionary process: a single individual with below-average evolutionary fitness might survive long enough to reproduce, and invader-invader interactions could occur between its offspring, increasing the evolutionary fitness of the offspring enough for them to invade. Another factor, which we will discuss in more detail below, is the possibility of communication: a mutant individual may convince members of the population being invaded to adopt its behavior patterns, or simply be imitated by members of the population.

The assumption of a single invader lies at the heart of the evolutionary FRPD paradox. To see this, we must first examine the Maynard Smith criteria in greater detail. As stated above, strategy Y invades strategy X if $W(Y|X) > W(X|X)$, or $W(Y|X) = W(X|X)$ and $W(Y|Y) > W(X|Y)$. We write $Y > X$ if Y invades X, and $Y \not> X$ otherwise. Since $Y > X$ is primarily dependent on $W(Y|X)$ and $W(X|X)$, and $X > Y$ is primarily dependent on $W(X|Y)$ and $W(Y|Y)$, it is clear that $X > Y$ and $Y > X$ may be true or false independently of each other. This leads to four possibilities:

I. $X > Y$ and $Y \not> X$. In this case, any initial proportion of strategy X can take over, and completely wipe out, strategy Y. We say that X *dominates* Y, and write $X \gg Y$.

II. $X \not> Y$ and $Y > X$. In this case, any initial proportion of strategy Y can take over, and completely wipe out, strategy X. We say that Y *dominates* X, and write $Y \gg X$.

III. $X > Y$ and $Y > X$. In this case, no matter what the initial proportions of strategies X and Y, the two strategies reach a stable equilibrium where the proportion of strategy X is given by:

$$p = \frac{W(X|Y) - W(Y|Y)}{W(X|Y) - W(Y|Y) + W(Y|X) - W(X|X)}$$

In this case, we write $X \overset{p}{\leftrightarrow} Y$.

IV. $X \not> Y$ and $Y \not> X$. This is a bistable equilibrium, in which either strategy X or strategy Y will take over the population depending on the initial proportions of the two strategies. X will take over if its initial proportion is higher than:

$$m = \frac{W(Y|Y) - W(X|Y)}{W(X|X) - W(Y|X) + W(Y|Y) - W(X|Y)}$$

In this case we write $X \overset{m}{\Leftrightarrow} Y$.

For the FRPD, we have noted that C_0, the strategy which always defects, is evolutionarily stable. Thus $C_i \not\succ C_0$ for all $i > 0$. However, assuming $R > 1$, it is also true that $C_0 \not\succ C_i$ for all $i > 1$. Thus C_0 is in a bistable equilibrium with C_i, for $i > 1$. We can use the above equation for m to compute the proportion of C_i needed to take over the population. Assuming $i > 1$, we know $W(C_0|C_0) = 0$, $W(C_i|C_i) = iR$, $W(C_0|C_i) = T$, and $W(C_i|C_0) = S$. Then $m = \frac{-S}{iR-T-S} = \frac{1}{(i-1)R}$. For large i, C_i can take over a population of C_0 even if it only makes up a very small fraction of the population. For example, assuming $R = 2$ and $i = 1000$, C_i will take over as long as its proportion is at least $\frac{1}{1998}$ of the population. Similarly, for $i \gg j$, C_i and C_j exist in bistable equilibrium, but C_i will take over even if it makes up only a very small fraction of the population: $m = \frac{1}{(i-j-1)R}$.

From this result, it is clear that the Finitely Repeated Prisoner's Dilemma is not adequately described by the standard evolutionary model. "Evolutionary backward induction," where C_i is taken over by C_{i-1}, can occur in the evolutionary FRPD. However, it should also be possible for C_i to be taken over by C_j where $j \gg i$, as long as our model abandons the unreasonable assumption of a single invader. In the following section, we propose an alternative evolutionary model, based on the dominance criterion of Neill (2001), which allows both of these two types of "invasion" to occur.

5 An alternative evolutionary model

Our evolutionary model is identical in form to the Nowak-Sigmund invasion model discussed above. First, we randomly select the initial strategy X. Then we repeat the following steps: a mutant strain Y is selected randomly, and we compute whether the mutant strain takes over the population (replacing strategy X) or dies off.

How do we use this model to calculate the evolutionary success of a strategy, and hence, whether that strategy is "rational" in an evolutionary sense? One option is to simulate the model for a large number of time steps T, and count the number of time steps t_i that each strategy C_i dominates the population. Then the evolutionary performance of C_i is approximated by the proportion of turns dominating the population, $\frac{t_i}{T}$.

A better alternative is to treat the process as a Markov chain. We compute in advance the probability of transitioning from each strategy to every other strategy on a given turn; of course, this is only possible if the number of strategies is finite, and only feasible if that number is relatively small. The probability P_{ij} of transitioning from C_i to C_j is the product of two probabilities: the prior probability p_j that strategy j will attempt to invade, and the probability q_{ij} that the invasion attempt will succeed. For a uniform prior, $p_j = 1/M$ for all strategies, where M is the total number of strategies. For the Nowak-Sigmund invasion model, $q_{ij} = 1$ if j invades i, and $q_{ij} = 0$ otherwise. Once we have computed P_{ij} for each i and j, we can compute the stationary distribution of the Markov chain by solving the matrix equation $Px = x$. Then x_i is the exact proportion of turns we would expect the population to be dominated by each strategy C_i, and hence, an exact measure of each strategy's evolutionary performance.

