J. theor. Biol. (2001) **211,** 159–180
doi:10.1006/jtbi.2001.2337, available online at http://www.idealibrary.com on IDEAL®

Optimality Under Noise: Higher Memory Strategies for the Alternating Prisoner's Dilemma*

DANIEL B. NEILL†‡

Department of Computer Science, Duke University, Durham, North Carolina 27708, *U.S.A.*

(*Received on* 12 *June* 2000, *Accepted in revised form on* 19 *April* 2001)

The Alternating Prisoner's Dilemma is a variant of the iterated Prisoner's Dilemma in which the players alternate in the roles of actor and recipient. We searched for strategies which are "optimal" in the Alternating Prisoner's Dilemma with noise (a non-zero probability that a player's decision will be transmitted incorrectly). 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). It must also have high evolutionary "dominance", a general measure of evolutionary performance which considers both resistance to invasion and the ability to invade other strategies. A strategy which meets these optimality criteria can evolve cooperation by invading a population of defectors and establishing a stable cooperative society.

Most of the strategies commonly discussed in the Alternating Prisoner's Dilemma literature are low-memory strategies such as Tit For Tat, Pavlov, and Firm But Fair, but none of these strategies can simultaneously meet all of the optimality criteria. However, we discovered a class of higher memory "Firm Pavlov" strategies, which 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.

Introduction

The Prisoner's Dilemma is a widely used mathematical model of interactions between individuals with partially conflicting goals. In this paradoxical 2×2 non-zero sum game, each player must choose whether to *cooperate* for mutual benefit or to *defect* for individual benefit (harming the other player). The outcomes of the Prisoner's Dilemma game are given in the following payoff table (Table 1).

If both players cooperate, each receives the payoff R as a *reward* for mutual cooperation. If both players defect, each receives the payoff P as a *punishment* for defecting. If one player defects while the other cooperates, the defector receives payoff T (as a *temptation* to defect), while the cooperator receives the *sucker* payoff S. The Prisoner's Dilemma is defined by $T > R > P > S$.

---

* This work is supported by NSF VIGRE grant number DMS-9983320.

† Author's note: This is a revised and abridged version of my 2000 undergraduate honors thesis of the same title. More complete data/results, as well as a more detailed exploration of some concepts, can be found in the original paper at *http://www.math.duke.edu/vigre/pruv/studentwork/neill.pdf.*

‡ Author to whom correspondence should be addressed. 1903, Cape Bend Avenue, Tampa, FL 33613, U.S.A.

0022–5193/01/140159 + 22 $35.00/0

Table 1
Payoffs to P1/P2

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

Additionally, $2R > T + S$, so cooperation achieves more points than alternating T and S payoffs.

What is the optimal strategy for the Prisoner's Dilemma? It is clear that mutual cooperation is preferable to mutual defection, but in the one-shot Prisoner's Dilemma game, no matter what the other player chooses, a player can achieve a higher payoff by defection. Hence the optimal strategy is to defect, and mutual defection is the only rational result of a one-shot Prisoner's Dilemma. An *Iterated Prisoner's Dilemma* (IPD) occurs when a Prisoner's Dilemma interaction is repeated by the same players over a number of rounds. Based on past results, reciprocal altruism can develop, enabling mutual cooperation to become a rational option. A *strategy* for the IPD is an algorithm for deciding whether or not to cooperate on a given round, based on the results of previous rounds. It is clear that some strategies are better than others. If a player cooperates too often, the other player can take advantage of him by defecting; if he defects too often, the other player is likely to retaliate, resulting in low scores for both players. Since a strategy's performance in the IPD is highly dependent on the other player's strategy as well as the specific parameters of the game, it is much less obvious as to what an "optimal" IPD strategy would be, or even how "optimal" should be defined. In 1980, Robert Axelrod conducted a computer "tournament" in which 14 strategies designed by leading game theorists were matched against each other. The winning strategy was *Tit for Tat* (*TFT*), submitted by Anatol Rapoport. This strategy cooperates initially, then echoes what the other player chose in the last round (Axelrod & Hamilton, 1981). Tit for Tat is very successful in many variants of the Prisoner's Dilemma game, and its success has sparked theories of the evolution of cooperation based upon reciprocity (Axelrod, 1984). These models have been applied to fields ranging from economics to biology (e.g. Maynard Smith, 1982; Axelrod, 1984; Milinski, 1987).

Tit for Tat, however, is clearly not successful in an iterated Prisoner's Dilemma with "noise", a non-zero probability that a player's decision will be transmitted incorrectly. In an IPD interaction between two *TFT* strategies, a single accidental defection can lead to an endless sequence of mutual recriminations (Nowak & Sigmund, 1990). Various other strategies have been examined for the the noisy iterated Prisoner's Dilemma, including the *Pavlov* (Win-Stay, Lose-Switch) strategy, which cooperates only after mutual cooperation or mutual defection. Pavlov is an "error correcting" strategy: against its clone, it can recover from an accidental defection, resuming mutual cooperation in a relatively small number of rounds (Kraines & Kraines, 1989, 1993).

The *Alternating Prisoner's Dilemma* (APD) is a variant of the iterated Prisoner's Dilemma in which the players alternate in the roles of actor and recipient rather than acting simultaneously (Nowak & Sigmund, 1994; Frean, 1994). In this game, players alternate turns, and on each player's turn he must choose whether to cooperate or defect. This game models many situations in which participants must take turns in helping each other, such as gift-giving, or bipartisan cooperation in politics. Alternating reciprocal altruism is also commonly observed in animal behavior: for example, South American vampire bats who have found a good meal will help hungry bats by donating some of their surplus food (Wilkinson, 1984). In this sort of interaction, only one bat (the one with food) makes a decision to cooperate; this cooperation is likely to be reciprocated by other bats during future feedings. Similarly, an Alternating Prisoner's Dilemma is observed in the guarding behavior of the dwarf mongoose (Rasa, 1989). When a group leaves the termite mound where it has roosted for the night, one member (a subadult male) will stay behind and maintain a watch for predators, increasing its own risk in order to protect the group. At the next mound, another individual will run ahead and establish the next watch, while the rearguard rejoins the pack. In this example, the guarding animal must make a choice between displaying vigilant guard behavior (cooperation) or focusing on its own safety (defection). Examples of alternating reciprocal altruism can also be seen in the fighting of young male baboons (Trivers, 1985). In these examples, it makes no sense for the players to cooperate simultaneously (Nowak & Sigmund, 1994), so the standard Iterated Prisoner's Dilemma does not model the situation as well as the Alternating Prisoner's Dilemma.

To simplify our calculations, we focus on a "strictly alternating" Prisoner's Dilemma between two players, A and B (i.e. A gets one turn, then B gets one turn, every round). Player A's choices are denoted in upper case (C or D), and player B's choices are denoted in lower case (c or d). To keep score, we treat each turn after the first one as if it were a complete Prisoner's Dilemma, and thus each player receives two payoffs every round. For example, consider an APD game which starts at CdDc. This means that, on the first turn, Player A chooses C. Since it is the first turn, neither player receives a payoff. On the second turn, Player B chooses d. Since Player A's last choice was C, A receives a sucker payoff S, and B receives a temptation payoff T. On the third turn, Player A chooses D. Since Player B's last choice was d, both players receive a punishment payoff P. On the fourth turn, Player B chooses c. Since player A's last choice was D, A receives a temptation payoff T, and B receives a sucker payoff S. Thus after the sequence CdDc, each player has received a total payoff of $T + P + S$. This process continues indefinitely, with players receiving payoffs based on the last two turns.

Though some have argued that the IPD and the APD are mathematically equivalent (Axelrod, 1984), the two games can result in dramatically different interactions between strategies. For example, in an IPD game between two Tit for Tat players, a single accidental defection leads to a sequence of alternating temptation and sucker payoffs, while in an APD game, an error leads to a sequence of punishment payoffs. As a result, strategies may be successful in IPD but not APD interactions, or vice versa. For example, two Pavlov players quickly restore mutual cooperation after an error in the IPD (CD DD CC CC ...), but in the APD an error leads to a 6-turn cycle of temptation, punishment, and sucker payoffs (CdDcDdCdDcDd ...). Thus the Pavlov strategy is "self-cooperating" in the IPD but not the APD. The *Firm but Fair* (*FBF*) strategy, which cooperates except after receiving a sucker payoff, is self-coooperating in the APD but not the IPD (Frean, 1994). Two *FBF* players can restore mutual cooperation after an error in the APD (CdDcCc ...), but in the IPD an error leads to alternating temptation and sucker payoffs (CD DC CD DC ...). From these examples, it is clear that the IPD and APD are distinct problems, and successful strategies for one game may perform poorly in the other.

n-ply strategies for the Alternating Prisoner's Dilemma

We now define some strategies which are commonly discussed in the Alternating Prisoner's Dilemma literature, and introduce the general notion of an "n-ply" strategy. The *n-move history* of a game is a string of C's and D's $H = h_n h_{n-1} \dots h_2 h_1$, where $h_k = C$ if a player cooperated k turns ago, and $h_k = D$ if a player defected k turns ago. Odd indices correspond to the opponent's moves, and even indices correspond to the player's moves. The opponent's moves are generally written in lower case, and the player's moves in upper case. For example, a 3-move history of cCd would mean that the opponent's most recent move was defection, the player's most recent move was cooperation, and the opponent's previous move was cooperation.

We can enumerate the 2^n possible history strings H in a lexicographic order, with $H_1 = C \dots Cc$, $H_2 = C \dots Cd, \dots$, $H_{2^n-1} = D \dots Dc$, $H_{2^n} = D \cdots Dd$. An *n-ply strategy* for the APD chooses whether to cooperate or defect based on the n-move history H. The strategy is defined by 2^n numbers $(a_1 \dots a_{2^n})$, where $0 \leqslant a_i \leqslant 1$ for all i, and each a_i corresponds to the probability of cooperation if $H = H_i$.

For example, the 2-ply strategy (1 0 1 0.2) will cooperate if $H = H_1$ or $H = H_3$, that is, if the 2-move history is Cc or Dc. Similarly, it will defect if the 2-move history is Cd, and cooperate with probability 0.2 if Dd.

The commonly discussed 1-ply strategies include *ALLC* (1 1), *ALLD* (0 0), *TFT* (1 0), and Random ($\frac{1}{2}$ $\frac{1}{2}$). The commonly discussed 2-ply strategies include Pavlov (1 0 0 1), Firm But Fair (1 0 1 1), and the "generous strategies" $g1$ ($1\ \frac{1}{3}\ 1\ \frac{1}{2}$), $g2$ ($1\ \frac{1}{2}\ 1\ \frac{1}{2}$), and $g3$ ($1\ \frac{2}{3}\ 1\ \frac{1}{2}$). We can also form a 2-ply strategy by a trivial extension of a 1-ply strategy: the 2-ply strategy ($a\ b\ a\ b$) behaves equivalently to the 1-ply strategy ($a\ b$). In general, the n-ply strategy S_n is equivalent to the $n+1$-ply strategy formed by concatenating S_n with itself.

