Protocols for Cooperation
Cultural Diversity of Strategies for the Alternating Prisoner's Dilemma

David P. Kraines and Vivian Y. Kraines

The Prisoner's Dilemma has long been the paradigm for modeling cooperation in a competitive environment. The classical version, in which both agents simultaneously choose to cooperate or to defect, is unnecessarily artificial. The stochastic Alternating Prisoner's Dilemma in which choices are made sequentially and mistakes are possible is more faithful to examples in the real world. Intelligent *k*-ply agents, who base their action on *k* previous outcomes, can achieve mutual cooperation in the stochastic Alternating Prisoner's Dilemma. Among the 4-ply agents that remember the previous two choices of the agent and of his opponent, several mutually evolutionarily stable clans will emerge in an evolutionary model based on Darwinian natural selection. In particular, the stochastic 4-ply strategy dubbed Tough Love maintains a stable cooperative environment by punishing misbehavior while quickly forgiving the repentant.

Key Words: Cooperation, Prisoner's Dilemma, cultural diversity, evolution, adaptive dynamical system.

1. Introduction

People tend to help their neighbors and even go out of their way for strangers, yet in an overly trusting society a crook or con man would gain at the expense of the majority. One might assume that others would learn to imitate this more successful strategy. The memes for cheating should spread and the cooperative structure would crumble causing the society as a whole to suffer. Why is it that most people in our societies continue to get along with each other?

Some have argued that, because groups of altruists help each other, those groups are more successful than groups of egoists. Thus by a process akin to natural selection, groups of altruists will flourish while those of egoists will die out. But as G. C. Williams (1966) and others have compellingly argued, simple altruistic behavior is unsustainable except among close relatives. Dawkins (1976) and others extend this reasoning to memetic adaptation in human society.

Game theoretic models have offered a possible solution to this paradox. Maynard Smith & Price (1973) applied techniques of this mathematical theory to evolutionary biology. They introduced the concept of evolutionarily stable strategies and explored how various behavioral patterns (strategies) could become evolutionarily stable even though they may be less successful than other strategies in one on one interactions. See also (Maynard Smith 1982).

The Prisoner's Dilemma has long been the paradigm in game theory for modeling the trade-off between cooperative and competitive interactions among human societies. Assume that two agents interact and that each may cooperate or defect. If each agent cooperates (plays C), then they each get a reward of R while if each defects (plays D) then they each receive a lesser penalty of P. If the first agent were to cheat or defect while the second meekly cooperated, then the first would gain the exploitation or temptation payoff of $T > R$ while the second would suffer the sucker payoff $S < P$. This information is traditionally encoded in the payoff matrix $\begin{bmatrix} R,R & T,S \\ S,T & P,P \end{bmatrix}$.

In most of this paper we will assume that the payoff entries G = [R, S, T, P] = [1, -2, 2, -1].

Consider a population of agents whose "fitness", or reproductive potential, is determined by the outcome of a one-shot PD with others in the society. If most are altruists (always play C) but a few are egoists (always play D), then the selfish minority will be far more successful and, in succeeding generations, will replace the altruists. When most of the population are egoists, then all lose.

Cooperation has some chance to develop if the PD is played repeatedly by the same agents - the so-called Repeated or Iterated Prisoner's Dilemma (IPD). It is still true that the altruistic agent who always cooperates (`AllC`) would continually lose to the selfish agent who always defects (`AllD`), but more sophisticated agents can escape from this dilemma. Work of Rapoport (1965) and others had suggested that a **Tit For Tat** (`TFT`) type strategy - repeat the other agent's previous move - is successful in many environments. `TFT` agents are neither altruists nor egoists. Although cooperative with other nice strategies, the `TFT` agent who is cheated will retaliate rather than turn the other cheek. These agents neither exploit nor are exploited; the total payoff to any agent `A` in a finite IPD with `TFT` must be either equal to the payoff to `TFT` or at most greater than that by only T–S. In round robin computer tournaments with many different strategies, Axelrod (1976) discovered that the `TFT` strategy scored the highest total amount. Moreover, in ideal environments, in a population of arbitrary agents, those using a`TFT` -like strategy may arise by a process akin to evolution.

The initial success for `TFT` agents in the IPD as a model for the evolution of cooperative behavior dimmed under closer examination. By accident or miscommunication, an agent using this `TFT` strategy might defect or be perceived to do so. An opponent using this `TFT` strategy would retaliate just as the original defector cooperated. The agents could get stuck in a vicious cycle of retaliations, CD DC CD ... or DD DD DD... , seen too often in ethnic conflicts around the world. An exploiting but repentant strategy dubbed `Pavlov` (also known as the Simpleton and Win Stay Lose Switch) (Kraines & Kraines 1993, Nowak & Sigmund 1993) was shown to be successful in many environments.

Many applications to the real world claimed for classic game theory are quite artificial and simulations and tests used to study the classical iterated Prisoner's Dilemma tend to be contrived. Actions in the real world are rarely simultaneous. More often each agent reacts in turn with knowledge of the other's action.

The alternating or sequential PD seems to be a more realistic model for social interactions. The alternating PD (APD) was investigated by Nowak & Sigmund in 1993 using methods from their work on IPD. Their results were generalized and extended later by Frean (1994) and Hauert & Schuster (1998) among others. Brams (1994) has introduced many examples from political science for which the simultaneous (iterated) PD is not as good a model as the alternating game. Wedekind & Milinski (1996) found that their students would use either a Generous `TFT` (`GTFT`) or a `Pavlov`, type strategy in simulations of the APD . Frean (1994) asserts that neither `TFT` nor `Pavlov` is stable in the APD. Rather a combination of these agents, dubbed **Firm But Fair** (`FBF`), emerges as a successful cooperator. Hauert & Schuster (1998) extend these results to 3 and 4-ply agents.

In this paper, we modify the evolutionary model of Kraines & Kraines (2000) for the IPD and apply it to the APD. More precisely, we study the natural selection (or adaptive) dynamical system mimicking Darwinian natural selection (Nowak & Sigmund 1989, Hofbauer & Sigmund 1990, Kraines & Kraines 2000) for the set of agents that consider the previous k decisions (plies). We conclude that cooperative, though not overly altruistic, behavior can evolve into a reasonably stable cooperative strategy provided the individuals are able to remember the result of several previous encounters.

Cooperative strategies that emerge among 2 and 3-ply are similar to those studied by Frean (1994) and Hauert & Schuster{1998}, but important differences emerge for 4-ply agents. In the latter case, cooperative evolutionary trajectories lead to several successful clans of agents. To maintain cooperation, each clan has a distinctive set of cultural rules or protocols that combine elements of FBF and GTFT. If a few agents using one of these protocols were to interact in a different clan, then both the visitors and the hosts would be worse off than if they interacted within their own clan. This cultural differentiation among clans is evolutionarily stable with respect to the other clans.

2. The alternating prisoner's dilemma:

Assume that two agents, A and B, alternately choose to cooperate or defect in the PD. Denote the choice of A by C or D and that of B by c or d. The game continues for an infinite or a large number of moves.

Several methods have been proposed to assign payoffs after the action of just one agent (Nowak & Sigmund 1994, Frean 1994, Hauert & Schuster 1998). In our model, we treat each ply, or half round, after the first one as if it were a complete Prisoner's Dilemma.

The game begins when the leader A either bestows a favor on B (cooperates) or deprives B of some asset (defects). No payoff takes place after this first ply. If A cooperated, then B may return the favor (plays c after C) or exploits A by taking more (plays d after C). If A defected, then B may meekly accept the loss (play c after D) or resist the demand and retaliate (play d after D). At this stage, the agent A receives a payoff of R, S, T, or P respectively according to the usual Prisoner's Dilemma scheme. A payoff of R, T, S or P, respectively, is given to agent B at that same time. At the third step (beginning of the second round), agent A either cooperates or defects and a second payoff of R, S, T or P accrues to A and to B based on the decisions of step 2 and 3 only. The process continues indefinitely with agent A choosing C or D at step $m = 2n - 1$ and agent B choosing c or d at step $m = 2n$. This means that each agent receives two (positive or negative) payoffs per round and this payoff is determined by the last two choices.

Ply:	1	2	3	4	5	6	7
A:	C		D		D		C
B:		c		c		d	
Payoff:							
A:	0	R	T	T	T	P	S
B:	0	R	S	S	S	P	T

Table 1 Payoff in the APD after the initial seven steps sequence of choices σ = CcDcDdC

After the initial seven steps sequence of choices σ = CcDcDdC, A receives a sum of *six* payoffs of R + T + T + T + P + S and correspondingly B receives R + S + S + S + P + T as in Table 1. Let w_n(A|B) be the payoff to A at the n^{th} step. The payoff w(A |B) to agent A in a match of N steps, or half-rounds, with B is $\frac{1}{N-1}\sum_{n=2}^{N} w_n(A|B)$. In the infinite APD, w(A |B) is defined to be the limit of this running average. The omission of an initial payoff will not significantly affect the average payoff in an infinite or very long running APD.

3. *k*-ply strategies.

Agent A is said to use a (mixed) *k*-ply strategy *T*, or that she is a (mixed) *k*-ply agent, if, for each sequence σ of the previous *k* choices, agent A plays C with a fixed probability Pr(σ) depending only on this sequence σ. Mathematically, a *k*-ply strategy is a function from the set of the 2^k possible sequences to the interval from 0 to 1.

Reactive or 1-ply agents were introduced in the simultaneous IPD in (Nowak & Sigmund 1989) and in the APD in (Nowak & Sigmund 1994). They can be encoded as a pair A=A(*p, q*) of probabilities where *p* is the probability that A will cooperate if his opponent cooperated in the previous round and *q* is the probability that he will cooperate if his opponent defected. A mixed 2-ply strategy is represented by the 4 tuple A = A(*r, s, t, p*) with *r* = Pr(Cc), *s* = Pr(Cd), *t* = Pr(Dc) and *p* = Pr(Dd), corresponding to the probability that A cooperates after the given sequence of the two previous choices. This input for a 2-ply agent in the APD is the same as that for the input of the memory-1 strategy (*r, s, t, p*) for the simultaneous IPD.

We distinguish three types of mixed *k*-ply agents. *Deterministic* agents will definitely cooperate or defect for each given sequence, i.e., Pr(σ) = 0 or 1. *Fallible* (almost pure or noisy) agents are determined to cooperate or defect for each sequence but occasionally make a mistake or are perceived to make a mistake, i.e., Pr(σ) = ε or 1-ε for some fixed small ε > 0. *Stochastic* agents will never cooperate nor defect with absolute assurance, i.e., ε ≤ Pr(σ) ≤ 1-ε.

3.1 Deterministic agents

The four deterministic 1-ply agents are AllC = (1,1), AllD = (0,0), TFT = (1,0) and the Bully = (0,1). Call an agent *nice* and *trusting* if he starts out cooperating and has Pr(Cc) = 1. Two nice and trusting agents such as AllC and TFT will continue to cooperate and so each will receive the reward R on each play. A selfish AllD agent (0, 0) will exploit the altruistic AllC and thrive but he cannot exploit a TFT agent.

.

Every *k*-ply agent may be extended to a *k*+1-ply agent by ignoring the first entry of the length *k*+1 sequence; for each length *k* sequence σ, the *k*+1-ply extension is given by Pr(Cσ) = Pr(Dσ) = Pr(σ). The 16 deterministic 2-ply strategies include the extensions of the 1-ply agents above and a few others of interest such as Pavlov = (1, 0, 0, 1), Grim or Trigger = (1, 0, 0, 0) and pure Firm But Fair FBF = (1, 0, 1, 1). Each of these agents are nice and trusting but will retaliate when wronged. FBF combines the cooperative and forgiving aspects of TFT with the contrition of Pavlov.