Now that we have outlined our model and the performance measure derived from it, we consider how our model will compute q_{ij}, the probability that strategy C_j succeeds in invading strategy C_i. As discussed above, the Nowak-Sigmund model relies totally on the Maynard Smith invasion criteria, which assume that the invading population is small and hence the invader's self-payoff $W(C_j|C_j)$ is nearly irrelevant. This is a poor assumption to make for the FRPD, since a high self-payoff by the invading strategy may allow cooperators to take over defectors with only a small proportion of the population initially cooperating. Thus, rather than using the Maynard Smith criteria, we use the "evolutionary dominance" criterion of Neill (2001), which takes the invader's self-payoff into account. Not surprisingly, this criterion is based on the four possible outcomes (X dominates Y, Y dominates X, stable equilibrium, bistable equilibrium) defined above.

5.1 The dominance criterion

Neill (2001) presents the dominance score $\text{dom}(X|Y)$ as a measure of the evolutionary performance of strategy X against strategy Y. The dominance score of X against Y is 1 if X dominates Y, 0 if Y dominates X, and between 0 and 1 otherwise. If X and Y exist in stable equilibrium, the dominance score of X against Y is the proportion of strategy X in the equilibrium population. If X and Y have a bistable equilibrium, the dominance score of X against Y is the proportion of strategy Y needed to take over the population. Thus the dominance score is defined in terms of the relationships and quantities above:

$$\text{dom}(X|Y) = \begin{cases} 1 & \text{if } X \gg Y \\ 0 & \text{if } Y \gg X \\ p & \text{if } X \overset{p}{\Leftrightarrow} Y \\ 1-m & \text{if } X \overset{m}{\Leftrightarrow} Y \end{cases}$$

In all cases, higher dominance scores $\text{dom}(X|Y)$ correspond to a better evolutionary performance of X against Y. Also, we note that the dominance score presents a "constant-sum" measure of evolutionary performance: $\text{dom}(X|Y) + \text{dom}(Y|X) = 1$. If X performs well against Y, then Y performs poorly against X, and vice-versa. Neill (2001) argues that this constant-sum property makes the dominance score a useful measure of performance: we would expect successful strategies to have high dominance scores against most other strategies, regardless of whether or not those strategies are cooperative. Thus, assuming a given space of possible strategies to choose from, each strategy's performance is measured by its mean dominance score against all strategies in that space, and the strategy with the highest dominance score is chosen. Moreover, the dominance score has the essential property that will be needed to resolve the Finitely Repeated Prisoner's Dilemma: it does not make the assumption of a single invader, and thus takes the self-payoff of the invading strategy into account.

However, there are several serious problems with relying on the average dominance score as a measure of evolutionary performance. First, it is not reasonable to assume that performance against every strategy is equally important; it is more important to perform well against successful strategies, since these are the strategies that will survive (and hence, interact with other strategies). Second, a strategy's mean dominance score does not directly correspond to the strategy's performance in any specific evolutionary model. Nevertheless, as we will demonstrate below, use of the dominance score as a transition probability has a clear interpretation as part of an evolutionary invasion model. Thus the dominance score, though not sufficient to measure the evolutionary performance of a strategy, will be a useful component of an evolutionary model which allows us to better accomplish this task.

5.2 Putting it all together

We have now presented all of the necessary components of our evolutionary model: an invasion model based on the Nowak-Sigmund model, and the dominance score which will be used as the probability of transitioning from one strategy to another. Thus the model operates as follows: on each time step, an invading strategy C_j is selected at random to invade the current population of strategy C_i. Then with probability $\mathrm{dom}(C_j|C_i)$, C_j takes over the population, replacing C_i; with probability $\mathrm{dom}(C_i|C_j) = 1 - \mathrm{dom}(C_j|C_i)$, the population of C_i remains. As discussed above, we can use Markov chains to compute the proportion of time that each strategy dominates the population. To do so, we find the stationary distribution of the Markov chain represented by the transition matrix P, where $P_{ij} = \mathrm{dom}(C_j|C_i)p_j$ for $j \neq i$. Assuming a uniform prior over M strategies, this reduces to $P_{ij} = \frac{\mathrm{dom}(C_j|C_i)}{M}$.

Note that for a transition matrix, each row must add to one, and hence $P_{ii} = 1 - \sum_{j \neq i} P_{ij}$. Expanding this expression, and simplifying, we get:

$$P_{ii} = 1 - \sum_{j \neq i} \frac{\mathrm{dom}(C_j|C_i)}{M} = \frac{1 + \sum_{j \neq i} \mathrm{dom}(C_i|C_j)}{M} = \frac{\frac{1}{2} + \sum_j \mathrm{dom}(C_i|C_j)}{M}$$

From this, we can see that the self-transition probability P_{ii} is related to the average dominance score of C_i by $P_{ii} = \mathrm{dom}(C_i) + \frac{1}{2M}$, where the extra $\frac{1}{2M}$ results from the probability of a strategy being randomly selected to "invade" itself. (In this case, the population of C_i remains whether or not the "invasion" is successful.) Thus the average dominance score, though not itself sufficient as a performance measure, does correspond to an important quantity in our model.