To fully specify an n-ply strategy, we should also specify its behavior at the start of the game (turns $1..n$). We generally assume that $h_k = C$ for $k \geqslant n$, that is, "moves" in the history list before the start of the game are assumed to be cooperation. For example, on the first turn of a game, a 3-ply strategy will assume that the history is $H = cCc$. Alternatively, we can average over all possible initial history sequences.

What Does it Mean to be Optimal?

The question of what makes a strategy "optimal" in the Iterated Prisoner's Dilemma or Alternating Prisoner's Dilemma is very difficult to answer. As many researchers have noted, the performance of a strategy is highly dependent on which other strategies it interacts with; this has led to several conflicting definitions of the term "optimality", with resulting differences in which strategies (if any) are considered optimal. According to one line of argument, optimality is defined relative to a given set of opponents: the optimal strategy is the one which achieves the highest score (with respect to some measure) against that set of opponents. One typical measure of performance is the average score in a round-robin tournament interaction. As we shall prove, no fixed strategy performs best against every given set of opponents in a round-robin tournament interaction, and this has caused many to argue that no optimal strategy exists.

However, there are three main problems with this argument. First, it is not clear that we should limit our search to "fixed" strategies: given a sufficiently long game length, it may be possible for a strategy to achieve success against an opponent by making a large number of exploratory moves, constructing and testing hypotheses which describe the opponent's behavior, and thus adapting over time. Such strategies are beyond the scope of this paper; we consider only those strategies whose behavior is governed by simple, fixed rules. Second, it is not clear if a round-robin interaction is the best measure of "success" on which to base our definition of optimality. Numerous other measures have been proposed, including various measures of success in evolutionary interactions. We consider many of these "evolutionary optimality criteria" later in the paper, and propose our own measure of evolutionary success. This measure, which we call "evolutionary dominance", is a combination of evolutionary stability (ability to resist invasion) and evolutionary potency (ability to invade other strategies). Nevertheless, all of these measures depend on the choice of opponents; it is clear that if we require an "optimal" strategy to be optimal with respect to every given set of opponents, most reasonable measures will conclude that no optimal strategy exists.

The third, and most profound, objection to this argument questions the idea that "optimal" strategies must be most successful with respect to every given set of opponents. In the real world, organisms must react to an uncertain environment; they are likely to have limited and incomplete knowledge, and must act in the best way they can, given this knowledge. Thus we ask the following question: what is the best fixed strategy for an organism to choose, assuming no prior knowledge of the set of opponents? An optimal choice of strategy is one which maximizes its expected success under uncertainty, given the information and options available to it. In other words, we consider a strategy optimal if no other known strategy is expected to achieve a higher performance in this uncertain environment. Thus we use an inductive approach to optimality, one which bears a distinct resemblance to the modern scientific method: a strategy is considered optimal until we discover a strategy which is demonstrably "better".

In order to apply this method, we must propose and examine "optimality criteria", standards for evaluating the general performance of a strategy. If we define a reasonable set of criteria, strategies which meet these criteria are clearly preferable to those which do not: by defining increasingly strict criteria, we can eliminate non-optimal strategies and gradually narrow down the possibilities for an optimal strategy. If we limit the domain of strategies under consideration (for example, all n-ply strategies), we may even be able to find a single strategy which is clearly optimal in that domain.

The question remains: how do we define these optimality criteria? It is clear from our definition that an optimal strategy must perform with a high degree of relative success in interactions with many other sets of strategies. Thus we search for criteria which are strongly correlated with a high average performance. In other words, it is not sufficient to propose criteria and consider a strategy "optimal" if it meets these standards. It is also necessary to show that strategies which meet the criteria are highly successful against a wide variety of strategies (with respect to such established criteria as round-robin tournament performance), for otherwise our chosen criteria would be useless.

Before defining the optimality criteria, we first attempt to gain a general idea of what it means for a strategy to be "successful" in an Alternating Prisoner's Dilemma interaction. We consider a single, infinitely long APD game between two strategies X and Y. The goal of each strategy is to maximize its average payoff per turn,

$$w = \lim_{N \to \infty} \frac{1}{N-1} \sum_{i=2}^{N} p(i),$$

where $p(i)$ is the payoff received by that strategy after turn i. Though the infinite game length is an idealization of real-life interactions, it serves as a good approximation of situations in which the probability of further encounters is high (Nowak & Sigmund, 1994). Kraines and Kraines show that average payoffs in 4-round and 8-round games are closely approximated by the infinite game (2000). Thus even short games can be modeled closely by this assumption, as long as the players believe that the game will continue.

We denote the expected value of the payoff w to strategy X in an infinitely long game against strategy Y as $w(X|Y)$. In particular, we define $w(X|X)$, the "self-payoff" of strategy X, to be the expected value of its payoff w in an infinitely long game against its clone. In certain cases (i.e. interactions between deterministic strategies with no noise), payoffs may vary depending on which strategy goes first. In these cases, we average the two possibilities. We assume that an APD strategy X will interact with many different strategies Y over time, and thus the strategy's goal is to maximize its average payoff against all strategies it interacts with. Assuming a population of opponents Y_i with varying frequencies $f(Y_i)$, with $\sum_i f(Y_i) = 1$:

$$w(X) = \sum_i f(Y_i) w(X|Y_i).$$

The relative success of two strategies X and Y can be evaluated by comparing $w(X)$ and $w(Y)$: strategies with higher w are considered to be more successful.

What characteristics must a strategy have for it to achieve relative success against a wide variety of other strategies? Two things influence a strategy's payoff on any given turn: a strategy scores higher if it defects, and if its opponent cooperates. Thus a strategy X should choose to cooperate with a strategy Y only if cooperating will make Y significantly more likely to cooperate on future turns. In particular, a strategy should be able to exploit unconditional cooperators, as well as prevent other strategies from exploiting it. It should also cooperate with nice, but unexploitable, strategies such as Tit for Tat. Thus one possibility for the optimality criteria would be to consider a strategy's scores against *ALLC*, *ALLD*, and *TFT*. However, since *TFT* performs poorly in games with noise, and choosing any specific variation of *TFT* adds a subjective bias to our criteria, we choose instead to consider a strategy's score against its clone. Strategies which can achieve cooperation with their clones are likely to cooperate with a wide range of other strategies, while strategies which do not cooperate with their clones are unlikely to achieve cooperation with most other strategies.

Combining these desired characteristics, we find that an "optimal" strategy must possess three essential properties. It must be "self-cooperating" (able to achieve mutual cooperation with its clone), "unexploitable" (able to resist exploitation by *ALLD* and other strategies), and "exploiting" (able to exploit unconditional cooperators). We will show that strategies which meet these simple criteria are able to achieve remarkable success with respect to a variety of measures, including a round-robin tournament simulation.

These properties are also essential for success in the "evolutionary" APD, in which a strategy's payoff w is used as a measure of its reproductive fitness (Maynard Smith & Price, 1973). Strategies which receive higher payoffs are able to produce more offspring, while those with poor performance quickly die off, and thus the population of strategies evolves over time. In the evolutionary APD, an "optimal" strategy should be able to invade and resist invasion by other strategies. According to Maynard Smith's definition (1982), strategy X *invades* strategy Y if $w(X|Y) > w(Y|Y)$, or if $w(X|Y) = w(Y|Y)$ and $w(X|X) > w(Y|X)$. We will discuss the evolutionary APD, defining a more precise measure of evolutionary "dominance", but for now we simply consider a strategy's evolutionary interactions with unconditional cooperators (the *ALLC* strategy) and unconditional defectors (the *ALLD* strategy). An optimal strategy should be able to evolve cooperation by invading a population of defectors and establishing a cooperative "society". Once established, this society should be "stable" enough to resist invasion by defectors, as well as resisting *genetic drift*, the infiltration of the population by unconditional cooperators. Strategies which are unexploitable and self-cooperating can establish a stable cooperative society of this sort, and strategies which are exploiting can resist genetic drift. We will also show that the strategies meeting these criteria are extremely successful in evolutionary interactions with a wide variety of other strategies, with success defined by the measure of "evolutionary dominance".

OPTIMALITY CRITERIA

We define three criteria for optimality, "self-cooperating", "C-exploiting", and "D-unexploitable". We first define a "self-cooperating" strategy in terms of its self-payoff $w(X|X)$, comparing this to the self-payoffs of *ALLD* (the lowest possible self-payoff) and *ALLC* (the highest possible self-payoff):

Definition 1. A strategy X is *self-cooperating* if $w(X|X) > w(ALLD|ALLD)$. A strategy X is *totally self-cooperating* if $w(X|X) = w(ALLC|ALLC)$.

Next, we define a "C-exploiting" strategy in terms of its payoff against *ALLC*, again comparing this to the payoffs of *ALLD* (maximum) and *ALLC* (minimum).

Definition 2. A strategy X is *C-exploiting* if $w(X|ALLC) > w(ALLC|ALLC)$. A strategy X is *totally C-exploiting* if $w(X|ALLC) = w(ALLD|ALLC)$.

Third, we define a "D-unexploitable" strategy in terms of its payoff against *ALLD*, again comparing this to the payoffs of *ALLD* (maximum) and *ALLC* (minimum).

Definition 3. A strategy X is *D-unexploitable* if $w(X|ALLD) > w(ALLC|ALLD)$. A strategy X is *totally D-unexploitable* if $w(X|ALLD) = w(ALLD|ALLD)$.

More generally, we define the "relative performance" of strategy X against strategy Y to be

$$\sigma_Y(X) = \frac{w(X|Y) - \inf_Z w(Z|Y)}{\sup_Z w(Z|Y) - \inf_Z w(Z|Y)}.$$

Thus $\sigma_Y(X) = 1$ if X achieves the maximum possible score against Y, and $\sigma_Y(X) = 0$ if X achieves the minimum possible score against Y. We now consider the "relative performance" of strategy X against *ALLC* and *ALLD*, which we denote by σ_C and σ_D, respectively:

$$\sigma_C = \sigma_{ALLC}(X) = \frac{w(X|ALLC) - w(ALLC|ALLC)}{w(ALLD|ALLC) - w(ALLC|ALLC)},$$

$$\sigma_D = \sigma_{ALLD}(X) = \frac{w(X|ALLD) - w(ALLC|ALLD)}{w(ALLD|ALLD) - w(ALLC|ALLD)}.$$

Thus, a strategy is "C-exploiting" if $\sigma_C > 0$, and "totally C-exploiting" if $\sigma_C = 1$. Similarly, a strategy is "D-unexploitable" if $\sigma_D > 0$, and "totally D-unexploitable" if $\sigma_D = 1$. Note that each of these quantities is the proportion of defections by X against *ALLC* or *ALLD*, respectively.

