Natural Selection and Social Learning in Prisoner's Dilemma: Co-adaptation with Genetic Algorithms and Artificial Neural Networks*

Michael Macy[1]

Abstract. Evolutionary game theory has been used to study the viability of cooperation in a predatory world. While previous studies have helped to identify robust strategies, little is known about how success translates into the reproduction of cultural rules. Analogs of genetic replication may be deceptive if social learning and natural selection engender different population dynamics. I distinguished selection and learning based on whether rules are hardwired or softwired in the organisms that carry them. I then used genetic algorithms and artificial neural networks to operationalize the distinction. Applied for the first time to iterated Prisoner's Dilemma, neural network experiments showed that researchers may need to be much more cautious in using Darwinian analogs as templates for modeling the evolution of cultural rules.

Keywords. Prisoner’s Dilemma, evolution of cooperation, genetic algorithm, neural network

Natural selection cannot select for behavior per se;
it can only select for mechanisms that produce behavior.

Cosmides and Tooby, 1987

1 Introduction

Evolutionary game theory has attracted the attention of sociologists interested in cooperation among self-interested free-agents (Lopreato 1990; Allison 1992; Kollock 1993; Nielsen 1994; Heckathorn forthcoming. For an overview, see Macy and Flache 1995). Axelrod's (1984) celebrated "computer tournament" led to the proliferation of evolutionary contests among strategies for playing iterated Prisoner's Dilemma. This game has two contestants, each with two choices, to “cooperate” or “defect.” The game is defined by the payoff inequality $T > R > P > S$, where T is the temptation to unilaterally defect, R is the reward when both

* Reprinted with permission of *Sociological Methods and Research.* This research was supported by a grant from the National Science Foundation, #SBR 95-11461.
1 Department of Sociology, Brandeis University, Waltham, MA 02254

cooperate, P is the punishment when both defect, and S is the payoff for suckers. In a single encounter, the best strategy is to defect, since $P > S$ (should the partner defect) and $T > R$ (should the partner cooperate). The dilemma is that the best choice for each means that both receive P, the worst possible collective outcome!

If the game is iterated, however, evolution might favor more enlightened solutions. Axelrod showed how a strategy of “Tit For Tat” (or TFT) might gain a foothold in a computational ecology infested with predators. TFT uses a simple protocol: "Cooperate on the first move and thereafter mimic your partner." The strategy is tough but fair. It never tries to exploit, is quick to forgive those who mend their ways, but never lets betrayals go unpunished. More aggressive strategies do better in a population of naive altruists, but by driving their prey to extinction, they dig their own graves. In contrast, TFT enjoys a decisive evolutionary edge: Unlike predators, it benefits from interacting with copies of itself.

Axelrod's work spawned numerous sociobiological offspring. Most of the subsequent research has focused on the identification of robust strategies (Schuessler 1989; Hirshleifer and Rasmusen 1989; Vanberg and Congleton 1992; Kollock 1993). Little attention has been directed to explaining how “fitness” translates into social influence, or how robust strategies might spread across a population. Analogs of natural selection may be a useful point of departure, but sociobiological models of cultural evolution may be overdue for a critical reconsideration. In evolutionary games, strategies like TFT are “hardwired”[2] in the contestants and can therefore only adapt as their hosts are replaced by organisms programmed with different protocols. Organisms are replaced when they lose out in the competitive struggle to survive and reproduce, hence the central importance of fitness in sociobiological models.

This is a reasonable assumption if rules are genetically inscribed as instinctual responses and selection pressures are tied to competition for survival. However, sociologists are often interested in extending evolutionary models to the spread of cultural rules, such as heuristic routines, conventions, rituals, moral sentiments, habits, and normative codes (Allison 1992; Kollock 1993; Heckathorn forthcoming). Unlike instinctual rules, cultural rules can be “softwired,” that is, they can be modified and passed from one organism to another without the need to replace the host.

If agents are hardwired, adaptation can occur only at the level of the population; an organism can be replaced (through selection at the individual level), but cannot be modified. However, if rules are softwired, adaptation can proceed through reinforcement instead of reproduction. Reproduction alters the

---

[2] Cognitive scientists use “hardwired” to refer to programming that minimizes the response time of an agent. Following Sigmund (1993, p. 177), I use the term to refer to programming that cannot be altered by the agent, with no implication of a faster response time. For example, reciprocity might be hardwired in an organism as a genetically encoded instinctual response, as distinguished from a culturally acquired norm of exchange.

frequency distribution of strategies within a population of actors, while reinforcement alters the probability distribution of strategies within the repertoire of each individual. In simple reinforcement, the organism modifies its behavior in response to positive and negative feedback, as predicted by the "law of effect" (Thorndike 1911). Positive outcomes increase the probability that the associated behavior will be repeated, while negative outcomes reduce it. The process closely parallels the replication of hardwired rules, in which positive outcomes increase the rule's chances for survival and reproduction, while negative outcomes reduce it. However, with reinforcement, rules compete *within* the individuals that carry them, not *between* them.