These definitions raise an obvious and important question: why is evolutionary dominance better than evolutionary invasion for determining which of the two strategies takes over the population? Note that the dominance score agrees with the Maynard Smith criteria for two of the four possible evolutionary outcomes: if C_j dominates C_i, C_j will take over C_i, and if C_i dominates C_j, C_j will not take over C_i. The differences occur when two strategies are in stable or bistable equilibrium. In stable equilibrium, the Maynard Smith criteria assume that the invasion is successful with probability 1, while the dominance score assumes that the invasion is successful with probability according to the equilibrium proportion of the invader in the population. Obviously, neither of these two models is completely realistic: in real-world interactions the two strategies would coexist, and on the next time step a third strategy would attempt to invade the mixture of the two. As discussed above, allowing this sort of multi-strategy interaction would dramatically increase the complexity of the model, making it both unpredictable and computationally infeasible. Assuming that only one strategy is allowed to dominate the population at a time, the dominance score corresponds to the most reasonable solution: randomly picking a single organism from the combined population at stable equilibrium, and carrying that organism's strategy on to the next time step.

The dominance score also differs from Maynard Smith's criteria in its treatment of bistable equilibria. Recall that in a bistable equilibrium $C_j \overset{m}{\Leftrightarrow} C_i$, C_j takes over a population of C_i if its initial proportion is greater than m. The Maynard Smith criteria assume a single invader: the proportion of the invader C_j in the population is negligible, so C_j cannot take over C_i. The dominance score, on the other hand, assumes that C_j takes over C_i with probability $1-m$. This assumption is equivalent to assuming that the initial proportion of C_j in the population is chosen randomly from a uniform probability distribution on (0,1). This does not make immediate sense for invasion by either mutation or migration: the number of mutants is likely to be small (though not necessarily negligible) compared to the underlying population, and the size of a migratory population is likely to be smaller than the size of the population it is migrating into.

However, the assumption makes much more sense when dealing with a *communicating* population. It is well known that the spread of ideas through a communicating population can occur at much shorter time scales than those required for natural selection. Thus our model views each evolutionary time step as a two-stage process: the "Lamarckian" spread of the invading strategy through the population, followed by Darwinian natural selection.

In this model, a single mutant conceives of an "idea" which is contrary to the established idea and thus prescribes a different pattern of behavior (i.e. a different strategy for the real-world interaction described by our model). This idea is then spread (by direct communication or by imitation of behavior) to various members of the population, who in turn spread the idea to others, until the idea has been distributed through the entire population. Some proportion of the population will adopt the invading idea, changing their behavior as a result, and pass the changed behavior on to their offspring by teaching or imitation. Since the learned behavior is passed directly to the offspring, this part of the evolutionary process is Lamarckian from the organism's perspective, though it can also be viewed as the natural selection of "fitter" ideas. In any case, we assume that the proportion of the population initially adopting the invading idea is independent of the actual effect of adopting that idea on the organism's evolutionary fitness; the idea may be attractive to a large segment of the population, or only a small minority, regardless of its long-term benefits or harms.[2] Since we do not have knowledge of the multitude of other factors which may make a given strategy attractive or unattractive, it makes sense to assume that a random proportion of the population initially adopts the invading strategy. We can then apply Darwinian natural selection to compute which strategy has higher evolutionary fitness in the combined population; this strategy will take over the population, and wipe out the other strategy, as above. Note that the initial communication and adoptation of an invading strategy does not affect the eventual evolutionary outcome unless it is in bistable equilibrium with the strategy being invaded; of course, adoptation would also affect the evolutionary outcome in other cases if the entire population adopted only one of the two strategies, but our model does not allow this to occur.

[2]In reality, the initial proportion of the population adopting an idea may be positively correlated with its long-term evolutionary success; this depends on whether the benefits and costs are immediately apparent. Allowing for correlation complicates the mathematics, but does not change our overall result. Also, a non-uniform probability distribution can be used for the initial proportion of invaders; we consider this generalization below.

5.3 Applying the model

Thus we have argued that our model is a more useful representation of real-world evolutionary interactions in a communicating population. It deals well with both stable and bistable equilibria, both of which are handled poorly by the Nowak-Sigmund model. We now consider how this can be applied to the Finitely Repeated Prisoner's Dilemma paradox. First, we note that no stable equilibria occur between C_i strategies in the FRPD game; we observe only dominance and bistable equilibria. For $R > 1$, C_j dominates C_i if $j = i - 1$, is dominated by C_i if $j = i + 1$, and is in bistable equilibrium otherwise. As derived above, for $j \gg i$, C_j can take over C_i with only a very small initial proportion of the population; thus in our model, the probability of C_j taking over C_i is nearly 1. Hence we expect to observe two phenomena: "evolutionary backward induction" (C_i invaded by C_{i-1}) and "evolutionary forward progression" (C_i invaded by C_j, $j \gg i$). For a large number of rounds N, and for most strategies C_i, there are a large number of strategies C_j with $j \gg i$. Under the typical assumption of a uniform prior, we would expect evolutionary progression to be much more common than evolutionary backward induction; as a result, we would expect the evolutionary process to be "pushed" toward continual cooperation rather than continual defection, resulting in domination of the population by strategies that cooperate until nearly the end of the game. If this expectation is satisfied, we have resolved the evolutionary FRPD paradox.