The Bully , also known as Tat for Tit, is a rather unsophisticated strategy who cheats a cooperator but gives in to a defector. The Bully exploits AllC, winning an average of T each play but is exploited by AllD, losing S each round to AllD's great advantage. Pavlov will also exploit the Bully except perhaps on the first play. When two Bullies play each other, one will exploit the other (CdCdCd.. or DcDcDc...) to the same extent that AllD exploits Bully. If a Bully plays TFT or FBF then the sequence becomes CcDdCcDd with average payoff of $\frac{1}{4}$ (R+S+T+P) to each. Neither TFT nor FBF is able to take advantage of this poor strategy.

3.2 Fallible agents

Even for very small $\varepsilon > 0$, the behavior of fallible agents can be quite different from that of their deterministic cousins. After the inevitable error, TFT(1-ε, ε, 1-ε, ε) gets into a mutual defection rut with her clone, CcC**d**DdD, until another error restores cooperation. Each receives R during the cooperation period, a T or S immediately after a mistake is made, and P after that. With probability ε, one of the agents will accidentally cooperate leading to another period of mutual cooperation. The average self-payoff is $\frac{1}{2}$ ((1-ε)(R+P) + ε(S+T)). Unlike in the IPD, two fallible TFT agents do not fall into the alternating CD DC CD ... rut in the APD.

Although fallible Pavlov = (1-ε, ε, ε, 1-ε) proved very successful in the simultaneous IPD, he does not fare well against his clone in the APD. Although they will each cooperate initially, eventually one will defect by mistake, the other will retaliate, the first will repent and cooperate, but the second will defect in an attempt to exploit. In other words, the sequence of plays is a cycle of length 6 that proceeds as C**d**DcDdCdDc … with two defections for every cooperation and an average payoff of $\frac{1}{3}$ (S+T+P) = $-\frac{1}{3}$ with the standard payoff G. Cooperation is harder to restore after this error than for TFT agents. Neither a "mistaken" cooperation by one of the fallible Pavlov agents rather than the expected retaliation (...cDdCd**C**dDc..) nor a "mistaken" defection instead of a repentant cooperation (...DdCdD**d**CdC....) will change the lose-lose pattern. Trust can be restored only if the retaliator accidentally cooperated rather than exploited her clone (..DcDdC**c**CcC..). Thus mutual cooperation can be restored by a second error at just two stages in the cycle of length 6. Pavlov and his clone are in this cycle three times as long as in the cooperative pattern and the average payoff to each in an infinite game is $\frac{1}{4}(1)+\frac{3}{4}(-\frac{1}{3}) = 0$ for the standard payoff.

In contrast to TFT and Pavlov, fallible FBF(1-ε, ε, 1-ε, 1-ε) will not get stuck in vicious cycles but will readily resume cooperation after a mistake. On the negative side, FBF and Pavlov will suffer $\frac{1}{2}$ (P + S) = $-\frac{3}{2}$ against AllD while AllD receives $\frac{1}{2}$ (P+T) = $\frac{1}{2}$ up to terms of order ε. TFT will do as well as any agent against AllD with an average loss of P = -1.

There are significant differences between the IPD and the APD in other matches. The IPD between fallible Pavlov and TFT degenerates into a lose-lose match (CC C**D** DC DD CD ..) (Kraines & Kraines 1993). In the APD, if Pavlov errs, the sequence of moves is Cc**D**dCcCc while if TFT were the first one to err, the sequence becomes C**d**DdCc… so that cooperation is quickly restored after a single error. Both TFT and Pavlov recover faster in a match against each other in an APD with noise than do either TFT or Pavlov in an APD against their clones.

Table 2 summarizes the payoff to fallible agents AllD, AllD, TFT, Pavlov and FBF against each other in the APD with noise level ε and with the standard payoff at least up to terms of order ε^2.

	AllD	TFT	FBF	PAV	AllC
AllD	-1 + 2ε	-1 + 5ε	0.5 – 0.25ε	0.5 – ε	2 – 4ε
TFT	-1 + ε	0	1 – 3ε	1 – 3ε	1 – ε
FBF	-1.5 + 3.75 ε	1 – 7ε	1 – 4ε	1 – 6ε	1 – ε
PAV	-1.5 + 3ε	1 – 7ε	1 – 6ε	0	1.5 – 3ε
AllC	-2 + 4ε	1 – 5ε	1 – 5ε	-0.5 + ε	1 – 2ε

Table 2 Round Robin payoffs among elementary fallible agents

3.3 Stochastic agents

Both memory-one strategies for the simultaneous IPD and 2-ply strategies in the APD can be written in the form (*r, s, t, p*). For example, the agent Random = $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ "flips a coin" at each decision point. Agents are called *stochastic* if the values of (*r, s, t, p*) are between ε and 1–ε for fixed *noise level* ε>0. Because of this noise, stochastic agents will never cooperate or defect with certainty. Agents are called *trusting* if *r* is near one and *cheating* if *r* is near 0. A trusting agent is called *generous* if *s* is moderately large. Generous agents will often forgive a defection. The *trait space*, the set of all these strategies, forms the 4-dimensional hypercube **H** with sides of length 1–2ε. Similarly, the trait space for a population of *k*-ply strategies playing an APD forms a 2^k dimensional hypercube.

4. Replicator equations

If the payoff to the strategies A and B satisfies *w*(A|B) > *w*(B|B), then we say that A *invades* B (Maynard Smith 1982) and we write A →B. Otherwise write A ¬B. If A → B but B ¬A then we say that A *dominates* B and write A > B. It may happen that A → B and B →A in which case we write A ↔ B or that A ¬B and B ¬A in which case we write A ⇔B. According to Table 2, AllD, Pavlov and FBF will each dominate AllC while AllC ↔TFT. If A ↔ B, then as in [Hofbauer & Sigmund 1998] a dimorphic population will tend toward a stable mix in which A makes up a proportion *q* of the population where *q* is given by the *stable equilibrium* ratio:

$$q = \frac{w(A|B) - w(B|B)}{w(A|B) + w(B|A) - w(A|A) - w(B|B)} \qquad (4.1)$$

In a dimorphic population of TFT and AllD, neither type can invade the other. If A ¬B and B ¬A, then the ratio *q* in (4.1) gives the *bistable equilibrium* ratio, or tipping point for domination by A. If the proportion of agents greater than *q* use the A strategy, then the B agents will die out and the population will tend toward a monomorphic A society.

Following Neill (2001), define the *dominance measure* dom(A |B) as follows. If A > B, then set dom(A |B) = 1 and dom(B |A) = 0. For example, dom(AllD |Pavlov) = 1. If A ↔ B so that a dimorphic population tends toward a stable equilibrium with q proportion of agents A, then dom(A |B) = q. For example, by Table 2 dom(AllC |TFT) = $\frac{\varepsilon}{1-4\varepsilon}$. If A ⇔B with bistable equilibrium q, then a dimorphic population with more than q proportion of agent A will become monomorphic of type A. In this case define dom(A |B) = 1–q. For example, by Table 2 dom(AllD |FBF) = $\frac{1}{2}+\frac{3}{2}\varepsilon$. In all cases, dom(A |B) =dom(B |A).

5. Payoff in the APD

Assume that a mixed 2-ply strategy A = (r, s, t, p) plays against another such B = (r', s', t', p') in an infinite (or very long) APD. Frean (1994) introduced the Markov transition matrix for A

$$T_A = \begin{bmatrix} r & 1-r & 0 & 0 \\ 0 & 0 & s & 1-s \\ t & 1-t & 0 & 0 \\ 0 & 0 & p & 1-p \end{bmatrix} \text{ and the analogous one, } T_B \text{ , for B.} \qquad (5.1)$$

Each row represents one of the four possible states, or sequences of plays, Cc, Cd, Dc, Dd, presented to agent A. Similarly, each column correspond to the states or sequences cC, cD, dC, and dD presented to B. The matrix entries correspond to the probability of transforming from a row state to a column state. For example, if A played C and B followed with d (row 2), then A will play D and move to state dD, column 4, with probability 1–s. From state dD, row 4 for T_B, agent B will in turn play d with probability 1– p'. This would present A with Dd.

The T_A matrix transforms states presented to agent A into states presented to agent B and so represents only half of a round or one ply. The transition matrix T_B transforms these states back to A's states using his opponent's mixed strategy. Thus the transition matrix $T_{AB} = T_A * T_B$ will transform one round (2-ply) histories presented to agent A through B 's choice and back into states presented to agent A. (Nowak & Sigmund 94, p222).

$$T_{AB} = \begin{bmatrix} rr' & r(1-r') & (1-r)s' & (1-r)(1-s') \\ st' & s(1-t') & (1-s)p' & (1-s)(1-p') \\ tr' & t(1-r') & (1-t)s' & (1-t)(1-s') \\ pt' & p(1-t') & (1-p)p' & (1-p)(1-p') \end{bmatrix} \qquad (5.2)$$

Similarly, the matrix $T_{BA}=T_B*T_A$ transforms states presented to B back to states presented to B. For example, after being presented with Cd, with probability (1–s)(1–p'), A will be presented with Dd on his next turn.

This Markov transition matrix can be used to find the expected payoff to the two agents in an infinite length APD as follows. T_{AB} has a left (row) eigenvector with eigenvalue 1 normalized so that the sum of the entries is 1. This vector eq_A = (q_1, q_2, q_3, q_4) is called an *equilibrium probability state* or *vector* for the Markov chain. If the probabilities (r, s, t, p) are strictly between

0 and 1, in particular if the agent is stochastic, then the entries of T_{AB} are all positive so that T_{AB} is a regular Markov matrix with unique equilibrium vector. (Kemeny and Snell, 1976).

The entries q_i of the equilibrium vector are numbers in [0, 1] that correspond to the proportion of rounds in an infinite game in which agent A is presented with a given state. If the equilibrium vector between the 2-ply agents A and B is eq_A = (0.1, 0.3, 0.2, 0.4), then sequential cooperations occur 10% of the time while A is suckered (his opponent responds with a d after his C) 30% of the time. The expected payoff to each from their mutual cooperation is 0.1·R = 0.1. The expected payoff to B from the times he exploits A is 0.3·T = 0.6 while A gets a negative value of 0.3·S = -0.6. More generally, the total expected payoff to A is 0.1·R + 0.3·S + 0.2·T + 0.4·P = -0.4 while the total expected payoff to B is 0.1·R + 0.3·T + 0.2·S + 0.4·P = -0.1. Every round consists of two separate payoffs. After A cooperates or defects, the payoffs paid to A and to B are given by a similar formula using eq_B.

The extension from 2-ply to *k*-ply agents is made simpler with an abbreviation.

Notation: Denote by [a, b, c, d]" the 2^k tuple vector (a, b, c, d, a, b, …c, d) for k>1.

Definition 5.1: In an infinite APD between *k-ply* agents A and B, the *payoff to* A is

$$w(A|B) = \tfrac{1}{2}(\langle [R, S, T, P]", eq_A \rangle + \langle [R, T, S, P]", eq_B \rangle)$$

and the *payoff to* B is

$$w(B|A) = \tfrac{1}{2}(\langle [R, T, S, P]", eq_A \rangle + \langle [R, S, T, P]", eq_B \rangle)$$

where < , > represents the inner product.

With only two states for 1-ply agents, it is not possible to calculate the payoff for a full round using this method. Instead we identify the 1-ply agent (*x, y*) with the 2-ply agent (*x, y, x, y*) and then compute the equilibrium state of the transition matrix and the associated payoff as above. See the appendix for a 3-ply example.