We also define the "relative self-performance" of a strategy to be

$$\sigma_S = \frac{w(X|X) - w(ALLD|ALLD)}{w(ALLC|ALLC) - w(ALLD|ALLD)}.$$

Note that a strategy is "self-cooperating" if $\sigma_S > 0$, and "totally self-cooperating" if $\sigma_S = 1$. Now we consider what σ values a strategy must have to be considered "optimal". A "perfectly optimal" strategy X should attain the maximum possible score against all strategies Y: that is, $\sigma_Y(X) = 1$ for all Y. However, a simple argument shows that this is impossible to achieve:

Proposition 1. *No perfectly optimal strategy exists.*

See Appendix A for proof. We now consider strategies which are optimal in their interactions with cooperators and defectors. To quantify this, we apply the Maynard Smith criteria. For X to invade $ALLC$, $w(X|ALLC) > w(ALLC|ALLC)$, which implies $\sigma_C > 0$. For X to invade $ALLD$, $w(X|ALLD) = w(ALLD|ALLD)$ and $w(X|X) > w(ALLD|X)$, which implies $\sigma_D = 1$ and $\sigma_S > 0$. For X to resist invasion by $ALLC$, $w(X|X) \geqslant w(ALLC|X)$. Solving for $w(X|X)$ in terms of σ_S, and $w(ALLC|X)$ in terms of σ_C, we obtain

$$\sigma_C \geqslant \frac{R-P}{R-S}(1-\sigma_S).$$

For X to resist invasion by $ALLD$, $w(X|X) > w(ALLD|X)$. Solving for $w(X|X)$ in terms of σ_S, and $w(ALLD|X)$ in terms of σ_D, we obtain

$$\sigma_D > 1 - \frac{R-P}{T-P}\sigma_S.$$

We define a "strongly optimal" strategy as one which can invade and take over a population of defectors, establishing total self-cooperation, and resist invasion by $ALLD$ and $ALLC$. A "weakly optimal" strategy can invade and take over a population of defectors, establishing some degree of self-cooperation, and resist invasion by $ALLD$ and $ALLC$. This implies:

Definition 4. A *strongly optimal* strategy is a strategy with $\sigma_D = 1$, $\sigma_S = 1$, and $\sigma_C > 0$.

Definition 5. A *weakly optimal* strategy is a strategy with $\sigma_D = 1$, $\sigma_S > 0$, and $\sigma_C > ((R-P)/(R-S))(1-\sigma_S)$.

We now consider these criteria for APD games withm varying amounts of noise.

OPTIMALITY CRITERIA FOR NO NOISE APD

In the no noise APD, we assume that errors do not occur; a strategy always makes the choice (cooperation or defection) that it intends to make. Thus we know that $w(ALLC|ALLC) = R$, $w(ALLD|ALLD) = P$, $w(ALLC|ALLD) = S$, and $w(ALLD|ALLC) = T$ for the no noise APD. Thus $\sigma_S = (w(X|X) - P)/(R-P)$, $\sigma_C = (w(X|ALLC) - R)/(T-R)$, and $\sigma_D = (w(X|ALLD) - S)/(P-S)$.

We now consider the strategies which are most commonly discussed in the Prisoner's Dilemma: Tit for Tat (TFT), Generous Tit For Tat ($GTFT$), $ALLC$, $ALLD$, Pavlov (PAV), and Firm But Fair (FBF). We compute σ_S, σ_C, and σ_D for each strategy in an APD with no noise (Table 2).

As can be seen from the table, none of these strategies are even weakly optimal. TFT is closest to optimal, since it can invade a population of defectors and establish complete cooperation. However, since TFT does not exploit unconditional cooperators, it is vulnerable to infiltration by $ALLC$, allowing defectors to reinvade. It should be noted that all of these strategies have a "2-ply memory" or less: they react based only on the last two turns (strategy's move

TABLE 2
σ values for no noise APD

Strategy	σ_S	σ_C	σ_D
TFT (1 0 1 0)	1	0	1
GTFT (1 g 1 g)	1	0	$1-g$
ALLC (1 1 1 1)	1	0	0
ALLD (0 0 0 0)	0	1	1
PAV (1 0 0 1)	1	0	$\frac{1}{2}$
FBF (1 0 1 g)	1	0	$\frac{1}{1+g}$

followed by opponent's move). In fact, we can prove:

Proposition 2. *No* 0, 1, *or* 2-*ply strategy is strongly optimal in the APD with no noise.*

See Appendix A for proof. We note that a 2-ply strategy may be weakly optimal: the *GRIM* strategy (1 0 0 0) with initial cooperation probability $\frac{1}{3}$ has $\sigma_S = \frac{1}{9}$, $\sigma_C = \frac{2}{3}$, and $\sigma_D = 1$, and thus meets the criteria for weak optimality. In order to find a strongly optimal strategy, however, we must consider strategies of 3-ply memory or higher.

OPTIMALITY CRITERIA FOR INFINITESIMAL NOISE APD

We now consider APD games with noise. Noise is defined as a non-zero probability ε that an error occurs on a given turn: a strategy will defect when it meant to cooperate, or cooperate when it meant to defect, with probability ε. This discrepancy between intent and result may be caused by performance errors or environmental factors; these errors in implementing a rule are unavoidable in any biological context (May, 1987).

In an APD game with "infinitesimal noise", it is possible for a player's decision to be transmitted incorrectly, but the probability of this occurring is extremely low. The infinitesimal noise case can be thought of as the limiting case of finite noise, with a noise probability $\varepsilon \to 0$, or the results can be applied as an approximation for small $\varepsilon > 0$.

We note that the presence of infinitesimal noise does not change the interactions between *ALLC* and *ALLD*: thus $w(ALLC|ALLC) = R$, $w(ALLD|ALLD) = P$, $w(ALLD|ALLC) = T$, and $w(ALLC|ALLD) = S$ as in the no noise case. However, it is important to note that $w(X|Y)$ can have a discontinuity at $\varepsilon = 0$: in other words, the performance of a strategy can differ significantly between the no noise case ($\varepsilon = 0$) and the infinitesimal noise case ($\varepsilon \to 0$). For example, consider the behavior of the Tit for Tat strategy, in an APD game with its clone. For the no noise case, *TFT* and its clone will cooperate continually, giving $\sigma_S = 1$. For the infinitesimal noise case, a single error will lead to a rut of mutual defection with payoff *P*, while a second error will restore mutual cooperation. In an infinitely long game we can expect *TFT* and its clone to be in the mutual defection rut half the time, giving each an average payoff of $(R + P)/2$. Thus $\sigma_S = \frac{1}{2}$ for the *TFT* strategy. We use similar reasoning to compute σ_S, σ_C, and σ_D for *TFT*, *GTFT*, *ALLC*, *ALLD*, *PAV*, and *FBF*. We assume that $g \gg \varepsilon$ for the *GTFT* and *FBF* strategies (Table 3).

TABLE 3
σ values for infinitesimal noise APD

Strategy	σ_S	σ_C	σ_D
TFT	$\frac{1}{2}$	0	1
GTFT	1	0	$1-g$
ALLC	1	0	0
ALLD	0	1	1
PAV	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
FBF	1	0	$\frac{1}{1+g}$

Thus, we can see that none of these lower memory strategies is even weakly optimal. As in the no noise case, we can prove that no strategy of less than three ply is strongly optimal. But for infinitesimal noise, we can go even further, proving that:

Proposition 3. *No* 0, 1, *or* 2-*ply strategy is (strongly or weakly) optimal in the infinitesimal noise APD.*

See Appendix A for proof. Thus, in order to find optimal strategies for the infinitesimal noise APD, we must examine strategies with 3-ply or higher memory.

OPTIMALITY CRITERIA FOR FINITE NOISE APD

We now consider APD games with a noise level $0 < \varepsilon < \frac{1}{2}$, examining as to how a strategy's payoff changes as a function of ε. This enables us to analyse games where the approximation $\varepsilon \approx 0$ does not hold.

First, it should be noted that the expected average payoff from continued "mutual cooperation" (ie. when both strategies intend to cooperate) is less than *R* in the presence of noise. If a strategy intends to cooperate with an error level

ε, it will actually cooperate with probability $1 - \varepsilon$, and defect with probability ε. Similarly, if a strategy intends to defect, it will cooperate with probability ε, and defect with probability $1 - \varepsilon$. Thus, the payoffs to *ALLC* and *ALLD* differ from the no noise and infinitesimal noise cases

$$w(ALLC|ALLC) = (1-\varepsilon)^2 R + \varepsilon^2 P + \varepsilon(1-\varepsilon)(T+S),$$

$$w(ALLD|ALLD) = (1-\varepsilon)^2 P + \varepsilon^2 R + \varepsilon(1-\varepsilon)(T+S),$$

$$w(ALLD|ALLC) = (1-\varepsilon)^2 T + \varepsilon^2 S + \varepsilon(1-\varepsilon)(R+P),$$

$$w(ALLC|ALLD) = (1-\varepsilon)^2 S + \varepsilon^2 T + \varepsilon(1-\varepsilon)(R+P).$$

Since the σ values of a strategy are defined in terms of these payoffs, we obtain the expected σ values for *ALLC* and *ALLD*. For *ALLC*, $\sigma_S = 1$ and $\sigma_C = \sigma_D = 0$. For *ALLD*, $\sigma_S = 0$ and $\sigma_C = \sigma_D = 1$. For nearly all strategies, however, the σ values vary as a function of ε. In most cases, σ_C increases, and σ_S and σ_D decrease, with increasing ε. For example, we can compute the σ values for *FBF* (1 0 1 1) as a function of ε (Table 4). Recall that for infinitesimal noise, $\sigma_S = 1$, $\sigma_C = 0$, and $\sigma_D = 0.5$.

Now, since the σ values of a strategy vary with ε, how do we determine if a strategy is self-cooperating, C-exploiting, or D-unexploitable with finite noise? One possibility is to choose a constant "noise resistance threshold" *NRT*. Then for any given ε, we define:

Definition 6. A strategy is *self-cooperating* at noise ε if $\sigma_S \geqslant NRT$.

Definition 7. A strategy is *C-exploiting* at noise ε if $\sigma_C \geqslant NRT$.

Definition 8. A strategy is *D-unexploitable* at noise ε if $\sigma_D \geqslant NRT$.

TABLE 4
σ values for FBF (1 0 1 1) *as a function of* ε

ε	σ_S	σ_C	σ_D
0.001	0.999	0.001	0.499
0.010	0.990	0.007	0.495
0.050	0.953	0.034	0.475
0.100	0.913	0.064	0.450
0.200	0.843	0.111	0.396

Using a noise resistance threshold of $NRT = 0.9$ and the $T = 5$ payoff table, we find that *FBF* is self-cooperating for $\varepsilon \leqslant 0.117$, but is not D-unexploitable. *TFT*, on the other hand, is D-unexploitable for $\varepsilon \leqslant 0.0697$, but is not self-cooperating.