6 Results

We now examine the evolutionary performance of each strategy C_i, using our new evolutionary model, for various game lengths N. Note that there are $N+1$ strategies, C_0 through C_N, to consider for a given N. We use Markov chains to compute the equilibrium probability that each strategy dominates the population, as discussed above. This gives us a transition matrix P with

$$P_{ij} = \begin{cases} \frac{1}{N+1} & \text{if } j = i-1 \\ 0 & \text{if } j = i+1 \\ \min(\frac{1}{(N+1)(k-1)(T-1)}, \frac{1}{N+1}) & \text{if } j = i-k,\ k > 1 \\ \max(\frac{kT-k-T}{(N+1)(k-1)(T-1)}, 0) & \text{if } j = i+k,\ k > 1 \\ 1 - \sum_{j \neq i} P_{ij} & \text{if } j = i \end{cases}$$

We initially assume the $T = 3$ payoff table, $[T\ R\ P\ S]$ = [3 2 0 -1]. For example, if $N = 3$ we obtain the 4x4 transition matrix

$$P = \begin{bmatrix} \frac{11}{16} & 0 & \frac{1}{8} & \frac{3}{16} \\ \frac{1}{4} & \frac{10}{16} & 0 & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{4} & \frac{10}{16} & 0 \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{4} & \frac{9}{16} \end{bmatrix}$$

Computing the stationary distribution x of this matrix by solving $Px = x$, we obtain x = [.32 .23 .24 .20]. Thus for $N = 3$, C_0 (continual defection) is the best performing strategy. We now present the results, assuming $T = 3$, for a variety of N. First, we define strategy C_i as "cooperative" if it cooperates at least 90% of the game ($i \geq .9N$), and "uncooperative" if it cooperates at most 10% of the game ($i \leq .1N$). Then for each N, we calculate the stationary distribution, and from this we find 3 values: the best strategy C_i (i.e. the strategy with the highest x_i), the proportion of the time the population is dominated by "cooperative" strategies, and the proportion of the time the population is dominated by "uncooperative" strategies. These results are given in Table 2.

N	stationary distribution	best C_i	cooperative	uncooperative
1	[1 0]	0	0	1
2	[.6 .2 .2]	0	.2	.6
3	[.32 .23 .24 .20]	0	.2	.32
4	[.18 .17 .21 .24 .20]	3	.2	.18
5	[.10 .11 .15 .21 .23 .20]	4	.2	.10
6	[.06 .07 .10 .15 .20 .22 .19]	5	.19	.06
8	[.03 .03 .04 .07 .10 .14 .19 .22 .19]	7	.19	.03
10	[.01 .02 .02 .03 .04 $\ldots$ .14 .18 .21 .18]	9	.39	.03
20	[.002 .002 .003 $\ldots$ .09 .13 .17 .19 .17]	19	.53	.007
30	[.001 .001 .001 $\ldots$ .09 .12 .16 .18 .17]	29	.63	.004
40	[4×10^{-4} $\ldots$ .05 .06 .09 .12 .16 .18 .16]	39	.71	.002
50	[3×10^{-4} $\ldots$ .05 .06 .09 .12 .15 .17 .16]	49	.76	.002
100	[6×10^{-5} $\ldots$ .05 .06 .09 .11 .14 .16 .15]	99	.88	7×10^{-4}
200	[1×10^{-5} $\ldots$ .05 .06 .08 .11 .14 .15 .14]	199	.94	3×10^{-4}
400	[3×10^{-6} $\ldots$ .05 .06 .08 .10 .13 .14 .14]	399	.97	2×10^{-4}
800	[8×10^{-7} $\ldots$ .05 .06 .08 .10 .12 .14 .13]	799	.99	8×10^{-5}

Table 2: FRPD results for $T = 3$

Thus defectors dominate the population for very short games, but for longer games, the population is dominated by strategies that cooperate until near the end of the game. For $N = 20$, less than 1% of dominating strategies were uncooperative, and more than 50% of dominating strategies were cooperative. For $N = 800$, less than .01% of dominating strategies were uncooperative, and 99% were cooperative. C_0, the strategy which always defects, was the most successful for $N \leq 3$, but for all $N > 3$, the most successful strategy was C_{N-1}.

Our empirical results clearly show that the proportion of "cooperative" strategies goes to 1 as N goes to infinity. In fact, we can prove a stronger result than this: for any constant k, $0 < k < 1$, the proportion of strategies cooperating through turn kN goes to 1 as $N \to \infty$. In other words, if we had chosen 99.999% instead of 90% as the threshold for a "cooperative" strategy, the population would still be dominated by cooperative strategies for large enough N. A proof of this result is given in the appendix.