To simplify the calculations and to get a better understanding of the relation between the two plies of a rounds, we establish a few results.

Proposition 5.2: If the transition matrix is regular, then $eq_A \cdot T_A = eq_B$ and $eq_B \cdot T_B = eq_A$

proof: Since the transition matrix is regular, for any probability vectors X and Y, $X(T_{AB})^n$ converges to the row vector eq_A and $Y(T_{BA})^n$ converges to eq_B as *n* gets large. If $Y=X \cdot T_A$, then $eq_A \cdot T_A = \lim X(T_{AB})^n \cdot T_A = \lim (X \cdot T_A)\cdot(T_{BA})^n = \lim Y \cdot (T_{BA})^n = eq_B$. The other statement is similar.

Theorem 5.3: The payoff to A at the first ply of a round minus that at the second ply of the round:

$\langle [R, S, T, P]", eq_A \rangle - \langle [R, T, S, P]", eq_B \rangle$ is equal to

$K(R + P - S - T)$ where $K = \sum_{i=0}^{2^{k-2}} (eq_A(4i+1) - (eq_B(4i+1))$

represents the average proportion of Cc before the first ply in a round minus the proportion of cC before the second ply in that round.

Proof is given in the appendix.

Corollary 5.4: If the payoff matrix is *decomposable*, i.e., R+P = S+T, then the expected payoff to either one of the agents is the same in each of the plies and so the payoff in the full round is twice that in each ply.

For a decomposable PD payoff matrix, the payoff to A can be computed as the dot product <[R S T P]" ,eq_A > rather than the average of two dot products. This reduces the computation time by about a factor of 2. Nowak & Sigmund (1994) use this result in their 1-ply and 2-ply APD analysis.

6. Evolution by adaptive dynamics:

Assume that the environment consists initially of a population of *k*-ply agents each using the same mixed *k*-ply strategy S. If each member of this population plays each other in an infinite or long run APD, then each will expect the same (large or small) payoff.

Nowak & Sigmund (1993) modeled the evolution of strategies by introducing a small number of other strategies at random. Each agent interacts with the others and the more successful of these increase in proportion while the weaker ones die out. Intuitively, the population is visited periodically by migrants or mutant strains that interact with the original population. Often, the mutant strains will die out. On occasion, newcomers that are able to invade the population in the sense of Maynard Smith (1982) will either replace the original or coexist with them. We call this the *migrant evolution model*. This approach has been extended to *k*-ply agents in the APD for $k \leq 4$ by Nowak & Sigmund (1994), Frean (1994), Hauert and Schuster (1998) and others.

As a model for natural or Darwinian evolution, this approach has serious deficiencies. Natural selection is a very gradual process requiring a near continuum of intermediate types. The migrant evolution model assumes that the newcomers bear little resemblance to the current population. Indeed to speed up the process, the probability distribution for introducing random migrants favors those near the boundary of the trait space. Even with this bias toward deterministic agents, most simulations of the migrant evolution model run for 10^7 generations or more with frequent discontinuous changes. It is true that in real life foreign species are occasionally introduced into a new environment. They usually die out soon or quickly replace native species as in the migrant evolution model. But these foreign species must have had the time and opportunity to evolve in their own land.

The adaptive dynamics model, introduced for the reactive strategies in the IPD by (Nowak & Sigmund 1989, Hofbauer & Sigmund 1990) and studied for memory-one strategies (Kraines & Kraines 2000) seems more realistic than the migrant evolution model. One may think of this model as follows. Assume that an initial monomorphic population of strategy S has numerous offspring, each slightly different from its parents. Mathematically, this means that the offspring strategies are uniformly distributed around the parent S and that the Euclidean distances between them are normally distributed with a small standard deviation.

To model "survival of the fittest" assume that each of the offspring interacts with all other offspring in an infinite APD and that only that variety with the greatest payoff survives. For the 2^k component *k*-ply agent S, let $F_S(X)$ be the payoff to an arbitrary *k*-ply agent X against agent S. The successor agent will be that nearby variation T of S for which $F_S(T)$ is greatest. This will now be made more precise.

Let the "genetic or memetic variation" between parent and offspring be the adjustable parameter $\delta > 0$. Let Grad be the 2^k component gradient vector of F_S evaluated at X = S. This vector Grad points in the direction from S toward its most successful descendant and its magnitude K is equal to the rate of change in that direction. In Kraines & Kraines (2000), we showed that the discrete version of the adaptive dynamics equations, $\frac{d}{dt}$ S = grad F_S, with the successor agent T = S + δGrad, is a reasonable approximation for natural selection.

This formula must be modified so that the offspring remains within the trait space. If S + δGrad lies outside the trait space **H**, then T is replaced by the orthogonal projection of S + δGrad onto the appropriate face of **H**. The *effective gradient* EffGrad is formed by modifying those components of Grad that would cause the corresponding components of the successor T to be less than ε or greater than 1−ε. For example, if δ = 0.01 and Pr(Cd) = 0.015 and Grad(Cd) = -2, then

$$T(Cd) = \min[\, Pr(Cd) + \delta Grad(Cd), \varepsilon] = 0.01 \qquad (6.1)$$

and so EffGrad(Cd) = -0.5. If when evaluated at this successor, Grad(Cd) < 0, then EffGrad(Cd) = 0, i.e, the grandchild has the same Cd component as the child. The (Euclidean) distance between parent and successor child is δK where K is the magnitude of EffGrad at S. In this discrete dynamical system, the evolutionary track is uniquely determined by the initial conditions and the adjustable parameters δ and ε.

7. Basins without noise

In the noiseless memory one IPD, nearly 80% of the 3 dimensional Pr(CC) = 1 face of the hypercube was shown to consist of stable cooperating agents (Kraines & Kraines 2000). In this region, the only non-zero component of the gradient is positive and parallel to the CC direction and so the effective gradient vanishes and the trajectory must stop. These agents, including variations of TFT and Pavlov but not of AllC, are evolutionarily stable in the sense of Maynard Smith. They continually cooperate with their clones and yet less altruistic nearby agents cannot exploit them. Although trajectories starting at random agents in **H** that get to this large portion of the cooperative face will terminate there, simulations for the memory one noiseless IPD indicate that only about 20% of trajectories do end in this region. Almost $\frac{2}{3}$ of trajectories terminate in the defecting basin while most of the rest tend toward stable *alternator* agents of the form (x, 0, 1, y) with outcomes like ... CD DC CD ... and average payoff S+T.

In the 1-ply noiseless APD, stable cooperators satisfy Pr(c) = 1 and Pr(d) < $\frac{2}{3}$ as will be shown in section 8.1. Thus 66% of the Pr(c)=1 face of the 2 dimensional trait space consists of stable cooperators. For k =2 about 75% of the Pr(Cc) = 1 face consists of stable cooperators while for k = 4 this drops to about 58%. Yet only about 19% of random trajectories of 1-ply and 30% of 2-ply agents converge toward this face while 42% of 3-ply and 44% of 4-ply trajectories stabilize on the cooperative region of the face. See Table 3.

Alternator cycles are unstable and rarely appear along a trajectory of noiseless 2-ply agents. The 3-ply versions of these alternator agents do arise infrequently along the trajectories and the 4-ply versions are more stable and more common. Table 3 lists results of simulations on 2, 3, and 4-ply agents in the noiseless APD starting from 1000 random initial agents.

	D basin	C basin	positive 3-cycles	negative 3-cycles	4 cycles	other
1-ply	652	191	0	0	0	157
2-ply	698	302	0	0	0	0
3-ply	542	424	22	6	5	0
4-ply	422	435	56	69	15	3

Table 3 Basin sizes for k-ply noise-less trait space

8. Cooperative, defecting and cyclic basins

In the stochastic IPD and in the stochastic APD, the stable regions on the Pr(cC..c) = 1−ε face of the trait space consist of a few isolated agents. Typically several components of the gradient in directions orthogonal to the CC or Cc direction do not vanish and so the effective gradient is non-zero although greatly diminished in magnitude. In the memory one stochastic IPD, trajectories that would have ended at `TFT` type agents with no noise continued to evolve into more and more generous varieties (Pr(Cd) increases). By natural selection several of the components of these agents continue to change but at a much slower rate. This fact reduces the size of the cooperative basin for stochastic agents considerably. For example, after the `TFT` type agents in the stochastic memory one IPD slowly evolved into very generous varieties (larger Pr(CD)), they became so forgiving that cheating variations were able to invade. Then the course of evolution (trajectory) changed radically and often approached an Alternator agent. (Kraines & Kraines 2000).

For this paper, we follow many trajectories in the stochastic APD of *k*-ply agents for $k \leq 4$. For a large set of initial stochastic *k*-ply APD agents, randomly chosen with uniform distribution in the stochastic trait space **H**, the trajectories have been calculated for a variety of step sizes, noise levels, and generations. In addition to the cooperating and defecting regions, a number of trajectories approach stable agents either interior to the hypercube or on "alternating" or cyclic basins. By choosing a large number of initial agents, we gain confidence that we have found all basins of attraction with the standard payoff and noise level. The size of these regions depends strongly on the size of the memory.

Trajectories typically reach an ε or 1−ε face of **H** after a few hundred generations and then move along this face for many more generations. Some trajectories will veer away from the face eventually toward an entirely different basin. Only 11% of trajectories converged to cooperative agents in the stochastic memory one IPD and almost all of these approached `Pavlov`. By contrast, nearly 29% of the trajectories in the 2-ply APD remain cooperative.

	APD with noise=0.01		
	D basin	C basin	others
1-ply	65%	20%	15%
2-ply	71%	29%	0%

3-ply	60%	30%	10%
4-ply	41.4%	38.6%	20%

Table 4 Sizes of the defector/cooperator basins for stochastic *k*-ply strategies
These figures are based on 2000 simulations of 4000 generations each.

Table 4 shows that more intelligent species, i.e., those with greater memory, are more likely to evolve protocols for mutual cooperation. Details are provided in the following sections.

8.1 Convergence for 1-ply agents

Reactive strategies have been analyzed by Nowak & Sigmund (1990) for the simultaneous game and Nowak & Sigmund (1994) and Frean (1994) and (1996) for the alternating game at least for "decomposable" payoff matrix, i.e., for R+P = S+T. For the standard payoff [1 –2 2 –1], the line $y = x - \frac{T-R}{T+R} = x - \frac{1}{3}$ that separates the trait space (unit square) into a lower triangle and an upper pentagonal region are equilibrium points for the dynamical system. Trajectories move along circular arcs with center `TFT` = (1, 0) as in Figure 1. Agents in the region of the square of area $1 - \frac{\pi}{9} = 0.6509$ that lies outside the quadrant of the circle of radius $\frac{2}{3}$ and center (1,0) evolve toward defectors with y = 0 and $x < \frac{1}{3}$. Agents in the region of the square of area $\frac{1+\frac{\pi}{4}}{9} = 0.1983$ that consists of the region inside the quadrant of the circle about (0, 1) of radius $\frac{\sqrt{2}}{3}$ together with the inside the triangle with vertices (1,0), ($\frac{2}{3}$, $\frac{1}{3}$) and (1, $\frac{2}{3}$) tend toward cooperators with x = 1 and $y < \frac{2}{3}$. Agents starting in the remainder region of area $\frac{\frac{3\pi}{4}-1}{9} = 0.1507$ stabilizes along the line $y=x-\frac{1}{3}$.