In addition to defining the optimality criteria for specific values of ε, we also examine how a strategy's self-payoff $w(X|X)$ changes as a function of ε. In particular, we define a "noise resistance coefficient" NRC_S, which gives a lower bound for $w(X|X)$ as a linear function of ε. This means that $w(X|X)_\varepsilon = e \geqslant R - (NRC_S)e$ for all e. We assume that the noise level can vary from 0 to $\varepsilon_{max} = 0.2$. Thus we define

$$NRC_S = \sup_{(e=0\ldots\varepsilon_{max})} \frac{R - w(X|X)_{\varepsilon=e}}{e}.$$

Once we have found NRC_S, this gives us a lower bound on the performance of a strategy for all $\varepsilon \leqslant \varepsilon_{max}$: we know $w(X|X) \geqslant R - (NRC_S)\varepsilon$. Thus if a strategy has $NRC_S = 5$, and assuming $R = 3$, we know $w(X|X) \geqslant 3 - 5\varepsilon$. It should also be noted that, for strategies which are not totally self-cooperating under infinitesimal noise, $NRC_S = \infty$.

It is clear that a strategy with lower NRC_S is more resistant to noise. How low must NRC_S be for a strategy to be considered "finite noise resistant"? To answer this question, we consider the results of a single error on the total self-payoff of a strategy. For *ALLC* against its clone, an error results in the pattern CdCcCc: the defector gains $(2T - 2R)$ and the cooperator loses $(2R - 2S)$ as a result of this error, for an average loss of $2R - (T + S)$ per player. For the (5,3,1,0) payoff table, each player loses an average of $2(3) - (5 + 0) = 1$ point per error, and thus

$w \approx 3 - \varepsilon$, or $NRC_S \approx 1$. For *FBF* against its clone, an error results in the pattern CdDcCc: each player loses $3R - (T + S + P) =$ three points per error, and thus $w \approx 3 - 3\varepsilon$, or $NRC_S \approx 3$. In general, if a single error leads to N straight defections on average, each player loses $(N + 1)R - (N - 1)P - (T + S)$ points per error. We define a strategy to be "finite noise resistant" if one error leads to an average of two defections or less. This implies:

Definition 9. A strategy is *finite noise resistant* if $NRC_S \leqslant 3R - P - T - S$.

Thus a strategy is finite noise resistant for the $(5, 3, 1, 0)$ payoff table if $NRC_S \leqslant 3$, and for the $(4, 3, 1, 0)$ payoff table if $NRC_S \leqslant 4$. We now propose a new standard for optimality under finite noise:

Definition 10. A strategy X is *finite noise optimal* if it is finite noise resistant, totally D-unexploitable under infinitesimal noise, and C-exploiting under infinitesimal noise.

Thus a finite noise optimal strategy must meet all the criteria for strong optimality under infinitesimal noise, as well as having a low noise resistance coefficient NRC_S. Since no strategy of less than 3-ply memory is even weakly optimal under infinitesimal noise, we know that none of these strategies are finite noise optimal.

A New Class of Strategies

Thus our main goal is to find a higher memory strategy which is "finite noise optimal". This strategy must be totally D-unexploitable, totally self-cooperating, and (at least weakly) C-exploiting under infinitesimal noise. In other words, it must have $\sigma_S = \sigma_D = 1$ and $\sigma_C > 0$ for $\varepsilon \to 0$. It must also be "finite noise resistant", having $NRC_S \leqslant 3$ for the $(5, 3, 1, 0)$ payoff table and $NRC_S \leqslant 4$ for the $(4, 3, 1, 0)$ payoff table.

FIRST DEFECTOR STRATEGIES

We now define the "first defector strategies", a class of infinite memory strategies. An *n-ply first defector strategy* for the APD chooses to cooperate or defect based on the n-move history H, together with a "memory bit" b. This bit records which player most recently made an unprovoked defection: it is set to 1 whenever the 2-move history is Cd (opponent made an unprovoked defection) and set to 0 whenever the 3-move history is cDc or cDd (player made an unprovoked defection). The strategy is defined by $2^n + 1$ numbers $(a_1, \ldots, a_{2^n+1})$, where $0 \leqslant a_i \leqslant 1$ for all i. For $i = 1 \ldots 2^n - 1$, a_i corresponds to the probability of cooperation if $H = H_i$. However, if $H = H_{2^n}$ (the n-move history is all defections) either a_{2^n} or a_{2^n+1} is used: the probability of cooperation is a_{2^n} if $b = 0$, and a_{2^n+1} if $b = 1$. For example, consider the 2-ply first defector strategy (1 0 0 [1 0]). This strategy will defect if the 2-move history is Cd or Dc, and cooperate if the 2-move history is Cc. If the history is Dd, it will cooperate if $b = 0$ (the player's unprovoked defection started the sequence of D's) and defect if $b = 1$ (the opponent's unprovoked defection started the sequence of D's). First defector strategies are denoted by FD_∞ to distinguish them from their finite memory approximations FD_n.

FINITE MEMORY APPROXIMATIONS TO *FD* STRATEGIES

Given a k-ply first defector strategy $FD_\infty = (a_1, \ldots, a_{2^k+1})$, an *n-ply approximation* FD_n to the first defector strategy is defined as follows. For $n = k$, $FD_n = (a_1, \ldots, a_{2^k-1}, g)$, where g is the strategy's "generosity" (probability of cooperation after n defections). Thus a 2-ply approximation to (1 0 0 [1 0]) would be the 2-ply strategy (1 0 0 g). For $n > k$, we define FD_n recursively by concatenating two copies of FD_{n-1}: $FD_n = FD^*_{n-1} \oplus FD_{n-1}$, where $\oplus$ is the concatenation operator and FD^*_{n-1} is defined as FD_{n-1} with:

$$g = \begin{cases} a_{2^k} & \text{if } n \text{ is odd,} \\ a_{2^k+1} & \text{if } n \text{ is even.} \end{cases}$$

Thus a 3-ply approximation to (1 0 0 [1 0]) would be (1 0 0 1)$\oplus$(1 0 0 g) = (1 0 0 1, 1 0 0 g). A 4-ply approximation would be (1 0 0 1, 1 0 0 0)$\oplus$(1 0 0 1, 1 0 0 g) = (1 0 0 1, 1 0 0 0, 1 0 0 1, 1 0 0 g). Thus, we can approximate a k-ply first defector strategy FD_∞ with an n-ply strategy FD_n for any desired $n \geqslant k$.

THE FIRM BUT FAIR STRATEGY CLASS

The Firm But Fair strategy class consists of the FBF_∞ first defector strategy, as well as its n-ply approximations FBF_n. FBF_∞ is the 2-ply first defector strategy (1 0 1 [1 0]), which behaves equivalently to "contrite TFT" (Wu & Axelrod, 1995) and "Reciprocity" (Leimar, 1997). The 2-ply approximation $FBF_2(g)$ is the standard Firm But Fair strategy (1 0 1 g). The 3-ply and higher approximations are defined as follows, where $\oplus$ is the concatenation operator:

$$FBF_3(g) = FBF_2(g=1) \oplus FBF_2(g)$$

$$= (1\ 0\ 1\ 1,\ 1\ 0\ 1\ g),$$

$$FBF_4(g) = FBF_3(g=0) \oplus FBF_3(g)$$

$$= (1\ 0\ 1\ 1,\ 1\ 0\ 1\ 0,\ 1\ 0\ 1\ 1,\ 1\ 0\ 1\ g),$$

$$FBF_n(g) = \begin{cases} FBF_{n-1}(g=1) \oplus FBF_{n-1} & \text{if } n \text{ is odd,} \\ FBF_{n-1}(g=0) \oplus FBF_{n-1} & \text{if } n \text{ is even.} \end{cases}$$

These strategies cooperate with a cooperating opponent, and defect in response to an unprovoked defection by the opponent. In the event of a sequence of defections, they will cooperate if they were the first unprovoked defector (cDd, cDdDd, etc.), defect if the opponent was the first unprovoked defector (Cd, CdDd, etc.), or cooperate with probability g if their memory capacity is exceeded (the first defection occurred at least n turns ago, so they do not know who defected first).

THE FIRM PAVLOV STRATEGY CLASS

The Firm Pavlov class of strategies is a variation of Firm But Fair which also can exploit an unconditionally cooperating opponent. The Firm Pavlov strategy class consists of the FP_∞ first defector strategy, as well as its n-ply approximations FP_n. FP_∞ is the 3-ply first defector strategy (1 0 0 1, 1 0 1 [1 0]). The 3-ply approximation $FP_3(g)$ is the strategy (1 0 0 1, 1 0 1 g). The 4-ply and higher approximations are defined as follows, where $\oplus$ is the concatenation operator:

$$FP_4(g) = FP_3(g=0) \oplus FP_3(g)$$

$$= (1\ 0\ 0\ 1,\ 1\ 0\ 1\ 0,\ 1\ 0\ 0\ 1,\ 1\ 0\ 1\ g),$$

$$FP_n(g) = \begin{cases} FP_{n-1}(g=1) \oplus FP_{n-1} & \text{if } n \text{ is odd,} \\ FP_{n-1}(g=0) \oplus FP_{n-1} & \text{if } n \text{ is even.} \end{cases}$$

These strategies are similar to FBF_n, but defect after they successfully exploit the opponent (3-move history cDc). They do not defect after a 3-move history of dDc, because this would prevent the strategy from breaking out of sequences of mutual defection against itself.

We now examine the performance of various *FBF* and *FP* strategies, searching for strategies which meet the criteria for infinitesimal noise optimality, finite noise optimality, and arbitrary noise optimality.

Optimality of *FBF* and *FP*

OPTIMALITY UNDER NO NOISE AND INFINITESIMAL NOISE

For the no-noise case, $FBF_n(g)$ has $\sigma_S = 1$ and $\sigma_C = 0$, since it will cooperate continually against its clone or *ALLC*. Its value of σ_D depends on both n and g. Against *ALLD*, $FBF_n(g)$ will cooperate with probability g if it has made at least $\lfloor n/2 \rfloor$ straight defections, and defect otherwise. Thus it cooperates once every $\lfloor n/2 \rfloor + 1/g$ turns. This gives us

$$\sigma_D = \frac{(\lfloor n/2 \rfloor - 1)g + 1}{(\lfloor n/2 \rfloor)g + 1}.$$

Thus, $FBF_n(g)$ has $\sigma_D = 1$ for $g = 0$, and the infinite memory FBF_∞ strategy also has $\sigma_D = 1$. $FP_n(g)$ with initial cooperation has σ values identical to $FBF_n(g)$. However, $FP_n(g)$ with initial defection will exploit *ALLC* continually, giving it $\sigma_C = 1$. Thus for all n, $FP_n(0)$ with initial defection has $\sigma_C = \sigma_D = \sigma_S = 1$, and thus it is strongly optimal for no noise. Similarly, the infinite memory FP_∞ with initial defection has $\sigma_C = \sigma_D = \sigma_S = 1$, and is strongly optimal for the no noise APD.