Next we examine the effects of changes in the payoff table on the distribution of dominating strategies. For a number of different values of T, we compute three values: the proportion of cooperative dominating strategies for $N = 20$, the proportion of uncooperative dominating strategies for $N = 20$, and the "break-even point" N_b, the smallest N such that the cooperator C_N outperforms the defector C_0. As above, we assume $R = T - 1$, $P = 0$, and $S = -1$.

T	cooperative ($N = 20$)	uncooperative ($N = 20$)	N_b
1.01	0	1	535
1.1	.055	.527	55
1.2	.106	.218	29
1.5	.214	.057	13
2.0	.348	.020	7
2.5	.458	.011	5
3.0	.534	.007	4
3.5	.588	.006	4
4.0	.628	.004	4
5.0	.685	.003	3
10.0	.797	.001	3
100.0	.896	1×10^{-4}	3

Table 3: FRPD results for varying T

As can be seen from these results, our choice of T strongly affects the proportion of cooperative strategies that dominate the population for a given N. Increasing T increases the proportion of cooperative strategies: for $T \approx 1$, defectors will dominate the population unless N is very large, but for $T \gg 1$, cooperators will dominate the population for all $N \geq 3$. For "normal" values of T ($2 \leq T \leq 4$), the break-even point between cooperation and defection varies from 4 to 7 rounds. This demonstrates that the choice of payoff table is important, but if the game is long enough, cooperation until near the end of the game can be achieved even if the reward for cooperating is small.

7 Generalizing the model

Thus we have presented an alternative model of the evolution of behavior, argued for its applicability to a communicating population, and demonstrated that it resolves the evolutionary FRPD paradox. We now consider the model's usefulness as a paradigm for "evolutionary rationality," that is, its applicability to any arbitrary game. First, rather than working with the total payoff $W(X|Y)$, we should use $w(X|Y)$, the expected value of the average payoff per round to X against Y. We were able to use the simpler quantity $W(X|Y)$ in our analysis of the FRPD because the number of rounds was finite, all strategies were deterministic, and our model assumed no noise; these assumptions may not be true in general. Second, we note that the model can only be used with a finite set of strategies: thus for infinitely long games such as the Iterated Prisoner's Dilemma, we must choose some finite space of strategies, such as the low-memory edge strategies (Neill, 2001) or finite automata of a given size. Even for finitely repeated games, the number of possible strategies M can be very large. For the Markov chain solution method, we must compute $O(M^2)$ dominance scores, and then compute the stationary distribution of an MxM matrix; this is infeasible for large M. Alternatively, we can run the simulation for a large number of rounds as described above, but this may take a long time to converge to values close to the stationary distribution. Thus, even for finitely repeated games, we may want to restrict the space of strategies we are considering. For example, restriction of the FRPD strategies to those of the class C_i reduced the size of the strategy space from $O(2^{2^N})$ to $O(N)$. Of course, if the strategy space is restricted arbitrarily, we must qualify our claim of rationality: the strategy is only "evolutionarily rational" with respect to the restricted strategy space under consideration.

Another generalization of the model is necessary if we have some meta-information about the "real" evolutionary model of the population of strategies we are considering. In particular, we are interested in the probability distribution $u(x)$ of the initial proportion of invaders in the population. This may vary based on the type of invasion (migration or mutation), frequency of invasion, amount of communication, and tendency to adopt new ideas. For example, use of the dominance score is only correct if $u(x)$ is uniform on (0,1), which may occur in populations with large amounts of communication and a reasonable likelihood of adopting new ideas. On the other hand, if we consider a population where invasion occurs by mutation, and communication is nonexistent, $u(x)$ may be Poisson with a low mean (proportional to the mutation rate). This would mean, in the case of a bistable equilibrium with $C_j \overset{m}{\Leftrightarrow} C_i$, that C_j is extremely unlikely to take over the population unless $m \approx 0$.

Thus we define the generalized dominance score $\mathrm{dom}_u(X|Y)$ as the probability that strategy X succeeds in invading strategy Y, given some probability distribution $u(x)$ of the initial proportion of invaders in the population. $\mathrm{dom}_u(X|Y)$ is identical to the standard dominance score, $\mathrm{dom}(X|Y)$, except in the case of a bistable equilibrium. In this case, assuming $X \overset{m}{\Leftrightarrow} Y$, the probability that X takes over Y is equal to the probability that the initial proportion of invaders is greater than m: $\mathrm{dom}_u(X|Y) = \int_m^1 u(x)dx$. Similarly, assuming $X \overset{m}{\Leftrightarrow} Y$, we know $Y \overset{1-m}{\Leftrightarrow} X$, so $\mathrm{dom}_u(Y|X) = \int_{1-m}^1 u(x)dx$. For the standard dominance score, $u(x) = 1$, so $\mathrm{dom}_u(X|Y) = 1 - m$ and $\mathrm{dom}_u(Y|X) = m$. Note that in general, we are *not* guaranteed that $\mathrm{dom}_u(X|Y) + \mathrm{dom}_u(Y|X) = 1$. Often, we assume a function u such that there is a smaller likelihood of accepting the invading idea than the idea being invaded, so $\mathrm{dom}_u(X|Y) + \mathrm{dom}_u(Y|X) < 1$ for almost all values of m.