If rules are softwired, two strategically interdependent individuals can evolve their relationship by learning how best to interact with someone who is trying to do the same, independently of the outcomes experienced by other elements of the population. The unit of co-adaptation is the set of actors directly influencing one another in response to the influence each receives. In a two-person interaction (or game), this unit is the dyad, in a three-person game, the triad, and so on. In contrast, if rules are hardwired, two interdependent players cannot adapt their protocols to one another's behavior. They cannot grow their relationship. In a two-person game, the unit of co-adaptation is not the dyad but the population of rule-carriers competing to reproduce.

Learning can also operate directly on the population distribution. For example, norms can jump from one organism to another by imitation (Dawkins 1976; Durham 1992; Boyd and Richerson 1985; Lopreato 1990). "Social learning" (Bandura 1977) refers to the combination of direct and vicarious reinforcement as the primary elements of human adaptation. Role-modeling can provide an efficient shortcut past the hard lessons of direct experience. If socially connected actors can learn from each other's experience, they can save themselves a good deal of time and grief.

Imitation is the principal rationale for modeling cultural evolution as an analog of natural selection (Boyd and Richerson 1985; Dawkins 1989). However, social influence differs decisively from sociobiological adaptation. Softwired rules can spread without replacement of their carriers, which means that reproductive fitness loses its privileged position as the criteria for replication. While "imitation of the fittest" is a reasonable specification of cultural selection pressures, it is clearly not the only possibility. For example, moral pressure against "free-riding" may be a function of the rate of compliance with the norm and the extent of social ties to deviants, as suggested by Latané's social impact theory (Nowak and Latané 1994) and simulated by Macy (1991b, 1993) and Gould (1993).

To sum up, sociobiology assumes that cooperation is grounded in instinctual rules that evolve in response to selection pressures that condition the chances for survival and reproduction. This hypothesis provided the point of departure for early efforts to model the cultural evolution of prosocial norms. The risk is that the replication of hardwired rules may be a misleading model for cultural evolution.

Social learning theory suggests an alternative model in which norms adapt to social feedback and then spread through imitation of significant others. If social learning converges with the evolutionary outcomes for hardwired actors, this will clearly increase the credibility of sociobiological approaches. If, on the other hand, the outcomes differ in interesting ways, then researchers will need to be much more cautious in using Darwinian analogs as templates for modeling the evolution of cultural rules.

This study tests the convergence of natural selection and social learning models when applied to the most widely studied problem in evolutionary game theory, the viability of cooperation in an iterated Prisoner's Dilemma. I modeled natural selection using genetic algorithms and social learning using artificial neural networks. Computer simulations showed how an emergent self-organizing system of interdependent co-adaptive agents might evolve very differently, depending on whether the protocols are hardwired.

2 The Evolution Of Hardwired Strategies

The genetic algorithm was proposed by Holland (1975) as a problem-solving device, modeled after the recursive system in natural ecologies.[3] The algorithm provides a simple but elegant way to write a computer program that can improve through experience. The program consists of a string of symbols that carry machine instructions. The symbols are often binary digits called “bits” with values of 0 and 1. The string is analogous to a chromosome containing multiple genes. The analog of the gene is a bit or combination of bits that comprises a specific instruction. The values of the bits and bit-combinations are analogous to the alleles of the gene. A one-bit gene has two alleles (0 and 1), a two-bit gene has four alleles (00, 01, 10, and 11), and so on. The number of bits in a gene depends on the instruction. An instruction to cooperate or defect requires only a single bit. However, an instruction to cooperate, defect, or refuse to play requires two bits.

When the gene's instructions are followed, there is some performance evaluation that measures the program's reproductive fitness relative to other programs in a computational ecology. For example, it might be the payoff that is received in a round of Prisoner's Dilemma. Relative fitness determines the probability that each strategy will propagate. Propagation occurs when two mated programs recombine through processes like "crossover" and "inversion."

---

[3] Holland (1992) provides a straightforward explanation of the design and implementation of genetic algorithms. Nonspecialists interested in a delightfully playful introduction may also enjoy Prata’s *Artificial Life Playhouse* (1993) or Sigmund’s *Games of Life* (1993). Bainbridge (1995) is the best sociological introduction to artificial neural networks. For a more technical introduction, see Hinton (1992). Gallant’s (1993) *Neural Network Learning* is a useful text. Bainbridge, Brent, Carley, Heise, Macy, Markovsky, and Skvoretz (1994) surveyed recent sociological applications of genetic algorithms, neural nets, and other models of what Bainbridge has appropriately termed “artificial social intelligence.”

In crossover, the mated programs (or strings) are randomly split and the "left" half of one string is combined with the "right" half of the other, and vice versa, creating two new strings. If two different protocols are each effective, but in different ways, crossover allows them to create an entirely new strategy that may combine the best abilities of each parent, making it superior to either. If so, then the new rule may go on to eventually displace both parent rules in the population of strategies. In addition, the new strings contain random copying errors. These mutations continually refresh the heterogeneity of the population, in the face of selection pressures that tend to reduce it.