For infinitesimal noise, we first note that an *FBF* or *FP* strategy's value of σ_D will be unchanged from the no noise case. The *FBF* strategies have $\sigma_C = 0$ as in the no noise case, but the *FP* strategies have $\sigma_C = \frac{1}{3}$. This results since, in cooperation, only an error by *FP* will cause it to start exploiting *ALLC*, but while exploiting, an error by either player will cause it to resume cooperation. Thus *FP* is twice as likely to leave the exploiting rut as to enter, and it will spend one-third of the time exploiting *ALLC*.

We now consider self-cooperation under infinitesimal noise, computing the proportion of time that an *FBF* or *FP* strategy will be in mutual cooperation with its clone. We first examine the $FP_3(0)$ strategy (1 0 0 1, 1 0 1 0), demonstrating that $\sigma_S = 1$. But how can $FP_3(0)$ be totally self-cooperating if its generosity (probability of cooperation after dDd) is 0? It is possible for $FP_3(0)$ to fall into a rut of mutual defection with its clone, but it requires two errors to enter the rut and only a single error to restore cooperation. It thus enters the rut once every $1/\varepsilon^2$ turns, and leaves the rut every $1/\varepsilon$ turns. Hence, in an infinitely long game, we can expect $FP_3(0)$ to be in the defection rut with probability:

$$\frac{1/\varepsilon}{1/\varepsilon^2 + 1/\varepsilon} = \frac{\varepsilon}{1+\varepsilon} \approx 0.$$

Thus $FP_3(0)$ has $\sigma_S = 1$, $\sigma_D = 1$, and $\sigma_C = \frac{1}{3}$. We have found a strategy that meets the criteria for strong optimality in the infinitesimal noise case!

We now compute the proportion of time that an *FBF* or *FP* strategy will be in a rut of mutual defection with its clone in the infinitesimal noise APD. In general, it takes $\lceil n/2 \rceil$ errors to enter the mutual defection rut, and the strategies can escape the rut with probability g ($g \geqslant \varepsilon$). Thus the proportion of time spent in this state is

$$\frac{\varepsilon^{\lceil n/2 \rceil}}{g + \varepsilon^{\lceil n/2 \rceil}}.$$

For $\varepsilon \approx 0$, this quantity is negligible unless $n \leqslant 2$ and g is $O(\varepsilon)$. Thus for *FBF* and *FP* strategies of 3-ply or higher memory, $\sigma_S = 1$. Similarly, $\sigma_S = 1$ for FBF_∞ and FP_∞. For the 2-ply *FBF* strategy, $\sigma_S = \frac{1}{2}$ for $g = \varepsilon$, and $\sigma_S = 1$ for $g \gg \varepsilon$.

A strategy which is strongly optimal for infinitesimal noise must have $\sigma_S = \sigma_D = 1$ and $\sigma_C > 0$. Thus $FP_n(0)$ is strongly optimal for all n, and FP_∞ is also strongly optimal for the infinitesimal noise APD.

OPTIMALITY UNDER FINITE NOISE

For the finite noise case, we first calculate the noise resistance coefficient NRC_S for *FP* and *FBF* strategies with various values of n and g. Recall that for a strategy X with noise resistance coefficient NRC_S, $w(X|X) \geqslant R - (NRC_S)\varepsilon$ for all $\varepsilon \leqslant 0.2$, and a strategy with $NRC_S \leqslant 3$ is noise resistant (Table 5).

For $T = 5$, all *FBF* strategies with $n \geqslant 5$ are noise resistant, as are all *FP* strategies with $n \geqslant 7$. FP_3 is noise resistant for $g > 0.4$, FBF_3 is noise resistant for $g > 0.25$, and FP_5 is noise resistant for $g > 0.047$. Compare these strategies to FBF_2, which is noise resistant only for $g = 1$.

For $T = 4$, however, FP_5 is noise resistant for all values of the generosity g. Thus the $FP_5(0)$ strategy meets all the criteria for finite noise optimality: it is finite noise resistant, and under infinitesimal noise, it is C-exploiting, totally self-cooperating, and totally D-unexploitable. We have thus accomplished the first part of our goal, to find a finite memory strategy which is finite noise optimal. $FP_n(0)$ is also finite noise optimal for $n > 5$, as is FP_∞.

We also note that the $FP_5(0)$ and $FBF_5(0)$ strategies are self-cooperating ($\sigma_S \geqslant 0.9$) and D-unexploitable ($\sigma_D \geqslant 0.9$) for noise levels up to approximately 0.09. Thus both strategies have nearly maximum payoff against *ALLD*, and

TABLE 5

Noise resistance coefficients of $FP_n(g)$ and $FBF_n(g)$, $T = 5$

g	FBF_2	FP_3	FBF_3	FP_5	FBF_5
0	∞	5	5	3.01	3
0.01	201	4.86	4.82	3.01	3
0.02	101	4.73	4.65	3.01	3
0.05	41	4.36	4.19	3	3
0.10	21	3.83	3.59	3	3
0.20	11	3.21	3.04	3	3
0.50	5	3	3	3	3
1	3	3	3	3	3

nearly maximum self-payoff, even for relatively high noise levels.

Next, we examine the strategies' evolutionary interaction with *ALLD*. At low noise levels, only a small proportion of $FP_3(0)$ or $FBF_3(0)$ is needed to invade a population of defectors: for $T = 4$ and $\varepsilon = 0.01$, these strategies can invade *ALLD* as long as they make up $\frac{1}{200}$ of the population. This proportion increases (approximately) linearly for increasing noise. The FP_3 and FBF_3 strategies can also resist an *ALLD* invasion for $T = 4$ and $\varepsilon \leqslant 0.1$. For $\varepsilon \geqslant 0.1$, some very generous versions of these strategies will allow *ALLD* to invade, but the 5-ply *FP* and *FBF* strategies are immune to *ALLD* invasion for all $\varepsilon \leqslant 0.2$.

OPTIMALITY UNDER ARBITRARY NOISE

We also note that the infinite memory *FBF* and *FP* strategies have another useful feature: each strategy can recover from any finite sequence of errors, and restore cooperation with its clone, within two turns. Such a strategy can be thought of as not merely "noise resistant", but "noise proof". This is in contrast to strategies such as $FP_3(0)$. If two errors occur in quick succession in a game between $FP_3(0)$ and its clone, the players can get stuck in a "Dd rut", unable to restore cooperation until another error occurs. If the probability of an error is constant over the course of a game, $FP_3(0)$ will be in the rut of mutual defection with probability 0. If the error probability varies widely, however, $FP_3(0)$ can be permanently stuck in a rut of mutual defection (for example, if errors occur on the first and third turns of a game, then never again). This argument applies to any non-generous, finite memory strategy. In fact, we can prove:

Proposition 4. *No strategy X with finite memory is noise proof and totally D-unexploitable.*

See Appendix A for proof. Thus in a rapidly changing environment, it may be essential for strategies to be noise proof, and infinite memory strategies would be expected to evolve. In most real life examples, however, the probability of a long string of errors is low enough (ε^k for a string of k errors) such that it is sufficient for a strategy to be "noise resistant", and infinite memory is unnecessary.

Performance of *FBF* and *FP*

We have shown above that higher-memory strategies of the *FP* and *FBF* strategy classes meet many or all of our optimality criteria. In particular, we have found that the $FP_n(0)$ strategies are finite noise optimal for $n > 5$ (for $T = 4$) or $n > 7$ (for $T = 5$). The corresponding *FBF* strategies meet almost all of the optimality criteria, but are not C-exploiting.

We must now ask an essential question: do strategies which meet these optimality criteria perform well in (evolutionary and non-evolutionary) APD simulations? To answer this question, we first ran several round-robin tournaments, using a format similar to Axelrod's (1981). The first four tournaments consisted of 63 strategies: $g1$, $g2$, $g3$, *PAV*, *RAND*, *ALLD*, *ALLC*, 11 FBF_2 strategies (generosity 0-1), 10 FBF_3 strategies (generosity 0-0.9), 11 FP_3 strategies (generosity 0-1), 11 FBF_5 strategies (generosity 0-1), 11 FP_5 strategies (generosity 0-1), FBF_∞, and FP_∞. Tournaments were run at noise levels of 0.01 and 0.05, and with the payoff tables $(4, 3, 1, 0)$ and $(5, 3, 1, 0)$. The FP_5 strategies won all four tournaments: for the two tournaments with noise 0.01, $FP_5(0.2)$ was the champion, and for the two tournaments with noise 0.05, $FP_5(1)$ was the champion. In general, *FP* strategies outperformed *FBF* strategies: $FP_n(g)$ outscored $FBF_n(g)$ for any given n and g. Also, for the finite memory strategies, performance improved with increasing n. However, the infinite memory strategies FBF_∞ and FP_∞ performed poorly in these tournaments. We hypothesized that this resulted from poor performance against the strategies with low memory and low generosity. For example, consider FBF_∞ against the Tit for Tat strategy, $FBF_2(0)$. An error by FBF_∞ would be quickly corrected, but an error by *TFT* would result in a rut of mutual defection. To test this hypothesis, a fifth tournament was run with 58 of the 63 strategies, excluding the five non-generous ($g = 0$) *FBF*/*FP* strategies. A 0.01 noise level, and the $(5, 3, 1, 0)$ payoff table, were used. As expected, FP_∞ was the champion. We also noted that the optimum value of g varied with the memory n.

While the 2-ply *FBF* strategy performed best for $g = 1$, the higher memory strategies performed best for $g = 0.2$ ($\varepsilon = 0.01$) or $g = 0.7$ ($\varepsilon = 0.05$).

Next we selected 29 of the *FBF* and *FP* strategies: 4 FBF_2 strategies, 5 FBF_3 strategies, 6 FP_3, 6 FBF_5 strategies, and 6 FP_5 strategies with varying generosities, as well as FBF_∞, and FP_∞. Each of the strategies was run individually in a 10-strategy round-robin tournament against *g*1, *g*2, *g*3, *PAV*, *RAND*, *ALLD*, *ALLC*, *TFT* ($FBF_2(0)$), and *FBF* ($FBF_2(1)$). A 0.01 noise level and $(4, 3, 1, 0)$ payoff table were used. Of the 29 strategies tested, 28 won their tournament: the only exception was $FBF_3(0)$, which finished second to $FBF_2(1)$. The $FP_5(0.2)$ strategy won its tournament by the largest margin, and *FP* strategies again outperformed *FBF* strategies.