In the stochastic APD, trajectories that reach (x, ε) will continue toward the fallible but stable agent `AllD` = (ε, ε). Similarly, after reaching the cooperative (1–ε, y) the trajectory will asymptotically approach `GTFT` = (1–ε, $\frac{2}{3}$ –ε). This later agent is not stable in the dynamical system. If $y > \frac{2}{3} - \varepsilon$, then (1–ε, y) will evolve toward `AllD`. The sizes of the cooperative and defector basins in the stochastic trait space are essentially the same as that in the noiseless trait space.

8.2 Convergence for 2-ply agents

Analytic methods are far more difficult to carry out for *k*-ply agents with k > 1. Using the migrant evolution model, Nowak & Sigmund (1994) and Frean (1996) found that 2-ply `FBF` type agents of the form (1–ε, ε, 1–ε, x) tended to emerge after many generations. Our natural selection approach confirms that such agents with x approaching $\frac{2}{3} - \varepsilon$ are the asymptotic limits of most cooperators. Roughly the same 30% of the cooperating trajectories in the noiseless APD will still cooperate with 0.01 noise level.

The alternating 2-ply agents (ε, ε, 1−ε, 1−ε), (*x,* ε, ε, 1−ε) and (ε, 1−ε, 1−ε, *x*) are each unstable with trajectories from the first and third going to `AllD` and from the second going to an `FBF` type. Not all agents end in the defecting or the cooperative basin. The analogue of the attracting equilibrium line in the 1-ply trait space is a plane of agents of the same form as that studied in Kraines & Kraines (2000) in the memory one IPD. This plane is the attractor for only about 1% of all agents in the trait space.

8.3 Convergence for 3-ply agents

As seen in Table 3, in the trait space of 3-ply agents with no noise, about 54% of the trajectories become defectors, 42% cooperators and most of the remaining fall into stable alternating patterns, chiefly CcDcCdCc.. with a payoff of $\frac{1}{3}$. More than 75% of the Pr(cCc) = 1 face consists of stable cooperators. With noise ε = 0.01, about 10% of the cooperating trajectories in the noiseless APD become too generous and eventually fall into the Alternator or Defector basin. A relative few reach a metastable extension of the unstable 2-ply alternator agents. Of those that remain cooperative after 4000 generations, the average of the final agents is

$$\begin{pmatrix} 0.99 & 0.08 & 0.44 & 0.76 \\ 0.98 & 0.14 & 0.96 & 0.48 \end{pmatrix} \tag{8.1}$$

This result is in general agreement with the simulations of Hauert & Schuster (1998) in the migrant evolution model.

Two families of this form are particularly effective in the trait space of stochastic 3-ply agents. The combination of `Pavlov` and `TFT` , **Firm Pavlov**

$$\mathtt{FP}(w) = \begin{pmatrix} 1-\varepsilon & \varepsilon & \varepsilon & 1-\varepsilon \\ 1-\varepsilon & \varepsilon & 1-\varepsilon & w \end{pmatrix} \tag{8.2}$$

and the less exploiting variety **Tough Love**

$$\mathtt{TL}(w) = \begin{pmatrix} 1-\varepsilon & \varepsilon & 1-\varepsilon & 1-\varepsilon \\ 1-\varepsilon & \varepsilon & 1-\varepsilon & w \end{pmatrix} \tag{8.3}$$

differ in their response to the sequence cDc. Both will reciprocate cooperation (Pr(xCc) = 1−ε), punish unprovoked defection (Pr(xCd) = ε), contritely cooperate if punished after originally defecting (Pr(cDd) = 1−ε) and forgive a contrite opponent (Pr(dDc) = 1−ε). Each agent is moderately firm after a sequence of defections if it is not certain who "started it" (Pr(dDd) = *w*). After "accidentally" defecting, Firm Pavlov will exploit overly generous agents (Pr(cDd) = ε) while the guilt-ridden Tough Love will cooperate (Pr(cDd) = 1−ε) .

Neill (2001) suggested optimality criteria for stochastic *k*-ply agents in the APD. "In order to achieve success against a variety of other strategies, a strategy must be ``self-cooperating'' (able to achieve mutual cooperation with its clone), ``C-exploiting'' (able to exploit unconditional cooperators), and ``D-unexploitable'' (able to resist exploitation by defectors).... [The] ``Firm Pavlov'' strategies ... not only meet our stringent optimality criteria, but also achieve remarkable success in round-robin tournaments and evolutionary interactions. These higher memory strategies are friendly enough to cooperate with their clone, pragmatic enough to exploit

unconditional cooperators, and wary enough to resist exploitation by defectors: they are truly ``optimal under noise" in the Alternating Prisoner's Dilemma."

An optimal agent should exploit altruists not just for its immediate selfish gain. The prime example of an otherwise optimal strategy is the fallible `TFT`. Although a small proportion of them may replace `AllD`, since `TFT` is not C-exploiting, a small population of them will be taken over by `AllC` agents or will evolve into overly generous versions. The resulting population becomes susceptible to invasion by `AllD` with all eventually losing (Boyd & Lorberbaum 1987). See Kraines & Kraines (1993) for similar optimality criteria

A trajectory starting at either `FP`(w) or `TL`(w) for $w < 0.6$ will converge toward the `TL`(w_o) agent $\begin{pmatrix} 1-\varepsilon & \varepsilon & 1-\varepsilon & 1-\varepsilon \\ 1-\varepsilon & \varepsilon & 1-\varepsilon & w_o \end{pmatrix}$ with $w_o = \frac{2}{3} - \varepsilon$ up to terms of order ε^2. If w is initially greater than w_o, i.e., for agents more like the fallible `FBF`, cheaters can invade and the trajectory turns toward the defecting basin.

9. Convergence for 4-ply agents

For the standard payoff G and with $\varepsilon = 0.01$ noise level and $\delta = 0.008$ step size, trajectories starting from 1000 random initial agents were tracked for 4000 generations each within the 16 dimensional hypercube. Another group of 1000 trajectories with the same noise and twice the step size was tracked for 2000 generations. Results in both cases were similar. Trajectories with the same initial values were also tracked in the noiseless ($\varepsilon = 0$) trait space. Further simulations with different levels of noise and greater number of generations extend and substantiate the main results.

In each case, roughly 40% of random agents evolved into defecting strategies as determined by payoffs below -0.9 and 40% become cooperators with payoffs above 0.9. About 15% trajectories converge toward one of three stable or metastable societies of agents that alternate cooperation and defection with self-payoff $-\frac{1}{3}$, 0 or $+\frac{1}{3}$ and that cannot be easily invaded by nearby agents. The remaining 5% were steadily evolving and would reach one basin or another within a few thousand more generations.

As noted in section 3.2, after the inevitable but rare error dislodges the agent `Pavlov` ($1-\varepsilon$, ε, ε, $1-\varepsilon$) out of a run of mutual cooperation with its clone, it enters a cycle of win-lose, lose-win interactions CdDcDdCd... with expected payoff $-\frac{1}{3}$ to each. A 4-ply agent that satisfies $\Pr(CdDc) = \varepsilon$, $\Pr(DdCd) = \varepsilon$ and $\Pr(DcDd) = 1-\varepsilon$ may also get stuck in the same win-lose CdDcDdCd... cycle, at least if $\Pr(CcCc) < 1-\varepsilon$. These negative 3 cycle agents will receive an average of $\frac{S+T+P}{3}$ up to terms of order ε. Dual positive 3 cycle agents with payoff of the order of $\frac{R+S+T}{3}$ occurs in 4-ply agents with $\Pr(DcCd) = 1-\varepsilon$, $\Pr(CcDc) = 1-\varepsilon$ and $\Pr(CdCc) = \varepsilon$. About 13% of the trajectories were identified as converging to a positive or a negative 3 cycle agent. In addition, a relatively few (2.5%) trajectories approach the metastable 4 cycle CcDdCcDd...agent with an average payoff of $\frac{R+S+T+P}{4}$.

Of the 40% of cooperative trajectories in the stochastic trait space, nearly all have converged toward recognizable patterns during the first thousand generations or less. Cooperation requires

that Pr(CcCc) = 1−ε. To minimize the chance of exploitation by cheating descendants, most cooperators will almost surely retaliate after an unexpected breach of trust, i.e., Pr(CcCd) = ε and so a typical sequence of plays with an error starts out CcC**d**D. As we saw earlier, the 4-ply extensions of `TFT` and `Pavlov` in the stochastic APD will get into a rut of mutual defection or win-lose cycles and so will eventually degenerate.

With a few exceptions, once the self-payoff to an agent on a trajectory is below -0.9, the trajectory continues toward `AllD`. In trajectories for cyclic or alternating agents, the secondary entries often continue to change. Many alternating trajectories will tend to remain quite stable for up to several hundred thousand generations and then enter the D basin.

10. Cooperative Protocols of 4-ply agents

In the stochastic simulations with ε = 0.01, the trajectories leading to cooperative agents, as determined by payoff > 0.9, tend to separate into three major and one minor *clans* each using different *protocols* or conventions for restoring cooperation after an unintentional defection. Agents score higher when they interact among their own clan than with those using different protocols. This implies that none of the clans can invade any other. See Table 8 below.

The largest clan uses variants of the Firm But Fair (`FBF`) strategy. In an APD with its clone, `FBF` retaliates for an unprovoked defection and then both parties resume cooperation. The Two Tits for a Tat (`2TFT`) agent is simultaneously harsher than `FBF` against one who cheats, whether or not intentionally, and more contrite or forgiving of retaliation against it for defecting. Another protocol for cooperation, called the **Grim Pavlov** (`GP`) strategy, results in a sequence of defections of average length 4 when an error is made. The rarest of the cooperative families, **Clemency** (`Clem`), threatens a possible delayed retaliation to the initial defection.

	%	normal route use to restore mutual cooperation	expected self-payoff
`FBF`	62	CcC**d**DcCc	0.9595
`GP`	23	CcC**d**Dd...DcCc ...	0.9411
`2TFT`	9	CcC**d**DcDcCc	0.9412
`Clem`	6	CcC**d**CcCcCc (✎ probability)	0.9687
		CcC**d**CcDcCc (🗎 probability)	

Table 5 relative percent of cooperators that belong to a particular clan. The normal route that agents use to restore mutual cooperation after an accidental defection is the distinguishing feature of each clan.

We now consider each of these protocols in more detail. The fallible 4-ply `FBF` agent is $\begin{pmatrix} .99 & .01 & .99 & .99 \\ .99 & .01 & .99 & .99 \\ .99 & .01 & .99 & .99 \\ .99 & .01 & .99 & .99 \end{pmatrix}$. As Frean (1994) observed, the protocol of `FBF` will quickly restore cooperation with its clone after just one tit for tat exchange. After the mistaken defection, the

sequence becomes CcDdCc with corresponding payoffs to each of T+P+S = S+P+T = -1 rather than 3R = 3 for a net loss of 4. As such a sequence occurs with probability ε, the expected payoff to each is 1−4ε up to ε^2 terms. With probability $4\varepsilon^2$, the agents will make a second mistake before cooperation is completely restored.

Few of the entries need to be exactly ε or 1−ε for this main recovery route to apply. By the FBF clan, we mean all agents using strategies similar to the form

$$\begin{pmatrix} 0.99 & 0.01 & * & 0.99 \\ * & * & 0.99 & * \\ 0.99 & * & * & * \\ 0.99 & * & * & * \end{pmatrix}$$ with typical recovery pattern CcC**d**DcCcC (10.1)

More precisely, agents with values within 0.1 of the specified values are lumped into the FBF clan. The * denotes entries that range over values with large standard deviations. Of the 2000 initial random agents traced, 24.4% of the trajectories reached agents that share the basic structure of FBF after 4000 generations.