Nevertheless, we can deduce several properties of $\mathrm{dom}_u(X|Y)$, again assuming $X \overset{m}{\Leftrightarrow} Y$. We assume that the initial population $u(x)$ always consists of a mixture of the two strategies: thus $P(u(x) = 0) = P(u(x) = 1) = 0$. This implies $\lim_{m\to 0} \mathrm{dom}_u(X|Y) = 1$. Similarly, $\lim_{m\to 1} \mathrm{dom}_u(X|Y) = 0$. Finally, since $u(x)$ is non-negative, we know that $\mathrm{dom}_u(X|Y)$ is non-increasing for increasing m.

These properties allow us to prove an important result: the evolutionary FRPD paradox is resolved for an evolutionary invasion model using any generalized dominance score $\mathrm{dom}_u(C_j|C_i)$ as the invasion probability q_{ij}. To see this, we consider q_{0N}, the probability that an attempted invasion of the defector C_0 by the cooperator C_N is successful. For large N, we know $C_N \overset{m}{\Leftrightarrow} C_0$ with $m = \frac{1}{(N-1)(T-1)}$. Thus $\lim_{N\to\infty} m = 0$. Now, if our invasion model uses a generalized dominance score, we know $q_{0N} = \mathrm{dom}_u(C_N|C_0)$, and thus $\lim_{N\to\infty} q_{0N} = \lim_{N\to\infty} \mathrm{dom}_u(C_N|C_0) = \lim_{m\to 0} \mathrm{dom}_u(C_N|C_0) = 1$. From this, it is clear that evolutionary forward progression can occur, and the population becomes dominated by cooperators for $N \to \infty$.

It should be noted that, under models where $X \overset{m}{\Leftrightarrow} Y$ implies $\mathrm{dom}_u(X|Y) = 0$ except for $m \approx 0$, very large N may be required for cooperation to succeed. For example, consider a model where the initial proportion of invaders is uniformly distributed on $(0,\epsilon)$, for $\epsilon \approx 0$. In this case, $q_{0N} = 1 - \frac{1}{\epsilon(N-1)(T-1)}$, so for $q_{0N} \geq .9$, we must have $N \geq 1 + \frac{10}{\epsilon(T-1)}$. For example, for $\epsilon = .001$ and $T = 3$, we obtain $N \geq 5001$.

Another interesting example occurs when the initial proportion of invaders is some fixed quantity $u(x) = u_0$. This may occur in models such as Bicchieri & Rovelli (1995), where most players are "adaptive" (readjusting their strategy at the end of each FRPD interaction, according to its relative success), but a small proportion of the population are "fanatics" who maintain their original strategy regardless of its relative success. In this case, q_{0N} is 1 if $m = \frac{1}{(N-1)(T-1)}$ is less than u_0, and 0 otherwise. For $u_0 = .01$ and $T = 3$, C_N can take over C_0 if $N > 51$.

The model can also be generalized to certain multi-player games, such as the repeated Tragedy of the Commons (an n-player variant of the FRPD). In this game, the payoff to each strategy on a given round depends on whether the strategy cooperates or defects, and the number of its opponents that cooperate. We consider here the variant in which a player receives the reward or temptation payoff only if every opponent cooperates, and the punishment or sucker payoff otherwise. Again assuming that the proportion of invaders is some fixed quantity $u(x) = u_0$, we find that q_{0N} is 1 if $m = \frac{1}{(N-1)(T-1)}$ is less than u_0^{n-1}, and 0 otherwise. For $u_0 = .01$, $T = 3$, and $n = 10$, C_N can only take over C_0 for $N > 5 \times 10^{17} + 1$.

Thus, for some generalized dominance scores, the game length must be very long for cooperation to be rational. Nevertheless, since cooperation is rational for long enough games, the evolutionary FRPD paradox is resolved.

8 Conclusions: toward an adaptive model of rational choice

Having resolved the evolutionary FRPD paradox, we now turn to the question of whether this solution resolves the original Finitely Repeated Prisoner's Dilemma paradox. Our answer, after some discussion, is a qualified "yes." Aumann (1995) argues convincingly that, when rationality is defined as "habitual payoff maximization" according to the subgame perfect equilibrium (Selten, 1975), common knowledge of rationality implies backward induction. Thus we have two alternatives: to accept the subgame perfect equilibrium as a normative definition of rationality, and therefore to accept the impossibility of cooperation in the FRPD with common knowledge of rationality, or to reject this definition of rationality as unreasonable. As is clear from this paper, this author prefers the latter approach; however, the creation of a single theory of rationality which both always prescribes reasonable behavior as "rational," and has good computational and theoretical properties, is a (very difficult!) open question.