From these results, it is clear that both the *FBF* and *FP* strategies perform well in round-robin tournaments, with *FP* performing significantly better than *FBF*. The ability to exploit unconditional cooperators (without significantly reducing the strategy's ability to cooperate with conditional cooperators) is essential for success in a round-robin tournament simulation. The $FP_5(0.2)$ strategy was particularly successful in the round-robin tournaments we conducted. It is perfectly self-cooperating, mostly D-unexploitable, and somewhat C-exploiting; the non-zero generosity value also allows it to cooperate with low-memory/low-generosity conditional cooperators such as *TFT*.

Evolutionary Optimality Criteria

A BRIEF OVERVIEW OF STANDARD EVOLUTIONARY MODELS

Theoretical biologists have proposed a number of evolutionary models for population dynamics in the iterated Prisoner's Dilemma. In all of these models, evolution is driven by natural selection: strategies which earn higher average payoffs have higher "evolutionary fitness", and are able to survive and reproduce. The "Genetic Algorithm" model used by Axelrod (1987) and others represents each strategy as a "chromosome", a collection of genes describing the strategy's behavior in any given situation. This model creates genetic variation in a strategy's offspring by both mutation (randomly occurring changes in an individual's genes) and crossover (genetic recombination of features from two different "parent" strategies). This approach enables continued evolution of new strategies, but relies heavily on chance: the evolution of a population is strongly dependent on which random mutations or recombinations occur, as well as the specific parameters of the genetic model.

Nevertheless, one advantage of genetic algorithm models is that they suggest methods through which strategies may evolve. Lindgren (1981) investigated the evolution of *n*-ply strategies through a genetic algorithm model which allowed for gene duplications and splits (increasing or decreasing the memory *n*) as well as point mutations. He found that higher memory strategies can evolve in this model, and these longer genomes can often be more successful than lower memory players. Thus Lindgren's work complements the results presented here: he presents the methods by which higher memory strategies can evolve, while we focus on the conditions under which higher memory is evolutionarily beneficial (and hence, likely to evolve). In particular, higher memory that is likely to evolve when meeting all three of the optimality criteria is important: when an organism is likely to interact with others employing a wide range of different strategies.

A second model is the invasion model proposed by Nowak and Sigmund (1992, 1993). This model assumes that a large homogeneous population of a given stategy X is visited periodically by mutant strains. In this model, a mutant strain Y is selected randomly from the hyperspace of all possible *n*-ply strategies, then Maynard Smith's invasion criteria (1982) are used to calculate whether Y can invade X. This invasion will occur if $w(Y|X) > w(X|X)$, or $w(Y|X) = w(X|X)$ and $w(Y|Y) > w(X|Y)$. The model assumes that if Y invades X, Y will take over the population; otherwise, the initial population will continue. Invasion models of this sort tend to result in cyclical or complex invasion behavior: for example, *ALLD* is invaded by *TFT*, which is invaded by *GTFT*, which is invaded by *ALLC*, which is invaded by *ALLD*. Though it is a fairly realistic model of population dynamics, this model is very dependent on which mutants are chosen to invade, and its complex time-dependent behavior makes it very difficult to decide as to which strategy is in some sense "optimal". One solution to this problem is given by the notion of an "evolutionarily stable strategy" (ESS). An ESS is a strategy X which cannot be invaded by any other strategy: that is, for all strategies Y, $w(Y|X) < w(X|X)$, or $w(Y|X) = w(X|X)$ and $w(Y|Y) \leqslant w(X|Y)$. An ESS is in one sense an "evolutionarily optimal" strategy: once it is established in a population, no other single strategy can invade. However, as shown by Boyd and Lorberbaum (1987), an ESS may be invaded by a combination of two mutant strategies. More importantly, an ESS may have difficulty establishing itself in a population, because it has difficulty invading other strategies. In fact, some evolutionarily stable strategies are "inaccessible" in the sense that they are unable to invade any other homogeneous population (Nowak, 1990).

Thus, it is clear that a measure of evolutionary "optimality" should take into account not only a strategy's resistance to invasion (evolutionary stability) but its ability to invade other strategies (evolutionary potency). One model which considers both of these factors is the "Adaptive Dynamics" model used in Nowak & Sigmund (1989) and Hofbauer & Sigmund (1990). This model assumes that an initial homogeneous population of strategy S has numerous offspring X, each slightly different from S. Each of these offspring interacts with all other offspring, and the variety with the highest total payoff survives. Assuming a symmetric distribution of offspring around S, the offspring with the highest total payoff will be the offspring with the highest payoff against the parent. We thus define the fitness function $F(X)$ to be $w(X|S)$. This leads to the adaptive dynamics equation

$$\frac{\mathrm{d}S}{\mathrm{d}t} = h\nabla F,$$

where the gradient ∇F is evaluated at $X = S$, and h is a constant corresponding to the rate of variation (Kraines & Kraines, 2000). Thus the population evolves in the direction which is most advantageous for the single mutant (Nowak & Sigmund, 1989). One use of this model is to examine an individual strategy by allowing it to evolve according to the adaptive dynamics equation: this process of "self-evolution" gives information about the stability of the strategy, as well as its possible evolutionary path. We will examine the self-evolution of FBF_2, FBF_3, and FP_3 using this model. Again, the behavior of this dynamical system is time-dependent and dependent on the initial strategy chosen, making it difficult to argue that a given strategy is "optimal". But, as in the Nowak–Sigmund model, we can find strategies which are "optimal" in the sense of resistance to invasion. A "relative evolutionarily stable strategy" (RESS) is a strategy which cannot be invaded (according to the Maynard Smith criterion) by any nearby strategy. In other words, strategy X is an RESS in some space H of strategies if, for some $\delta > 0$, for all strategies Y with $|X - Y| < \delta$, either $w(Y|X) < w(X|X)$, or $w(Y|X) = w(X|X)$ and $w(Y|Y) \leqslant w(X|Y)$. Once we have found an RESS X, we can also examine its "basin of attraction": the set of strategies in H which evolve into X. An RESS with a large basin of attraction is in some sense "optimal", since it is resistant to (local) invasion and can eventually take over many self-evolving populations. As Kraines and Kraines state, the adaptive dynamics model is "consistent with modern evolutionary theory for a population with limited genetic diversity and frequent minor mutations" (2000). However, it excludes the possibility of invasion by strategies which are not similar to the original strategy, and thus does not allow for major mutations, migrants, or interactions of multiple populations. We will propose another evolutionary model which addresses these shortcomings, but first we examine the self-evolution of various FBF and FP strategies.

SELF-EVOLUTION OF *FBF* AND *FP*

To examine the self-evolution of the FBF and FP strategies, we use a discrete approximation to the adaptive dynamics equation as in Kraines & Kraines (2000). This algorithm takes three parameters: the starting strategy $S(0) = [S_1(0), \ldots, S_{2^n}(0)]$, the step size δ, and the rate of variation h. For each time step t, it modifies S using the following equations:

$$\Delta S = S(t+1) - S(t) = h\Delta F,$$

$$\Delta F[i] = F(S + \delta\hat{x}_i) - F(S)$$

where $\hat{x}_i$ is the vector with a 1 in position i and zeros elsewhere. For each simulation, we self-evolve the strategy with $T = 4$, $\varepsilon = 0.01$, $\delta = 0.01$, and $h = 0.1$.

We first self-evolved the $FBF_2(0)$ strategy, (1 0 1 0). Its probabilities of cooperation after Cd and Dd quickly increased, leveling off after 2×10^5 generations to (1 0.2607 1 0.6554). Next, we self-evolved the $FBF_3(0)$ strategy, (1 0 1 1, 1 0 1 0). Its probability of cooperation after dDd increased rapidly at first, then much more slowly, leveling off after 10^7 generations. The resulting strategy was (1 0 1 1, 1 0 1 0.6566).

We note several interesting facts from the self-evolution of these two strategies. Both strategies increase in generosity, evolving toward increasing self-cooperation and decreasing D-unexploitability. Both generosity parameters seem to level off at approximately $\frac{2}{3}$, but the $FBF_2(0)$ strategy also evolves some unconditional cooperation (cooperating over $\frac{1}{4}$ of the time after Cd). Thus $FBF_2(0)$ evolves away from the FBF family of strategies, but $FBF_3(0)$ remains an FBF strategy under evolution. The FBF_3 strategy also evolves much more slowly than FBF_2, suggesting a higher degree of evolutionary stability.

Finally, we self-evolved the $FP_3(0)$ strategy, (1 0 0 1, 1 0 1 0). Its probability of cooperation after dDd increased rapidly at first, then much more slowly, leveling off after 10^7 generations. However, its probability of cooperation after cDc also increased (slowly and at a fairly constant rate), to $1 - \varepsilon$ after 5×10^6 generations. The resulting strategy is (1 0 1 1, 1 0 1 0.6566).

Thus both the $FP_3(0)$ and $FBF_3(0)$ strategies evolve to the relative evolutionarily stable strategy $FBF_3(0.6566)$. We can conclude from this that FBF_3 is much more stable than FP_3 under self-evolution. This result is likely to apply in the general case: there is no evolutionary advantage to exploitation when all strategies are sufficiently similar.

EVOLUTIONARY FITNESS AND THE DOMINANCE CRITERION

We now consider the question of a general measure of evolutionary performance. Are any of the previously mentioned models suitable for this measure? To answer this question, we consider what properties the measure should have, and apply these to the five models/techniques discussed above: genetic algorithms, Nowak–Sigmund invasion model, ESS theory, self-evolution, and RESS theory. We argue that three properties are essential. First, the measure must take into account both a strategy's evolutionary stability (resistance to invasion) and evolutionary potency (ability to invade other strategies). Second, the measure must be able to evaluate a strategy's performance against a large number of different strategies, not only a carefully chosen few. In particular, it must consider interactions with strategies which are substantially different from the given strategy, not only its evolutionary kin. Third, the measure must not be time-varying: it should not depend on the current composition of the population, but only on the set of strategies which is initially present.

ESS theory violates the first property, since it only takes evolutionary stability into account. Self-evolution violates the second property, since it evaluates only the given strategy, not its relation to other strategies. The RESS theory also violates the second property, since it considers the interaction of a strategy only with strategies very similar to it. Using the genetic algorithm or Nowak–Sigmund invasion models, the proportion of a strategy tends to vary significantly (and possibly chaotically) over time, so it is difficult to postulate a non-time-varying measure which accurately reflects a strategy's performance. Thus none of the commonly used models of the evolutionary APD lend themselves easily to a general measure of evolutionary performance.

In order to derive our measure, we return to the Maynard Smith invasion criteria (1982). From Maynard Smith, strategy X invades strategy Y if $w(X|Y) > w(Y|Y)$, or $w(X|Y) = w(Y|Y)$ and $w(X|X) > w(Y|X)$. If strategy X invades strategy Y, we write $X > Y$. Otherwise, we write $X \not> Y$. We must first observe that whether X invades Y and whether Y invades X are independent: thus there are four possibilities:

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

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

III. $X > Y$ and $Y > X$. In this case, no matter what the initial proportions of strategies X and Y, the two strategies reach a balance in which the population is a mix of X and Y. We say that X and Y are *stable*, and write $X \leftrightarrow Y$. To be more precise, we can calculate the proportion p of strategy X at equilibrium

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

Thus we write $X \overset{p}{\leftrightarrow} Y$.