We now consider the advantages and disadvantages of the special varieties

$$\text{FBF}(u,v,x,y) = \begin{pmatrix} .99 & .01 & u & .99 \\ .99 & .01 & .99 & x \\ .99 & .01 & v & .99 \\ .99 & .01 & .99 & y \end{pmatrix}$$ in more detail. (10.2)

The agents FBF(*u,v,x,y*) are trusting but vengeful. They are contrite when reprimanded (Pr(XcDd) = 1−ε) and forgiving to those who are contrite (Pr(XdDc) = 1−ε). Abbreviate FBF(*x*) = FBF(*x,x,x,x*) so that FBF(1−ε) is the fallible FBF. Among the FBF type agents, the 4-ply generalizations of the 3-ply families of **Firm Pavlov** agents FP(y) = FBF(ε, ε, ε, *y*) and **Tough Love** agents TL(*x,y*) = FBF(1−ε, 1−ε, *x, y*) are particularly effective. They balance greater resistance to AllD with rapid resumption of cooperation with their clones after a series of accidental defections than their 3-ply cousins.

An optimal agent exploits altruists, resists egoists and resumes mutual cooperation with its own type after any series of errors. Although all varieties of FBF will invade and dominate AllC, those with smaller *u* and *v* will do so faster while avoiding the potentially destructive genetic drift by unconditional altruists. As in Table 6, agents with smaller *x* and *y* are less exploitable by AllD. On the other hand, agents with smaller parameters take longer to resume cooperation with their clone and thus are replaced by more cooperative offspring with larger *u* and *v* and corresponding higher self-payoff.

FBF(*u, v, x, y*)	*u*	*v*	*x*	*y*	self payoff	dom(FBF|AllD)
FP(ε)	ε	ε	ε	ε	0.9416	0.9984
TL(🗁, ε)	1 − ε	1 − ε	$\frac{1}{2}$	ε	0.9513	0.9851
TL(ε, 🗁)	1 − ε	1 − ε	ε	$\frac{1}{2}$	0.9604	0.8387
TL(👌, 👌)	1 − ε	1 − ε	$\frac{2}{3}$	$\frac{2}{3}$	0.9607	0.6553

FBF(1-ε)	1 – ε	1 – ε	1 – ε	1 – ε	0.9608	0.4847

Table 6 Varieties of FBF and domination measure against AllD for ε = 0.01

An ideal strategy A should not only resist invasion by AllD, but should be able to entirely replace a minority of AllD agents in the population, i.e., dom(A|AllD) should be close to 1. TFT is very successful in this regard; as noted in section 4, dom(TFT|AllD) = $1 - \frac{\varepsilon}{1-2\varepsilon}$ in the stochastic APD. The dominance index dom(FBF(*u,v,x,y*)|AllD) depends greatly on the secondary components. For the fallible Firm But Fair agents, dom(FBF (1–ε)|AllD) = 0.5–ε = 0.49 while dom(FP (ε)|AllD) = $1-\frac{1}{2}\varepsilon = 0.995$ up to terms of order ε^2 so FP (ε) dominates AllD with an even smaller toe-hold than TFT.

Unlike TFT, FP(ε) has a quite respectable self-payoff of about 1–6ε and will dominate both the undiscriminating altruist AllC and even TFT itself. FP(*y*) will quickly resume cooperation with its clone after any series of random moves (lots of noise). Moreover, Neill (2001) shows that it fares well in interactions with a very wide variety of other agents in the trait space. The only serious drawback is that FP(ε) will evolve into generous versions with greater Pr(CcCd), slightly increasing its self-payoff and recovery time but simultaneously greatly decreasing its advantage over AllD, AllC and TFT. One might say that successful offspring develop a sense of guilt that will encourage resumption of cooperation after an accidental defection. But this in turn makes their descendants more vulnerable to outsiders.

The varieties of FBF(*u, v, x, y*) with large *u* and small *x* and *y* evolve toward a stable TL(*x, y*) strategy with *y* = .6566 and *x* between .01 and .65. If *x* and *y* are initially large, then the trajectory decays toward the D basin. The boundary between the cooperative basin and the defecting basin is a curve in the (*x, y*) square going through the points (.99,.01), (.77, .3), (.6566,.6566) and (.64, .99). The value of *v* has less significance.

These and related results suggest that the variety TL($\frac{2}{3} - \varepsilon$, $\frac{2}{3} - \varepsilon$) of the FBF family forms the ideal clan of 4-ply agents in the stochastic APD. TL($\frac{2}{3} - \varepsilon$, $\frac{2}{3} - \varepsilon$) dominates FP(🗁) and FBF($\frac{1}{2}$). A value of about $\frac{2}{3}$ for *y* = Pr(DdDd) is sufficiently low to protect TL($\frac{2}{3} - \varepsilon$, $\frac{2}{3} - \varepsilon$) against AllD while high enough to minimize the length of any mutual defection rut with one's clone. Although FP has a higher dominance index than TL($\frac{2}{3} - \varepsilon$, $\frac{2}{3} - \varepsilon$) against AllD, it has a smaller self-payoff. As noted in sections 8.3 and 14, the evolutionary trajectory from FP tends toward TL type agents while the variety TL($\frac{2}{3} - \varepsilon$, $\frac{2}{3} - \varepsilon$) is very close to being evolutionarily stable against nearby agents.

Grim Pavlov or GP is a variety of cooperators that emerges from about 9% percent of all trajectories. If GP accidentally defects and his opponent retaliates, GP will show no contrition but will defect again. From then on, each agent will defect with a moderate probability until one of them cooperates, after which mutual cooperation is restored. The basic structure of a GP(*x,y*) agent is

$$\mathtt{GP}(x,y) = \begin{pmatrix} 0.99 & 0.01 & * & 0.01 \\ * & * & * & x \\ 0.99 & * & * & * \\ 0.99 & * & 0.99 & y \end{pmatrix} \tag{10.3}$$

with typical recovery pattern CcC**d**Dd...dCcCc. The * entries do not significantly affect the strategy or the payoff. The expected number of defections between two GP agents before cooperation is restored can be computed as the infinite series

$$3 + (1-x)y \sum_{k=1}^{\infty} k(1-y)^k = 3 + \frac{1-x}{y}, \text{ up to terms of order } \varepsilon. \tag{10.4}$$

Although this agent arises naturally in the natural selection model and cannot be invaded by nearby agents with lower Pr(CcCc), GP is the only cooperator clan that can be invaded by AllD. For example, since $w(\mathtt{AllD}\,|\mathtt{GP}(\frac{1}{2},\frac{1}{2})) = .9627 > w(\mathtt{GP}(\frac{1}{2},\frac{1}{2})\,|\mathtt{GP}(\frac{1}{2},\frac{1}{2})) = .9411$, $\mathrm{dom}(\mathtt{GP}(\frac{1}{2},\frac{1}{2})\,|\mathtt{AllD}) = 0$.

Two tits for a tat or 2TFT is a vindictive variant of FBF that punishes a single unprovoked defection with two defections, yet is contrite enough to accept the two penalties if it makes the first defection. The typical route to restoring cooperation is CcC**d**DcDcCcC. A clan of such agents averages a self-payoff of 0.9416, somewhat lower than that of the FBF clan and about the same as GP. Like the pure FBF and GP, the 2TFT agents must be in a majority in order to drive out AllD in a dimorphic population. 2TFT is somewhat more exploiting of unconditional altruists than FBF.

Of 2000 initial random agents, 3.4% satisfied the basic conditions for 2TFT with average structure

$$\mathtt{2TFT} = \begin{pmatrix} 0.99 & 0.01 & * & 0.99 \\ 0.99 & * & 0.01 & 0.01 \\ 0.99 & * & 0.99 & * \\ * & 0.99 & * & * \end{pmatrix} \tag{10.5}$$

with values of * having large variance and little effect on the payoff.

Immediate retaliation for an unprovoked defection is not an absolute requirement to maintain a relatively stable cooperative society. Out of 2000 initial populations of strategies, 30 (1.5%) evolved into the moderately tolerant and forgiving strategy Clemency. Clem cooperates after the first defection by his opponent and if Clem accidentally defects and his opponent cooperates, then Clem will not try to exploit but will cooperate in turn. To maintain some modest protection against invasion by cheating varieties, the protocol calls for the threat of a belated retaliation.

A typical representative of Clem is the following.

$$\texttt{Clem} = \begin{pmatrix} 0.99 & 0.99 & 0.99 & * \\ 0.66 & * & * & * \\ 0.99 & 0.99 & * & * \\ * & * & * & * \end{pmatrix} \tag{10.6}$$

with typical recovery pattern of CcC**d**CcCcCc with 66% probability and CcC**d**CcDcCc with 34% probability. As before, * denote various probabilities with large standard deviation and little effect on the payoff.

Because agents using `Clem` rarely defect, they have a very high average payoff in a match with each other. During the period that this dynamical system was in the generous realm, the total payoff approached 0.9687, the highest among the various stable and metastable protocols. This protocol is just firm enough to give the same level of protection, or dominance index, against `AllD` as `FBF` and `2TFT`; those in the majority are able to drive out the other in a dimorphic population. `Clem` is somewhat less exploiting of unconditional altruists than the retaliatory clans.

11. Recovery time

Another measure of the effectiveness of a cooperative strategy is the number of steps, after a mistake or even a series of mistakes, that it takes for the agent and his clone to resume mutual cooperation (return to the CcCc state). To compute this *recovery time* for agents in the stochastic APD, let Q be the 15x15 *fundamental* matrix formed from the regular transition matrix T_A by eliminating the first row and first column (corresponding to CcCc). Then (I–Q) is invertible and the (*i*,*j*) entry of its inverse (= I + Q + Q^2 + ...) represents the probability of moving from state *i* to state *j* in some number of plies *without* first going through CcCc (Kemeny & Snell 1976). The sum of the column entries gives the answer we seek.

	recovery time from CcCd	max recovery time	from state
`FBF`($\frac{1}{2}$)	1.25	4.09	XdDd
`FP`($\frac{2}{3}$, $\frac{2}{3}$)	1.16	2.13	XxCd
`TL`($\frac{2}{3}$, $\frac{2}{3}$)	1.13	1.13	XxCd
`GP`	2.24	3.80	XdCd
`2TFT`	3.23	3.56	XdDd
`Clem`	1.78	6.23	XcDc

Table 7 Recovery time for 4-ply clans

For all varieties of `FBF`, the ordinary route for restoration of cooperation after an initial error, CcC**d**DcCcC.., takes just one more defection unless a second error occurs. A reasonably stable strategy must include a protocol for the ε^2 eventuality that two errors occur in close succession. For example, instead of retaliating, the opponent will "accidentally" cooperate with probability ε. Then with probability *u* = Pr(cC**d**C), cooperation is immediately restored while with probability 1–*u*, the sequence of plays most likely continues as cC**d**CdDcCcC. If the second error occurs after the original defector fails to show contrition, CcC**d**D*d*, then the agents get into a sequence of defections considered in (9.4) whose length is $\frac{1-x}{y}$. Other double error sequences that must be

considered are CcC**d**Dc*D*dCcC, CcC**d**DcC*d*DcC, and CcC**d**DcCc*D*dCcC. The expected number of additional defections if a single error is made soon after CcC**d** is seen to be $(2-3u)+(1+\frac{1-x}{y})+3+4+5=15-3u+\frac{1-x}{y}$ up to terms of order ε^2. Thus recovery time from state CcCd is $5+\varepsilon(15-3u+\frac{1-x}{y})$ with v basically irrelevant.