One avenue of research that may lead toward this goal is a paradigm we call *adaptive rationality*. Adaptive rationality assumes a multitude of possible *rational decision procedures*, each based on different assumptions about the *meta-game*. The meta-game is the larger context in which the decision situation is embedded: this includes a performance measure (possibly based on some evolutionary model), and the distribution of opponent strategies. Neither of these is known for certain within the context of the game unless additional meta-information I is supplied, and hence assumptions must be made. A rational decision procedure is a method for choosing a strategy: it must be optimal with respect to some set of assumptions A that is compatible with the given meta-information I. A strategy picked according to any rational decision procedure is defined to be "rational," and it is impossible to know which decision procedure a given "rational" opponent will choose. There is no way of determining theoretically which one of these procedures is "better" for a given real-world interaction: this simply depends on which set of meta-game assumptions more accurately models the specific situation, and this can only be determined empirically, on a case-by-case basis. Also, though a rational decision-maker must form some initial assumptions about the meta-game, it can modify these assumptions over the course of a game, "adapting" to the behavior it has observed. In fact, if the opponent's behavior contradicts the assumptions underlying the current strategy of an adaptively rational decision-maker, that decision-maker must modify its assumptions (and if necessary, its strategy) to achieve compatibility with its observations.

For example, if no meta-information is given, the subgame perfect equilibrium is a rational decision procedure. It is optimal with respect to the assumption that all opponents are also rational in the "subgame perfect" sense (thus assuming deviations from equilibrium play are due to a "trembling hand"). For the FRPD, it is also optimal with respect to the assumption of a Nowak-Sigmund evolutionary model. On the other hand, the subgame perfect equilibrium is not optimal with respect to any assumption that implies the rationality of cooperation in the FPRD. If we were given incompatible meta-information of this sort, the subgame perfect equilibrium would not be a rational decision procedure.

Similarly, if no meta-information is given, various rational decision procedures can be derived from our evolutionary model. Under the assumptions that our evolutionary model corresponds well to the meta-game of the real-world interaction we are considering, and that it is rational to choose a strategy with high evolutionary success, "evolutionary rationality" is a rational decision procedure. In this case, the rational decision-maker performs a process of "fictitious evolutionary play," simulating the evolutionary model given some space of strategies, and chooses the strategy which dominates the population the highest proportion of the time; this is only feasible for a decision-maker with massive computational powers. Alternatively, one can assume that the opponent has evolved according to our model (and thus is selected at random from the stationary distribution), and maximize expected payoff based on this assumption. Another rational decision procedure, usable by a rational decision-maker with more limited computational resources, could be "changing one's mind" using the model: starting with a random strategy, and repeatedly deciding whether to keep that strategy or change to another random strategy based on the (generalized) dominance score. In this case, we choose the strategy that dominates the population at the end of the simulation; for long enough simulations, this is equivalent to computing and sampling from the stationary distribution.

Additionally, we note that the evolutionary solution method presented here gives some hints at possible modifications of the subgame perfect equilibrium, in order to create a (non-evolutionary) rational decision procedure which is compatible with the rationality of cooperation in the FRPD. The idea here is not to prevent the backward induction argument from occurring, but to also allow the possibility of forward progression: a rational player decides C_j makes more sense than C_i, $j \gg i$, and thus switches from C_i to C_j. This would require a definition of the subgame perfect equilibrium which can take non-equilibrium moves into account, adapting to the opponent's play; more research is necessary to determine whether this can be achieved.

Finally, we note an interesting corollary of our evolutionary model: as compared to non-communicating populations, communicating populations are dominated more often by cooperative strategies, resulting in higher average payoffs to all members of the population. In other words, the communicating population does not get stuck in poor but evolutionarily stable strategies ("local maxima" of the evolutionary space), resulting in better average performance. This may allow a communicating subpopulation to out-compete another subpopulation which does not communicate among themselves: in other words, natural selection will select for the evolution of communication.

These philosophical arguments aside, we briefly summarize our main result: an alternative model of the evolution of behavior, based on our rejection of the Maynard Smith assumption of a single invader. This model leads to a performance measure, the proportion of time a strategy dominates the population under our model. If we define rationality based on optimizing this performance measure, our model resolves the Finitely Repeated Prisoner's Dilemma paradox, demonstrating that cooperation is always possible for sufficiently long games. Additionally, the model can be generalized and applied to many other decision situations, and thus it serves as a possible standard for rational decision-making under uncertainty.

9 Bibliography

• Aumann, R. (1995). "Backwards induction and common knowledge of rationality," *Games and Economic Behavior* **8**, 6-19.

• Axelrod, R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

• Axelrod, R. & Hamilton, W.D. (1981). "The evolution of cooperation," *Science* **211**, 1390-1396.

• Bicchieri, C. & Rovelli, C. (1995). "Evolution and revolution: the dynamics of corruption," *Rationality and Society* **7**(2), 201-224.

• Binmore, K.G. (1987). "Modeling rational players I," *Economics and Philosophy* **3**, 179-214.

• Boyd, R. & Lorberbaum, J. (1987). "No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game," *Nature* **327**, 58-59.

• Kreps, D. et al. (1982). "Rational cooperation in the finitely repeated prisoner's dilemma," *Journal of Economic Theory* **27**, 245-252.