IV. $X \not> Y$ and $Y \not> X$. In this case, either strategy X or strategy Y will take over the population, driving the other strategy to extinction, depending on the initial proportions of the two strategies. We say that X and Y are *bistable*, and write $X \Leftrightarrow Y$. To be more precise, we can calculate the minimum proportion m of strategy X needed to take over the population

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

Thus we write $X \overset{m}{\Leftrightarrow} Y$. We must also deal with the case $w(X|Y) = w(Y|Y)$ and $w(Y|X) = w(X|X)$, in which case m is defined as 0.5.

We now define the *dominance measure* $\mathrm{dom}(X|Y)$, which is a measure of the relative evolutionary performance of strategies X and Y.

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

Thus the measure $\mathrm{dom}(X|Y)$ is between 0 and 1, with higher dominance scores corresponding to a better evolutionary performance of X against Y. It should also be noted that $\mathrm{dom}(X|Y) + \mathrm{dom}(Y|X) = 1$, thus the dominance score presents a "constant-sum" measure of relative evolutionary performance (i.e. if X performs well against Y, Y performs poorly against X).

The measure $\mathrm{dom}(X|Y)$ measures evolutionary fitness in a substantively different manner than other models. Most evolutionary models assume that a homogeneous population is invaded by a small number of mutants: these mutants can be nearly identical to the original population (as in the adaptive dynamics model) or very different (as in the Nowak–Sigmund invasion model). The dominance measure, on the other hand, assumes the mixing of two large homogeneous populations, and the evolution of the resulting population until it reaches a stable state.

Finally, we define the dominance score of a strategy X to be its average dominance score against all strategies it interacts with. Assuming a population of opponents Y_i with varying frequencies $f(Y_i)$, with $\sum_i f(Y_i) = 1$

$$\mathrm{dom}(X) = \sum_i f(Y_i)\,\mathrm{dom}(X|Y_i).$$

Thus a strategy's dominance score is a measure of its average relative evolutionary fitness. It has several advantages over other fitness measures: since it is time-invariant, we can compute a single score (independent of the current proportions of each strategy) for any two strategies. This allows us to measure the evolutionary performance of a strategy in a round-robin tournament format similar to Axelrod's (1984), except that dominance scores (rather than average payoffs) are computed for each pair of strategies. Since the dominance measure is a constant-sum, the performance of a strategy is not affected as much by the set of strategies it interacts with: we expect successful strategies to have high dominance scores against most other strategies, while in the normal round-robin tournament format, even "successful" strategies will have poor average payoffs against strategies such as *ALLD*. The disadvantage of the dominance measure is that it does not take into account the more complex interactions that result from the mixing of more than two strategies: in a multi-strategy interaction, a strategy's dominance score is essentially its performance after the first round of evolutionary "battles". But once the system evolves and weaker strategies die off, the strategy may be interacting with a very different set of strategies, dramatically affecting its evolutionary performance. However, there appears to be no way of taking this evolutionary complexity into account without sacrificing our main objective: a simple and useful measure of overall evolutionary fitness. While a strategy with a very high dominance score may not succeed in every evolutionary interaction, it is likely to perform well in the great majority of such interactions. Thus we consider evolutionary dominance as another criterion for optimality: an "optimal" strategy must have a high dominance score against a large and varied set of opponents. In particular, we consider an n-ply strategy to satisfy this criterion if it has a high dominance score against the set of all n-ply strategies.

Evolutionary Performance of *FBF* and *FP*

We now examine the evolutionary performance of the FBF and FP strategies through several "round-robin dominance tournaments". In a round-robin dominance tournament, we assume that the frequency of each strategy is equal, and compute $\mathrm{dom}(X)$ for each strategy X. We then compare each strategy's dominance score; the strategy with the highest dominance wins the tournament.

ROUND-ROBIN DOMINANCE TOURNAMENTS

We first conducted round-robin dominance tournaments among the same 63 strategies as for our initial round-robin tournaments. The tournaments were run with the $(4, 3, 1, 0)$ and $(5, 3, 1, 0)$ payoff tables, and a noise level of 0.01 was used.

Both tournaments were won by the FBF_∞ strategy: this strategy dominated most of the other 62 strategies, and had $\mathrm{dom}(X|Y) > 0.5$ against every other strategy. FBF strategies outperformed FP strategies: $FBF_n(g)$ outscored $FP_n(g)$ for all n and g. Additionally, performance improved with increasing memory: $FBF_m(g)$ outscored $FBF_n(g)$, and $FP_m(g)$ outscored $FP_n(g)$, for $m > n$. In addition, strategies with relatively high generosities tended to perform better. The optimum value of g varied with the memory n: $g = 1$ was optimal for 2-ply strategies, $g = 0.6$ was optimal for 3-ply strategies, and $g = 0.7$ was optimal for 5-ply strategies.

These results are not unexpected: since FP evolves into FBF in the adaptive dynamics model, we would expect that FBF strategies have a higher dominance score than FP in a head-to-head competition. That is, $\mathrm{dom}(FBF|FP) > \mathrm{dom}(FP|FBF)$, so $\mathrm{dom}(FBF) > \mathrm{dom}(FP)$ in tournaments consisting mainly of FBF and FP strategies.

DOMINANCE VERSUS CORNER AND EDGE STRATEGIES

We now ask a more interesting question: how do the FBF and FP strategies perform against the set of all n-ply strategies? It is very difficult to compute the dominance score of a given strategy against all n-ply strategies. An n-ply strategy has 2^n parameters, each of which can vary continuously from 0 to 1, so to compute this score we must integrate $w(X|Y)$ over a continuous 2^n-dimensional subspace of R^{2^n}. Rather than performing the integration, we approximate the strategy's performance by considering only "corner strategies" (n-ply strategies for which each parameter is either 0 or 1), or "edge strateges" (n-ply strategies for which each parameter is either 0, $\frac{1}{2}$, or 1). There are 2^{2^n} n-ply corner strategies, and 3^{2^n} n-ply edge strategies.

We considered the performance of each of the 63 strategies in the previous tournament: for each, we calculated an average dominance score against the 256 3-ply corner strategies, and the 6561 3-ply edge strategies. A 0.01 noise level was used.

The $FP_5(0.1)$ strategy was the champion for three of the four tournaments (vs. corner strategies with $T = 5$, vs. edge strategies with $T = 5$ and 4). $FP\infty$ won the other tournament (vs. corner strategies with $T = 4$). Both strategies performed consistently well, with dominance scores between 0.88 and 0.96 for all four tournaments. In general, FP strategies outperformed FBF strategies: all of the top 15 strategies or each tournament were FP strategies. Performance generally improved with increasing memory for the finite memory strategies, though in some tournaments the 5-ply strategies (with generosity 0.1 or 0.2) could beat the infinite-memory strategies. The optimum values of g for each strategy were significantly lower than for the previous round-robin dominance tournaments: $g = 0.4$ for 2-ply strategies, and $g = 0.1$ for 3-ply memory and higher.

It is clear from these results that the FP strategies (particularly those with low but non-zero generosities) perform extremely well on an average in evolutionary interactions against 3-ply strategies. Though FBF strategies outperform FP strategies in head-to-head dominance tournaments, FP has a significantly higher dominance score when averaged over all strategies.

DOMINANCE OF 3-PLY CORNER AND EDGE STRATEGIES

Next, we examine the space of 3-ply strategies in more detail, comparing the dominance scores of all corner and edge strategies rather than focusing on a specific group of strategies. Round-robin dominance tournaments were run for the 256 3-ply corner strategies, and for the 6561 3-ply edge strategies. The tournaments were run for both $T = 4$ and 5, and a 0.01 noise level was used.

Interestingly, the top performer out of all the corner strategies for $T = 4$ was (1 0 0 1, 1 0 1 0): this is the $FP_3(0)$ strategy! Its average dominance score was 0.9380, significantly (0.04) higher than any other strategy. The second place strategy was $FBF_3(0)$. Since dominance score is a measure of overall evolutionary fitness, this suggests that $FP_3(0)$ is in some sense evolutionarily optimal among the 3-ply corner strategies. Even more interestingly, $FP_3(0)$ was also the top performer among the 6561 3-ply edge strategies, with a dominance score of 0.9514. This suggests that it is also likely to have the highest dominance score over the entire space of 3-ply strategies, implying that $FP_3(0)$ is the *evolutionarily optimal* 3-ply strategy for $T = 4$.

For $T = 5$, $FP_3(0)$ was again the top corner strategy, finishing significantly (0.034) higher than any other strategy with a dominance score of 0.8910. The second place strategy was a slightly more cooperative version of GRIM (1 0 0 0, 1 1 0 0). However, the top edge strategy was not $FP_3(0)$ but a slightly more exploiting variant (1 0 0 1, 1 0 0.5 0), with a dominance score of 0.9327. $FP_3(0)$ finished eighth, with a dominance score of 0.8796. This suggests that the 3-ply edge strategy with the highest dominance score over the entire space of 3-ply strategies is likely to be of the form (1 0 0 1, 1 0 k 0), where $0.5 < k < 1$. To test this hypothesis, we next compared the dominance scores of strategies (1 0 0 1, 1 0 k 0) against all 3-ply edge strategies. The winner of this tournament was $k = 0.59$, with a dominance score of 0.9361. This strategy, (1 0 0 1, 1 0 0.59 0), is likely to be evolutionarily optimal in the space of all 3-ply strategies. Thus $FP_3(0)$ appears to be optimal for $T = 4$, but for $T = 5$, strategies which sometimes defect after dDc perform slightly better.

Conclusions

When is a strategy "optimal" in the Alternating Prisoner's Dilemma with noise? Since no strategy can achieve a maximum score against every other strategy, we search instead for *optimality criteria* which are strongly correlated with a high average performance in various Prisoner's Dilemma models (including round-robin tournaments and evolutionary invasion models). In order to perform well against a variety of other strategies, an "optimal" strategy must be able to achieve mutual cooperation with its clone, resist exploitation by defectors, and exploit unconditional cooperators. These three criteria (*self-cooperating*, *D-unexploitable*, *C-exploiting*) were defined and examined for Alternating Prisoner's Dilemma games with no noise, infinitesimal noise, finite noise, and arbitrary noise. Most of the strategies commonly discussed in the Alternating Prisoner's Dilemma literature are "2-ply strategies" such as Pavlov and Firm But Fair, "low memory" strategies which make decisions based only on the last two turns. However, none of these strategies can simultaneously meet all three of the optimality criteria: a 2-ply strategy which is totally unexploitable by defectors, and can exploit unconditional cooperators, cannot attain any level of self-cooperation in the infinitesimal noise APD.