The recovery pattern for the successful agent TL(x,y) is represented graphically in Figure 2. The number at each state position is the component of the corresponding equilibrium vector for the Markov process. More precisely, this number is the percentage of rounds in which TL (x,y) is in that state in an infinite game with its clone. For example, the TL agents are in the state of mutual cooperation CcCc over 95% of the game and they are in each of the five states along the main recovery route, CcCdDcCcC, for 0.96% of the game. The varieties of FBF are distinguished by how the agents react after a secondary error that occurs with probability ε^2. Such a double error among TL(x, y) agents will push the players into paths that will quickly connect back into the main recovery route. The extra arrow leaving CdDd and the loop at DdDd indicate a likely branching that depends on the particular values of x and y. States CdCd and DcDc are only visited after an appropriate sequence of three errors. This occurs with probability ε^3.

Mutual cooperation is resumed among FBF ($\frac{1}{2}$) agents after just 1.25 moves from CcCd but is expected to take over 4 steps if FBF($\frac{1}{2}$) and his clone get into state XdDd as a result of multiple errors (see Table 7). Agents of the TL(x, y) types are clear winners in this contest since no matter what string of errors occurs, cooperation is restored in not much more than one move. To minimize recovery time, it is best for the agent to have $x = 1-\varepsilon$, i.e., TL($1-\varepsilon, y$) is optimal in this regard. On the other hand, this agent is somewhat more vulnerable to AllD.

The Clemency strategy is relatively unstable in an evolutionary sense. The trajectory will often turn toward the defecting basin after a few thousand generations. Part of the reason is that the recovery route after a sequence of two errors can be quite long. If Clem makes an accidental defection then, with $\frac{2}{3}$ probability, cooperation is restored immediately, but with $\frac{1}{3}$ probability cooperation takes at least four more steps CcC**d**CcDcCcC... If both Clem and his clone accidentally defect in close succession, the recover time increases to 6.23. See also section 14.

12. Evolutionarily stable clans

Assume that several different trajectories evolved independently into different clans each consisting of agents of the type FBF, GP, 2TFT or Clem. If these clans were to merge, it is reasonable to ask which would dominate. As shown in Table 8, the self-payoff of an agent against a clan member is always greater than that against one from a different clan. This implies that no clan can invade any other in the sense of Maynard Smith & Price (1973).

In anthropomorphic terms, one may envision a population that has a genetic propensity to reciprocate trust and retaliate against cheaters. These traits are necessary but not sufficient to ensure cooperation. In addition, different clans or cultures within the population must develop cultural mores or memes to quickly restore cooperation if cheating, intentional or not, ever occurs. Different societies may have social customs that smooth relations among themselves but incidentally make it more difficult for outsiders with their own customs. The cooperative clans FBF, GP, and 2TFT share the same strict justice; don't cheat but do retaliate for misdeeds.

They differ in the methods for restoring cooperation after an accidental defection. A member of FBF is contrite and forgiving. A member of GP is somewhat belligerent after he defects and is punished, but yet avoids a lasting feud. A member of 2TFT will punish a defector more severely while meekly accepting the punishment were he to defect first. A Clem member will overlook a misdeed on most occasions but retains the option to retaliate at a later time.

In Table 8A, the payoffs to row agents from column agents are given for the typical representatives of the four clans FBF, GP, 2TFT and Clem as well as AllD and AllC. In Table 8B, the number represents the proportion of row agents in a dimorphic population that is necessary to dominate the column agents. For example, in a population with more than 38% GP and less than 62% FBF, the GP agents will replicate faster and replace the FBF.

Payoff	FBF	GP	2TFT	Clem	AllD	AllC
FBF	0.9593	0.9031	0.8832	0.9230	-1.4713	0.9991
GP	0.936	0.9411	0.9006	0.9219	-1.6276	1.0079
2TFT	0.9242	0.8762	0.9412	0.9233	-1.4688	1.0036
Clem	0.8963	0.9017	0.8797	0.9687	-1.4773	0.9873
AllD	0.4938	0.9627	0.4865	0.4992	-0.98	1.96
AllC	0.9227	0.8963	0.9093	0.9693	-1.96	0.98

Table 8A Round robin tournament among various 4-ply clans

Bistable index	FBF	GP	2TFT	Clem	AllD
FBF	-	0.6194	0.6229	0.4205	0.5135
GP	0.3806	-	0.3853	0.5426	x
2TFT	0.3771	0.6147	-	0.4243	0.5181
Clem	0.5795	0.4574	0.5757	-	0.5123
AllD	0.4865	x	0.4819	0.4877	-

Table 8B Bistable index among various 4-ply clans

One drawback of *k*-ply agents is that because agents have limited memory, when presented with the DdDd state, it is not clear who "started it". To overcome this problem, Neill (2001) introduced the notion of a *first defector* strategy that *remembers* who first defected in a long string of defections - just one extra bit of information. But this is outside the trait space of *k*-ply agents and is not considered here.

13. Rate of evolution

In the discrete dynamical system, each agent is replaced by the most successful offspring in the next generation. By the rate of evolution of strategies we mean the Euclidean distance between an agent and its surviving offspring. As described in section 6, this variation is equal to $K\delta$ where K is the magnitude of the effective gradient of the payoff function and δ is the (adjustable) step size of the discrete dynamical system. For random agents in the stochastic hypercube, i.e., for initial points of trajectories, the gradient and the effective gradient are most often the same with

magnitude averaging 0.98. Initially, each of the components of the strategy will change monotonically with most changing much more slowly than the maximum rate of Kδ. If a trajectory converges toward a cooperative agent in the noiseless APD then, when the *mutual trust* index m = Pr(CcCc) reaches 1, the other components of the gradient vanish and the trajectory terminates at a (locally) evolutionarily stable state with self-payoff of 1. This implies that the effective gradient K = 0.

With $\delta = 0.01$ and noise level $\varepsilon = 0.01$, 100 random initial stochastic 4-ply agents were tracked for 4000 generations, the magnitude of the rate of change was computed. The trajectories leading to cooperators often followed a similar pattern with m = Pr(CcCc) increasing most rapidly of all components. As m approached the (stochastic) mutual cooperation level of $1-\varepsilon$, the rate of evolution Kδ increased by a factor of up to 10 often within the first 50 generations. After m reached the maximum of $1-\varepsilon$, evolution ceased in that direction. The magnitude of the effective gradient and the rate of evolution may discontinuously drop to as little as 3% of the former level. Between occasional discontinuities, K tends to gradually decrease, often reaching a level of 0.00035 after 10,000 generations. Such slowly evolving agents will be called ε stable.

In figure 3 the rate of growth of an initial 4-ply agent is plotted for its first 100 generations. After an initial drop, the rate of growth rapidly increases to a peak of 0.027 distance between parent and offspring and then instantly falls to a rate of 0.0009. It continues to decreases with some discontinuities reaching 0.0003 after 1000 generations.

The standard representatives for the four cooperative clans in Appendix A are better than ε stable. The protocols for restoring cooperation after a series of errors require that between 5 and 10 components of the agents have stabilized at the ε or $1-\varepsilon$ boundary greatly reducing the magnitude of the effective gradient. Most varieties of `FBF` evolve at the rate of about 0.0002 requiring roughly 500 generations for any component to change by 0.1. The agents `GP`(0.5, 0.5) and `2TFT` each evolve at a rate on the order of 0.00001 requiring roughly 10,000 generations for any component to change by 0.1. For general step size δ and noise level ε, the rate of evolution for representatives of the retaliatory clans is roughly proportional to $\delta\cdot\varepsilon^2$. It is closer to $\delta\cdot\varepsilon$ for `Clem`.

14. Further evolution of cooperative agents

As described in section 8.1, trajectories among stochastic 1-ply agents can be shown analytically to converge to the fallible `AllD`, the stochastic 1-ply `GTFT` = $(1-\varepsilon, \frac{2}{3}-\varepsilon)$ or an agent on the equilibrium line. Similarly, most 2-ply and 3-ply trajectories asymptotically approach either `AllD` or a cooperative limiting agent of the type `FBF`$(\frac{2}{3}-\varepsilon)$.

In contrast, the four cooperative clans of stochastic 4-ply agents are rarely completely stable in the adaptive dynamical system; secondary entries that correspond to reactions after two errors in close succession will change at the rate of ε^2. Trajectories starting at various representatives of cooperative clans exhibit many different behaviors. The trajectories starting from most representatives of the cooperative clans of stochastic 4-ply agents will often reach critical points and then radically change course. If one of the secondary probabilities passes a critical value, then it pays for the opponent to switch from D to C or vice versa. At that juncture, the trajectory rapidly changes course, sometimes toward the defecting basin, sometimes bouncing back into a possibly different cooperative variety, and sometimes entering a metastable alternating cycle of cooperation and defection. Even if a protocol is stable with respect to two errors, the ε^3 chance

that three errors occur in succession will cause a few of the entries to change and can lead eventually to radical changes.

Although overly altruistic agents suffer in the stochastic 4-ply APD, `Clem` threatens just enough of a deterrent to protect itself from invasion by cheaters. The key component of `Clem` is the *forgiveness* index f = Pr(CdCc). If a basically cooperative agent `A` defects against `Clem`, then `Clem` almost surely cooperates on the next play and forgives the defection entirely with probability f, i.e., up to terms of order ε, the sequence of choices is cC**d**CcC...with probability f, and the payoff to `A` is R+T+T+R+R+R rather than the safe 6R that `A` would have gained had he not defected. This potential gain to `A` for cheating of 2T–2R = +2 must be balanced by the potential cost to `A` if `Clem` carries out his threat to defect on the second round. Up to terms of order ε, the sequence cC**d**CcDcC.. occurs with probability $(1-f)$ with payoff R+T+T+R+S+S+R rather than 7R if `A` had not defected for a potential loss of 2T+2S-4R = -4. Thus, up to terms of order ε, it pays to cheat `Clem` only if

$$f(2T-2R) + (1-f)(2T+2S-4R) > 0, \text{ i.e., } \quad f < \frac{2R-T-S}{R-S} = \frac{2}{3}. \tag{14.1}$$

A more careful analysis shows that the critical value for f equals $\frac{2}{3} - \varepsilon$ up to terms of order ε^2.

The gradient driving the adaptive dynamical system starting at `Clem` has only positive components and so the successor agents are somewhat more cooperative than their ancestors and have slightly higher self-payoff. If f is less than its critical value of $\frac{2}{3} - \varepsilon$, more forgiving varieties with larger f will have a slightly larger payoff and will gradually replace the existing varieties. Eventually f exceeds its critical value and the trajectory rapidly degenerates toward `AllD`. The average number of generations before this degeneration was 8200 with a standard deviation of 2700 generations. In one case, degeneration took 15,000 generations. Thus too much "turning of the other cheek" will eventually lead toward an overly altruistic society that becomes open to invasion by cheaters.

The evolutionary trajectory starting at the average `FBF` (0.5, 0.5, 0.5, 0.5) is relatively stable for many more generations, but eventually it too degenerates. Slightly less exploiting agents with greater Pr(CcDc) will score somewhat higher and the trajectory reaches the `TL` type `FBF`(0.99,0.512,0.664,0.619) after about 375,000 generations. After agents along the trajectory become sufficiently cooperative, an opponent who defects after CcDd can more easily get away with a slight gain. In 45000 more generations, Pr(CcDd) dips below $\frac{2}{3}$. Then less vengeful variants of `FBF`, with larger *generosity* g = Pr(CcCd) will tend to avoid the CcDd state and thereby will score marginally higher against their parents. In the evolutionary trajectory starting at `FBF`, g increases at a relatively rapid rate of 0.1 per 2000 generations. When g exceeds $\frac{2}{3}$ after about 13200 generations, the trajectory has become a very generous variety of `FBF`:

$$\begin{pmatrix} .99 & .686 & .99 & .551 \\ .99 & .01 & .99 & .469 \\ .99 & .168 & .99 & .987 \\ .99 & .01 & .99 & .836 \end{pmatrix}$$ with self payoff increasing from 0.9604 to .9674.