• Maynard Smith, J. (1982). *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

• Milinski, M. (1987). "Tit for tat in sticklebacks and the evolution of cooperation," *Nature* **325**, 433-435.

• Neill, D.B. (2001). "Optimality under noise: higher memory strategies for the alternating

prisoner's dilemma," *Journal of Theoretical Biology* **211**, 159-180.
• Nowak, M.A. & Sigmund, K. (1992). "Tit for tat in heterogeneous populations," *Nature* **355**, 250-253.
• Nowak, M.A. & 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.
• Selten, R. (1975). "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory* **4**(1), 25-55.

10 Appendix

Theorem: For our evolutionary model, using transition probability $P_{ij} = \frac{\text{dom}(C_j|C_i)}{N+1}$, and for any constant k $(0 < k < 1)$, the proportion of strategies cooperating through turn kN goes to 1 as the number of turns $N \to \infty$.

Proof: Let x_i denote the proportion of strategy C_i in the stationary distribution. For simplicity of notation, we assume that kN is not an integer; we call a strategy C_i "cooperative" if $kN < i \leq N$, and "uncooperative" if $0 \leq i < kN$. Let x_{high} denote the proportion of cooperative strategies in the stationary distribution, and x_{low} denote the proportion of uncooperative strategies. Let P_{low} denote the probability of switching to an uncooperative strategy, given that the current strategy is cooperative:

$$P_{low} = \frac{\sum_{i=\lceil kN \rceil}^{N} x_i \sum_{j=0}^{\lfloor kN \rfloor} P_{ij}}{\sum_{i=\lceil kN \rceil}^{N} x_i}$$

Similarly, let P_{high} denote the probability of switching to a cooperative strategy, given that the current strategy is uncooperative:

$$P_{high} = \frac{\sum_{i=0}^{\lfloor kN \rfloor} x_i \sum_{j=\lceil kN \rceil}^{N} P_{ij}}{\sum_{i=0}^{\lfloor kN \rfloor} x_i}$$

We know that $x_{low} = \frac{P_{low}}{P_{low}+P_{high}}$. From this, we can find an upper bound for x_{low} by computing an upper bound for P_{low} and a lower bound for P_{high}. To do this, we rely on the following observation: for $j > i$, P_{ij} is decreasing with i and increasing with j, and P_{ji} is decreasing with j and increasing with i. Thus P_{low}, the weighted average of $\sum_{j=0}^{\lfloor kN \rfloor} P_{ij}$ for all $i = \lceil kN \rceil \dots N$, is bounded by:

$$P_{low} \leq \sum_{j=0}^{\lfloor kN \rfloor} P_{\lceil kN \rceil j}$$

Similarly, P_{high}, the weighted average of $\sum_{j=\lceil kN \rceil}^{N} P_{ij}$ for all $i = 0 \dots \lfloor kN \rfloor$, is bounded by:

$$P_{high} \geq \sum_{j=\lceil kN \rceil}^{N} P_{\lfloor kN \rfloor j}$$

We now simplify each expression using $P_{ij} = \frac{\mathrm{dom}(C_j|C_i)}{N+1}$, where $\mathrm{dom}(C_j|C_i)$ equals 1 if $i-j=1$, $\frac{1}{2(i-j-1)}$ if $i-j>1$, and $1-\mathrm{dom}(C_i|C_j)$ if $i-j<0$. Thus:

$$P_{low} \leq \sum_{j=0}^{\lfloor kN \rfloor} P_{\lceil kN \rceil j} = \frac{1}{N+1} + \sum_{j=0}^{\lfloor kN \rfloor - 1} \frac{1}{2(N+1)(\lfloor kN \rfloor - j)} = \frac{1}{N+1}\left(1 + \sum_{j=1}^{\lfloor kN \rfloor} \frac{1}{2j}\right)$$

Similarly for P_{high}:

$$P_{high} \geq \sum_{\lceil kN \rceil}^{N} P_{\lfloor kN \rfloor j} = \sum_{\lceil kN \rceil + 1}^{N} \left(\frac{1}{N+1} - \frac{1}{2(N+1)(j - \lceil kN \rceil)}\right) = \frac{1}{N+1}\left(\sum_{j=1}^{N-\lceil kN \rceil}\left(1 - \frac{1}{2j}\right)\right)$$

We now look at the asymptotics using $\sum_a^b \frac{1}{j} = O(\log b - \log a)$. Thus $P_{low} \leq \frac{1}{N+1} O(\log kN) = O\left(\frac{\log N}{N}\right)$ for fixed k. Similarly, $P_{high} \geq \frac{1}{N+1} O((N-kN) - \log(N-kN)) = O(1)$ for fixed k. Thus $x_{low} = \frac{P_{low}}{P_{low}+P_{high}} \leq \frac{O\left(\frac{\log N}{N}\right)}{O(1)} = O\left(\frac{\log N}{N}\right)$, which converges to 0 as $N \to \infty$. Thus $x_{high} = 1 - x_{low}$ converges to 1 as $N \to \infty$. $\square$