Higher memory strategies such as *Firm Pavlov*, however, can meet all the three optimality criteria: the $FP_3(0)$ strategy (1 0 0 1, 1 0 1 0) is "totally self-cooperating", "totally D-unexploitable", and "C-exploiting", and thus is optimal for the infinitesimal noise APD. Other members of the Firm Pavlov strategy class (5-ply memory and higher) can meet even more rigorous criteria for finite noise optimality.

These optimality criteria can be applied to analyse other Prisoner's Dilemma games with or without noise, including the simultaneous iterated Prisoner's Dilemma. The strategies which meet the criteria are likely to be successful in a variety of other performance measures. The Firm Pavlov and Firm But Fair strategy classes were shown to perform well in standard round-robin tournament simulations, adaptive dynamics models, and evolutionary invasion models. In addition to considering the standard models, we also proposed the *dominance criterion*, a general measure of evolutionary performance. This measure takes into account both evolutionary stability (ability to resist invasion) and evolutionary potency (ability to invade other strategies), producing a combined measure which can be used to examine the relative evolutionary performance for any set of strategies. We conducted a number of "dominance tournaments", a round-robin tournament format similar to Axelrod's, except that dominance scores (instead of average payoffs) are computed for each pair of strategies. In particular, the Firm Pavlov strategies achieved extremely high dominance scores against the set of all 3-ply edge strategies, suggesting that these strategies will be extremely successful in a wide variety of evolutionary interactions. In fact, when a round-robin dominance tournament was conducted among all 3-ply edge strategies, the winner was a 3-ply Firm Pavlov strategy.

To succeed in the Prisoner's Dilemma, a strategy must be "friendly" enough to cooperate with its clone, "pragmatic" enough to exploit unconditional cooperators, and "wary" enough to resist exploitation by defectors. The Firm Pavlov strategy $FP_3(0)$ can be thought of as a model for this type of behavior: it responds to cooperation with cooperation, except when it can continue exploiting an unconditional cooperator. Similarly, it responds to defection with defection, except when the opponent's defection is in response to its provocation. This 3-ply strategy combines the advantages of several 2-ply strategies: like Firm But Fair, it is self-cooperating under noise; like Tit for Tat, it is unexploitable; and like Pavlov, it can exploit unconditional cooperators. 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 truly "optimal under noise" in the Alternating Prisoner's Dilemma.

Many thanks to David and Vivian Kraines for their helpful comments and suggestions.

REFERENCES

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

AXELROD, R. (1987). The evolution of strategies in the iterated prisoner's dilemma. In: *Genetic Algorithms and Simulated Annealing* (Davis, D., ed.), London: Pitman.

AXELROD, R. & HAMILTON, W. D. (1981). The evolution of cooperation. *Science* **211,** 1390–1396.

BOYD, R. & LORBERBAUM, J. (1987). No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature* **327,** 58–59.

FREAN, M. R. (1994). The prisoner's dilemma without synchrony. *Proc. Roy. Soc. London B* **257,** 75–79.

HOFBAUER, J. & SIGMUND, K. (1990). Adaptive dynamics and evolutionary stability. *Appl. Math. Lett.* **4,** 75–79.

KRAINES, D. & KRAINES, V. (1989). Pavlov and the prisoner's dilemma. *Theory and Decision* **26,** 47–79.

KRAINES, D. & KRAINES, V. (1993). Learning to cooperate with Pavlov. *Theory and Decision* **35,** 107–150.

KRAINES, D. & KRAINES, V. (2000). Natural selection of memory-one strategies in the iterated prisoner's dilemma. *J. theor. Biol.* **203,** 335–355.

LEIMAR, O. (1997). Repeated games: a state space approach. *J. theor. Biol.* **184,** 471–498.

MAY, R. M. (1987). More evolution of cooperation. *Nature* **327,** 15–17.

MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

MAYNARD SMITH, J. & PRICE, G. R. (1973). The logic of animal conflict. *Nature* **246,** 15–18.

MILINSKI, M. (1987). Tit for tat in sticklebacks and the evolution of cooperation. *Nature* **325,** 434–435.

NOWAK, M. A. (1990). An evolutionarily stable strategy may be inaccessible. *Theoret. Popul. Biol.* **142,** 237–241.

NOWAK, M. A. & SIGMUND, K. (1989). Game dynamical aspects of the prisoner's dilemma. *J. Appl. Math. Comp.* **30,** 191–213.

NOWAK, M. A. & SIGMUND, K. (1990). The evolution of stochastic strategies in the prisoner's dilemma. *Acta Appl. Math.* **20,** 247–265.

NOWAK, M. A. & SIGMUND, K. (1992). Tit for tat in heterogeneous populations. *Nature* **355,** 250–253.

NOWAK, M. A. & SIGMUND, K. (1993). A strategy of win-stay, lose-shift that outperforms tit for tat in the prisoner's dilemma game. *Nature* **364,** 56–58.

NOWAK, M. A. & SIGMUND, K. (1994). The alternating prisoner's dilemma. *J. theor. Biol.* **168,** 219–226.

RASA, O. E. A. (1989). The costs and effectiveness of vigilance behavior in the dwarf mongoose: implications for fitness and optimal group size. *Ethol. Ecol. Evol.* **1,** 265–282.

TRIVERS, R. (1985). *Social Evolution*. Menlo Park: Benjamin Cummings.

WILKINSON, G. S. (1984). Reciprocal food-sharing in the vampire bat. *Nature* **308,** 181–184.

WU, H. & AXELROD, R. (1995). How to cope with noise in the iterated prisoner's dilemma. *J. Conflict Resolution* **39,** 183–189.

Appendix A: Proofs of Propositions 1–4

Proposition 1. *No perfectly optimal strategy exists.*

Proof. Consider the *GRIM* strategy, which cooperates until the opponent defects once, then defects for the rest of the game. No strategy can achieve the maximum possible score against both *GRIM* and *ALLC*: if the strategy ever defects after mutual cooperation, it will perform poorly against *GRIM* ($\sigma_{GRIM}(X) = 0$), and if it never defects after mutual cooperation, it will have a low relative performance against *ALLC* ($\sigma_C = 0$). Thus a strategy cannot be optimal against both strategies, and hence cannot be perfectly optimal.

Lemma 1. *No generous strategy is totally D-unexploitable.*

Proof. We consider an n-ply strategy X with generosity $g > 0$, and show that X has $\sigma_D < 1$. Recall that g is the probability of cooperation after a sequence of n defections. We can compute an upper bound on σ_D by assuming that the strategy will defect against *ALLD* except after a sequence of n defections (any additional cooperation against *ALLD* will lower σ_D). Thus each exploitation by *ALLD* results in $\lfloor n/2 \rfloor$ defections by strategy X to return to the n defections state. Once in the n defections state, strategy X will cooperate with probability g, and thus there will be an average of $1/g - 1$ defections before X cooperates again. Thus X will cooperate once for each $\lfloor n/2 \rfloor + 1/g - 1$ defections, and therefore it will defect against *ALLD* with probability

$$\sigma_D = \frac{\lfloor n/2 \rfloor - 1 + 1/g}{\lfloor n/2 \rfloor + 1/g} = \frac{(\lfloor n/2 \rfloor - 1)g + 1}{\lfloor n/2 \rfloor g + 1}.$$

Thus if $g > 0$, the strategy has $\sigma_D < 1$, and is not totally D-unexploitable.

Proposition 2. *No* 0, 1, *or* 2-*ply strategy is strongly optimal in the APD with no noise.*

Proof. For any 0-ply strategy $S_0 = (a)$, S_0 is equivalent to the 2-ply strategy $(a\ a\ a\ a)$. For any 1-ply strategy $S_1 = (a\ b)$, S_1 is equivalent to the 2-ply strategy $(a\ b\ a\ b)$. Now consider the 2-ply strategy $(a\ b\ c\ d)$ with an initial cooperation probability e. If $b = 1$, *ALLD* can exploit the strategy continually. Similarly, if $d \neq 0$, we know from Lemma 1 that $\sigma_D \leqslant 1/(1 + d)$. Thus for $\sigma_D = 1$, we know $b < 1$ and $d = 0$. But how can a strategy with $b < 1$ and $d = 0$ be totally self-cooperating? Any defection has a non-zero probability $1 - b$ of leading to an unending rut of mutual defection, so we must ensure that no defections occur: $a = 1$ and $e = 1$. But a strategy with $a = 1$ and $e = 1$ cooperates continually against *ALLC*, so no 2-ply strategy with $\sigma_D = \sigma_S = 1$ can have $\sigma_C > 0$.

Proposition 3. *No* 0, 1, *or* 2-*ply strategy is optimal in the infinitesimal noise APD.*

Proof. As in the no-noise case, we consider the 2-ply strategy $(a\ b\ c\ d)$. We will show that the strategy cannot have $\sigma_D = 1$, $\sigma_S > 0$, and $\sigma_C > 0$, and hence is not even weakly optimal. For $\sigma_D = 1$ to hold, the strategy's probability of entering the Dd rut against *ALLD* must be greater than its probability of leaving the Dd rut by $O(\varepsilon)$. Thus we must have $1 - b = O(1)$ and $d = O(\varepsilon)$, which implies $b < 1$ and $d = 0$. For $\sigma_C > 0$ to hold, the strategy's probability of entering the Dc rut against *ALLC* must *not* be less than its probability of leaving the Dc rut by $O(\varepsilon)$. Thus we must have $1 - a = O(1)$ or $c = O(\varepsilon)$, which implies $a < 1$ or $c = 0$. This gives us three possibilities for $\sigma_D = 1$ and $\sigma_C > 0$:

1. $a < 1, b < 1, c > 0, d = 0$. Against its clone, this strategy has probability $O(1)$ of entering a Dd rut and probability $O(\varepsilon)$ of leaving the Dd rut, so $\sigma_S = 0$.
2. $a < 1, b < 1, c = 0, d = 0$. Against its clone, this strategy has probability $O(1)$ of entering a Dd rut and probability $O(\varepsilon^2)$ of leaving the Dd rut, so $\sigma_S = 0$.
3. $a = 1, b < 1, c = 0, d = 0$. Against its clone, this strategy has probability $O(\varepsilon)$ of entering a Dd rut and probability $O(\varepsilon^2)$ of leaving the Dd rut, so

$\sigma_S = 0$. Thus no totally D-unexploitable 2-ply strategy can be both self-cooperating and C-exploiting.

Proposition 4. *No strategy X with finite memory is noise proof and totally D-unexploitable.*

Proof. Assume that strategy X has an n-ply memory for some finite n. Now consider the strategy's generosity g, its probability of cooperating after a sequence of n defections. If $g = 0$, then in a game against its clone, any sequence of n defections will result in a rut of mutual defection. Thus some sequence of errors with length $\Omega \leqslant n$ will result in a continued mutual defection, and hence the strategy is not noise proof. If $g > 0$, we know by Lemma 1 that the strategy is not totally D-unexploitable.