With such a weak tendency to retaliate, if an opponent `A` cheats, then with probability g, the sequence will be CcC**d**CcC.. and the cheater `A` gains a net of 2T–2R = 2 for his cheating while

with probability 1–*g*, cooperation will take longer to resume. The expected payoff to A will depend on the precise value of *g* and on the values of Pr(XxDd) but is in the range of -4. As the trajectory moves toward more generous agents, eventually *g* exceeds a critical value and the expected gain to an offspring A who cheats outweighs the expected long term cost of his action. Once this critical juncture is passed, in just 1100 more generations the cheaters dominate the population and the self-payoff plummets to -0.979.

In section 8.3, we noted that the 3-ply FP evolves into a variety of TL. In a similar manner, trajectories starting at varieties of the 4-ply FP pass through varieties FBF sharing the essential features of TL with Pr(CcDc) increasing nearly linearly at a rate of about 0.1 per 50,000 generations. In contrast with the 3-ply situation, trajectories starting at some varieties of the 4-ply FP become unstable and eventually approach AllD.

Not all changes in trajectory turn from cooperation to defection. In its evolutionary trajectory, the midrange entries of 2TFT change gradually and linearly for about 100,000 generations, some increasing and others decreasing at the rate of ε^2. Starting with all * = 0.5 in 2TFT (see 10.5), the agents evolve into 2TFT* = $\begin{pmatrix} 0.99 & 0.01 & 0.77 & 0.99 \\ 0.99 & 0.57 & 0.01 & 0.01 \\ 0.99 & 0.04 & 0.99 & 0.67 \\ 0.99 & 0.99 & 0.23 & 0.33 \end{pmatrix}$ after 86,000 generations. Then the trajectory changes rapidly for 2000 generations finally converging toward the agent

$\begin{pmatrix} 0.99 & 0.01 & 0.87 & 0.99 \\ 0.99 & 0.54 & 0.99 & 0.09 \\ 0.99 & 0.01 & 0.99 & 0.99 \\ 0.99 & 0.17 & 0.39 & 0.5 \end{pmatrix}$ in the FBF clan.

Major changes take far fewer generations to occur along the Clem trajectory than for FBF and 2TFT. This decreased stability follows from the fact that the critical state CdCc for Clem is on the primary recovery route whereas the critical states for these other strategies are on the secondary routes and so grow at the rate of ε^2.

Several of the cooperative trajectories appear to stabilize in the cooperative basin. The gradient at TL ($\frac{2}{3} - \varepsilon$, $\frac{2}{3} - \varepsilon$) has positive components in the XxXc and XxDd directions (first, third and fourth columns) and negative components in the XxCd direction (second column). The components in the XdDd directions increase very gradually from $\frac{2}{3} - \varepsilon$ and asymptotically approach the agent TL (0.6578, 0.6578), up to computer accuracy. The effective gradient at this agent vanishes and so is an equilibrium point on the trajectory.

The trajectory from GP with all * = .5 (see 10.3) asymptotically approaches a stable agent that we call **Grim But Relenting**

$$\text{GBR} = \begin{pmatrix} 0.99 & 0.01 & 0.01 & 0.01 \\ 0.66 & 0.01 & 0.99 & 0.99 \\ 0.99 & 0.01 & 0.33 & 0.01 \\ 0.99 & 0.01 & 0.99 & 0.65 \end{pmatrix} \qquad (14.2)$$

with relatively low self payoff of 0.94147 after 100,000 generations

To confirm these simulations, the gradient and the effective gradient was estimated for agents near these apparently evolutionarily stable strategies.
Similar results hold for GBR. For this agent, cooperation is soon restored after two errors in close succession. The states, CdCC and DcDc for which probability of cooperation is of the order of $\frac{1}{3}$ and $\frac{2}{3}$ respectively, can only be entered after three errors in close succession and so have ε^3 effect on the payoff and on the evolution. In distinction with their ancestors, a group of GBR agents cannot be invaded by AllD and will supplant a dimorphic population with AllD if they are initially in the majority but will succumb otherwise.

15. Robustness of the model

To confirm the robustness of the method, one must check that no significant changes occur in the trajectory as the noise level, the step size, initial agents, and the payoff entries are slightly altered.

To show that the step size δ does not greatly affect the dynamical system, trajectories were re-computed for 100 random starting points in the hypercube with step size 0.005 (half of the original) but number of generations (steps) doubled from 4000 to 8000. Only one differed by more than 0.1 and only three others by more than 0.01. Results were similar with step size 0.04.

The effect of the noise level ε is just a bit more significant. Trajectories were run from 100 initial agents with noise level 0, 0.005, 0.01 and 0.02 for 4000 generations. In the model with 0 noise, the cooperative trajectories most often reach an evolutionarily stable state of pure cooperation and self-payoff of 1 without the need to develop the secondary recovery protocols that distinguish various clans. With various levels of positive noise, the self-payoffs to the descendent agents of most trajectories are in close agreement. The payoff to FBF agents tends to average $1-4\varepsilon$ while those in the GP and 2TFT clans have payoffs closer to $1-6\varepsilon$.

Of those trajectories that converged to cooperative agents with the three different noise levels, very few changed clans. By increasing the noise level from 0.01 to 0.02, only 6 trajectories went to different basins; two defectors and three cooperators went to alternators and one alternator went to a defector. Decreasing the noise level had a marginally greater effect; three defectors and one cooperator became an alternator, three alternators and three cooperators became defectors, and two defectors and two alternators became cooperators. Reducing the noise from 0.005 to 0 had the effect of pushing one alternator and four defectors to cooperation, three alternators to defection and one defection to alternation. Less noise means fewer alternators, somewhat more defectors and even more cooperators.

To show that trajectories are not sensitive to initial conditions, a 4×4 *perturbation* matrix P with entries normally distributed with mean 0 and standard deviation 0.1 was generated. Trajectories were calculated for 200 random initial matrices for 4000 generations and then P was added to the initial matrices and the trajectories were calculated again. Only 10 (5%) of the end agents had changed significantly and most of those had changed from one of the 3 or 4 cycles to another one or to C or D basin. Only one final agent had shifted from a cooperator to a defector.

The migrant evolution method used by Nowak & Sigmund (1993) is quite different in spirit from the adaptive dynamical system model, yet the two models give qualitatively similar results. In the migrant evolution model, if a few percent of TFT or FP agents are introduced by chance into a defecting population, they will rapidly displace the defectors. These agents are in turn replaced by less exploiting varieties. Haurert & Schuster (1998) found that after $5\cdot10^7$ such generations, the

populations in their models were dominated by FBF types with high values of u, v and values of x and y approaching $\frac{2}{3}$.

16. Differences with memory size

Assume that a species of *k*-ply agents acquires more memory and becomes a species of *k*+1-ply agents. The trajectories of these agents considered as *k*-ply and as *k*+1-ply often diverge with one becoming cooperators and the other defectors. Occasionally larger memory can make the trajectory less stable.

For example, the cooperative 1-ply agent (0.99, 0.5) asymptotically converges to the agent (0.99, 0.656). If its memory doubles, the corresponding 2-ply agent will evolve into the very generous agent (0.99, 0.662, 0.99, 0.620) after 6000 generations, quite similar to its 1-ply cousin. But a few generations later, cheating offspring with slightly lower Pr(Cc) and Pr(Dc) will invade and the trajectory collapses to AllD in about 400 generations.

Some agents will remain stable with increased memory. The 2-ply agent (0.99, 0.01, 0.5, 0.5) will asymptotically approach the stable FBF ≈ (0.99, 0.10, 0.99, 0.65). When extended into a 3-ply agent, quite rapidly, Pr(dDc) rises to 0.99; successful strategies must accept contrition by others to avoid the mutual defection rut. At the same time, Pr(cDc) increases slowly but linearly, reaching 0.99 after 160,000 generations; less exploiting offspring have higher self-payoff. The entries Pr(xDd) have less effect and the trajectory asymptotically approaches
$\begin{pmatrix} 0.99 & 0.01 & 0.99 & 0.741 \\ 0.99 & 0.01 & 0.99 & 0.655 \end{pmatrix}$.

A further increase in memory for this agent has the opposite effect. The trajectory of the 4-ply extension of (0.99, 0.01, 0.5, 0.5) changes in a much more dramatic way. As for the 3-ply extension, Pr(XdDc) rises rapidly to 0.99. The trajectory initially evolves toward a FP type with Pr(CcDd) increasing rapidly and the other intermediate entries growing at a far slower rate. But unlike the 3-ply extension, after about 20,000 generations, Pr(CcDd) falls again as other routes to restoring cooperation become established. Eventually, after about 53,000 generations, the guilt index Pr(CcDc) increases over the critical value of $\frac{2}{3}$ allowing benevolent agents with higher generosity Pr(CcCd) to move in. This index rapidly increases and passes beyond the $\frac{2}{3} - \varepsilon$ threshold at 60,000 generations. Immediately thereafter, cheating offspring invade and in 1000 more generations, the trajectory has all but collapsed to AllD where it remains. See Figure 4.

More mathematically, various components of the gradient vanish when critical values are reached as detailed with Clem above. The self-payoff during this evolutionary track has several twists and turns. It rises and falls gradually with Pr(CcDd) and rises rapidly as Pr(CcCd) begins to increase. But as soon as Pr(CcCc) starts to decrease, the self-payoff plummets as the trajectory stabilizes in the D basin. As noted in section 14, the trajectories from the 3-ply and the 4-ply FP agents are sometimes quite distinct, with the former converging to a TL type and the latter eventually turns to the defecting basin.

These results show that adding more memory may not necessarily be advantageous. Several stable *k*-ply trajectories eventually collapse when doubled to *k*+1 agents. In some situations, one may be too clever for one's own good. Similar conclusions about variations of the stochastic Pavlov agents in the IPD were reached in Kraines & Kraines (1995).

17 Finite versus infinite match

Since infinite games are impossible in nature, to model real life applications, it may seem necessary to find the expected payoff in an n round APD with given initial states u_0 and v_0 for `A` and `B` respectively. This payoff is the inner product of the payoff [R S T P]" with $u_0(T_{AB})^n$ and [R T S P]" with $v_0(T_{BA})^n$ respectively. There is a computational disadvantage in this process. For small n, this payoff can be quite sensitive to the initial states of each agent and for large values of n, it involves considerably CPU time to make these evaluations. If one assumes that agents interact for an infinite number of rounds in the APD, standard Markov methods can be used to calculate the payoff quite rapidly

Fortunately, the use of Markov methods is an acceptable approximation. In a simulation of 1000 length n matches between randomly chosen 4-ply agents, the set of payoffs resembles a normal distribution between –1 and 1. These matches were replayed for infinite length games using the Markov methods discussed above. The absolute difference between the expected payoff in an n = 8 round game and the (infinite round) average payoff is about 0.01. The difference falls to 0.002 for n = 20. This difference in payoff approaches 0 at nearly an exponential rate in the number of rounds. Even in a 5 round APD, the majority of the pairs receive a payoff within 0.01 of what they would receive in an infinite game, although for about 5% of these pairs, the differences are greater than 0.1.

For agents with shorter memory, the difference between the average payoff in a short match and an infinite match is even smaller. For 3-ply agents, the average absolute difference drops below 0.01 after 7 rounds and to 0.001 by 15 rounds. For 2-ply agents, the average absolute difference falls to 0.01 after 6 rounds and to 0.001 by 12 rounds while for 1-ply agents, the average absolute difference falls to 0.01 after just 3 rounds and to 0.001 by the fifth round.

18. Conclusion

In human society, interactions tend to be governed by finely tuned learning and tradition (memes). Elaborate rituals and conventions exist in all societies for restoring harmony in the village after crimes or sins have been committed. These patterns vary widely among different cultures. The rituals of visitors will occasionally conflict with that of their hosts and interactions may cause trouble for all parties involved (lose-lose). The simple APD model in this paper shows that such cultural differentiation may arise by a method analogous to Darwinian natural selection.

Although several relatively stable protocols for cooperation emerge in the model, among the various clans 4-ply stochastic agents, the ideal one is arguably the Tough Love strategy. Unlike `TFT`, `TL` will avoid vicious cycles of defections with its clone. Unlike `Pavlov`, `TL` will not exploit altruistic agents and she will quickly resume cooperation with others after any series of miscommunications or errors. Unlike the fallible `FBF`, a sizable minority of `TL` agents will replace a majority of `AllD` agents. And unlike `FP`, `TL` with appropriate parameters is an evolutionarily stable strategy.

For genetic and/or cultural reasons, we feel pleasure after doing a good deed. We also resent those who take advantage of our good will and retaliate or set up a judicial system to retaliate for us. Yet we do tend to forgive those who repent/apologize. The emotion of guilt deters us from exploitation and shortens the return to cooperation. These feelings correspond to what would drive the Tough Love agent. Our emotional responses may well have evolved in part to get through prisoner dilemma type interactions.

Bibliography

Axelrod, R. (1984) *The evolution of cooperation*, New York: Basic Books

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

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

Brams, SJ (1994) *Theory of Moves* Cambridge University Press

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

Frean, M (1994) the prisoner's dilemma without synchony. *Proc R. Soc London B* **257** 75-79

Frean, M (1996) The evolution of degrees of cooperation, *J. Theoretical Biology.* **182** 549-559

Hauert, Ch. and Schuster H.G. , (1998) Extending the Iterated Prisoner's Dilemma without Synchrony. *J. Theoretical Biology.* **192** 135-166

Hofbauer, J. and Sigmund, K (1990) Adaptive Dynamics and Evolutionary Stability*, Appl Math Lett.* **3** 75-79

Hofbauer, J. and Sigmund, K. (1998) *Evolutionary Games and Population Dynamics.* Cambridge UK Cambridge University Press

Kemeny, J & Snell, J (1976) *Finite Markov Chains.* New York Springer Verlag.

Kraines, D. and Kraines, V (1993). Learning to Cooperate with Pavlov. *Theory and Decision*, **35**: 107:135

Kraines, D. and Kraines, V. (1995), *Evolution of learning among Pavlov strategies*, *J. Conflict Resolution* **39**: 439-466

Kraines, D. and Kraines, V (2000) Natural selection of memory-one strategies in the Prisoner's Dilemma. *J. Theoretical Biology.* **203**: 335-355

Maynard Smith, J. (1982) *Evolution and the theory of Games*. Cambridge: Cambridge Univ Press

Maynard Smith, J. and Price,G.R. (1973) The logic of animal conflict. *Nature*. **246**:15-18

Neill, D.B. (2001) Optimality under noise. *J of Theoretical Biology* **211**: 159-180

Nowak, M.A. and Sigmund, K. (1990). The evolution of stochastic strategies in the prisoner's dilemma. *Acta Appl. Math*., **20**:247-265

Nowak, M.A. and 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

Nowak, M.A. and Sigmund, K. (1994) The alternating Prisoner's Dilemma *J. Theoretical Biology.* **168** 219-226

Rapoport, A. and Chammah, A.M., (1965) *Prisoner's Dilemma*. Ann Arbor: University of Michigan Press

Williams,G. C. (1966) *Adaptation and Natural Selection*. Princeton: Princeton University Press

Wedekind, C., & M. Milinski (1996): Human cooperation in the simultaneous and the alternating Prisoner's Dilemma: Pavlov versus Generous Tit-for-Tat *Proc. Natl. Acad. Sci.* USA **93**: 2686-2689.

David Kraines
Department of Mathematics
Duke University
Box 90320
Durham NC 27701

Vivian Kraines
Department of Mathematics and Computer Science
Meredith College
Raleigh NC 27607

Appendix A

Agents considered in the paper:

Firm But Fair

$$\mathbf{FBF} = \begin{pmatrix} 0.99 & 0.01 & 0.5 & 0.99 \\ 0.5 & 0.5 & 0.99 & 0.5 \\ 0.99 & 0.5 & 0.5 & 0.5 \\ 0.99 & 0.5 & 0.5 & 0.5 \end{pmatrix}$$

Firm Pavlov

$$\mathbf{FP} = \begin{pmatrix} 0.99 & 0.01 & 0.01 & 0.99 \\ 0.99 & 0.01 & 0.99 & 0.66 \\ 0.99 & 0.01 & 0.01 & 0.99 \\ 0.99 & 0.01 & 0.99 & 0.66 \end{pmatrix}$$

Tough Love

$$\mathbf{TL} = \begin{pmatrix} 0.99 & 0.01 & 0.99 & 0.99 \\ 0.99 & 0.01 & 0.99 & 0.66 \\ 0.99 & 0.01 & 0.99 & 0.99 \\ 0.99 & 0.01 & 0.99 & 0.66 \end{pmatrix}$$

Grim Pavlov

$$\mathbf{GP} = \begin{pmatrix} 0.99 & 0.01 & 0.5 & 0.01 \\ 0.5 & 0.5 & 0.5 & 0.99 \\ 0.99 & 0.5 & 0.5 & 0.5 \\ 0.99 & 0.5 & 0.99 & 0.5 \end{pmatrix}$$

Grim But Relenting

$$\mathbf{GBR} = \begin{pmatrix} 0.99 & 0.01 & 0.01 & 0.01 \\ 0.66 & 0.01 & 0.99 & 0.99 \\ 0.99 & 0.01 & 0.33 & 0.01 \\ 0.99 & 0.01 & 0.99 & 0.66 \end{pmatrix}$$

Two Tits for Tat

$$\mathbf{2TFT} = \begin{pmatrix} 0.99 & 0.01 & 0.5 & 0.99 \\ 0.99 & 0.5 & 0.01 & 0.01 \\ 0.99 & 0.5 & 0.99 & 0.5 \\ 0.5 & 0.99 & 0.5 & 0.5 \end{pmatrix}$$

Clemency

$$\mathbf{Clem} = \begin{pmatrix} 0.99 & 0.99 & 0.99 & 0.5 \\ 0.66 & 0.5 & 0.5 & 0.5 \\ 0.99 & 0.99 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$$

Appendix B

***Matlab* calculation of the Markov transition matrix, equilibrium vector and payoff**

Given *k*-ply agents A and B, the $2^k \times 2^k$ Markov stochastic matrices T_A and T_B are easily constructed in Matlab. The m^{th} row of T_A is [0, 0, .. 0, u, (1-u), 0, .. 0] where u = u(σ) is the probability for cooperating after the sequence σ corresponding to the binary representation of the number m-1. This entry u occurs in column $2m-1$ if $m \le 2^{k-1}$ and in column $2m-2^k-1$ otherwise.

Example

For $k = 3$ and $A = \begin{bmatrix} 0.1 & 0 & 0.6 & 1 \\ 0.3 & 0.8 & 0 & 1 \end{bmatrix}$, then T_A will be an 8×8 matrix, with rows indexed by the 8 sequences of length 3 from cCc to dDd. and the columns indexed by CcC to DdD.

$$T_A = \begin{bmatrix} 0.1 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0.3 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The (3,5) entry of 0.6 corresponds to the 60% probability that from (third) state cDc, agent A will choose C, resulting in the (fifth) state DcC presented to B.

The products $T_{AB} = T_A * T_B$ and $T_{BA} = T_B * T_A$ are stochastic matrices that represent a full round in the APD. The length 2^k normalized left eigenvectors eq_A for T_{AB} and eq_B for T_{BA} are calculated, using *Matlab,* by dividing the generator of the null space of $(I - T_{AB})^T$ by the sum of its entries.

The expected payoff in Definition 5.1 is computed as follows. If the 3-ply equilibrium vector in a match between 3-ply agents A and B is eq_A = (0.1, 0.3, 0, 0.2, 0, 0.1, 0.2, 0.1), the second entry 0.3 corresponds to the frequency of the sequence cCd while the 6th entry 0.1 represents dCd. In other words, B has cheated A, 30% of the time after B originally cooperated and 10% of the time after B originally defected. Similarly, both have cooperated (xCc) with frequency 10%, A cheats B (xDc) with frequency 20% and both defect (xDd) with frequency 20%+10%. After the first ply of a round, B receives $0.1 \cdot R + (0.3+0.1) \cdot T + 0.2 \cdot S + 0.3 \cdot P = 0.2$ while A gets a negative value $0.1 \cdot R + (0.3+0.1) \cdot S + 0.2 \cdot T + 0.3 \cdot P = -0.6$.

Proof of Theorem 5.3

It is useful to establish the following.

Lemma: a) <[1, 1, 1, 1]", eq_A> = <[1, 1, 1, 1]" , eq_B> = 1

b) <[1, 0, 1, 0]",eq_A > = <[1, 1, 0, 0]", eq_B > and <[1, 0, 1, 0]",eq_B > = <[1, 1, 0, 0]", eq_A>

c) <[0 0 1 1]", eq_A > = <[0 1 0 1]",eq_B>

d) <[0 1 -1 0]",eq_A > = <[0 -1 1 0]" eq_B>

e) <[1 0 0 -1]", eq_A > = <[1 0 0 -1]" ,eq_B>

Statement a) says that if all payoffs are the same, then the outcome is always equal. The mathematical verification comes from noting that <[1 1 1 1]", eqA> is the sum of the entries of the equilibrium vector, and this has been normalized to be 1.

In b), the LHS represents the proportion of *k*-ply states presented to agent A in which the sequence ends in a c by agent B, i.e., all sequences of the form (...XxXc). The RHS represents the proportion of *k*-ply states presented to agent B in which the last subsequence of length 4 has a c in position 3 (...xXcX). In other words, B chose to cooperate on the next to last step. In the equilibrium (long run), these sets are in one to one correspondence. Similarly, in c), the LHS represents XxDx for A and the RHS represents xXxD for B.

To get equation d) subtract equation b) and c) from a). To get equation e), subtract equation c) from b).

Using the equations in the Lemma and vector algebra,

<[R S T P]", eq_A > – <[R T S P]", eq_B > =

<([R+P, S+T, 0, 0]"– P[1 0 0 -1]" – T[0 1 -1 0]) , eq_A >
– <([R+P, 0 S+T, 0]" – P[1 0 0 -1]" – T[0 -1 1 0]) , eq_B >

= <[R+P–S–T, 0 0 0]", eq_A > +(S+T) < [1 1 0 0] ,eq_A > – < [R+P–S–T, 0 0 0]", eq_B > +(S+T)< [1 0 1 0] , eq_B >

$$= (R+P-S-T)\langle [1\ 0\ 0\ 0]",(eq_A - eq_B)\rangle = (R+P-S-T)\sum_{i=0}^{2^{k-2}} (eq_A(4i+1) - (eq_B(4i+1))$$