To illustrate, consider the eight-bit string **10011010** mated with *11000101*. (The typefaces might represent gender, although the algorithm does not require sexual reproduction.) Each bit could be a specific gene, such as whether to cooperate under eight different conditions. In mating, the two parent strings are randomly broken, say after the third gene. The two offspring would then be **100***00101* and *110***11010**. However, a chance copying error on the last gene might make the second child a mutant, with *110***11011**.

2.1 Prisoner’s Dilemma Experiments

Axelrod (1987) showed that genetic algorithms can be used to look for Prisoner's Dilemma strategies that do not evolve into a competitor that outperforms it. Axelrod used a genetic algorithm with a three-round memory. The chromosome contained 64 genes (four payoffs in each of three rounds, or 4 x 4 x 4), plus six more genes for what to do in the first three rounds. That yields 2^{70} possible strategies. “If a computer had examined these strategies at the rate of 100 per second since the beginning of the universe, less than one percent would have been checked by now” (Axelrod 1987, p. 34). However, by building on partial solutions, a genetic algorithm can chew through this problem in a few hours.

Axelrod used a population fixed at 20 players and allowed only 50 generations. He found that cooperation eventually flourished, “based upon an evolved ability to discriminate between those who will reciprocate cooperation and those who won’t” (p. 38). However, his design used deterministic rules and no background noise. Protocols for reciprocity are known to be vulnerable to “echo effects” triggered by random miscues when interacting with other reciprocators (Axelrod and Dion 1988; Kollock 1993; Sigmund 1993). If either side should mistakenly defect, the result can be an indefinite series of recriminations, as noted by Sigmund. “Even the smallest background noise can come very costly to a Tit For Tat population...Mistakes cause feuds, and feuds do not pay” (1993, p. 192).

In a follow-up to Axelrod's research, Nowak and Sigmund (1993) relaxed the assumption of perfectly predictable behavior. They modeled iterated Prisoner's Dilemma as a Markov process among stochastic strategies consisting of four conditionally cooperative propensities, one for each of the four possible game outcomes. Their model had only one moment of memory and random cooperation on the first move. This model is equivalent to a genetic algorithm with a 4-gene chromosome but without crossover (which is not needed with very short strings) and “gray bits” instead of 0’s and 1’s. Unlike Axelrod, they allowed evolution to unfold for hundreds of thousands of generations.

Based on Axelrod’s experiments, they had expected to see a "generous" variant of Tit For Tat flourish. This rule rewards cooperators by reciprocating but does not always retaliate against defectors. Instead, the clear evolutionary winner was “Win-Stay, Lose-Shift” (WSLS), a pragmatic rule based on the law of effect: repeat rewarded choices, otherwise switch. In Prisoner’s Dilemma, a contestant is assumed to “win” if their partner cooperates and “lose” if their partner defects. TFT thus translates as “win-cooperate, lose-defect,” or alternatively, “win-reward, lose-punish.” TFT and WSLS have identical protocols for cooperation after *R* (a win) and defection after *S* (a loss). However, TFT cooperates after *T* (to reward the partner's cooperation), while WSLS defects (to repeat the rewarded behavior). Conversely, WSLS cooperates after *P* (to avoid further punishment), while TFT retaliates.

Conceptually, TFT and WSLS are complements: TFT teaches, WSLS learns. TFT rewards cooperation and punishes defection, as if trying to modify the partner's behavior. Conversely, WSLS modifies its own behavior by repeating choices that are rewarded and avoiding those that are punished. In effect, TFT promotes adaptation by teaching its partner a lesson, while WSLS contributes by getting the message.[4]

The conditions assumed by Nowak and Sigmund may have unduly favored WSLS over TFT. Nowak and Sigmund allowed only one moment of memory. A second memory register is required to implement more complex strategies that respond not to the partner’s previous choice but to recent changes in the partner’s behavior.

Meanwhile, Axelrod limited evolution to only 50 generations and used deterministic rules that never made mistakes. As noted above, previous researches have shown that TFT is highly vulnerable to blood feuds triggered by random miscues, such as faulty memory, misperception, or Selten’s (1975) “trembling hand” that occasionally pushes the wrong button (Kollock 1993; Bendor 1993). However, this problem can be easily solved by simply mutating a little higher threshold for retaliation. A more subtle threat to cooperation arises only in the *absence* of noise. Consider a world without error in which TFT is flourishing and predators have been driven to extinction. Universal cooperation creates a stable equilibrium where almost everyone cooperates after *R*. As long as their partners are cooperative, forgiving altruists cannot be distinguished from their streetwise and vengeful cousins, allowing the population to drift toward naive cooperation. Eventually, a mutant predator finds itself surrounded by suckers who have lowered their guard. The infection quickly spreads and social order collapses. As Sigmund put it, “The meek shall endanger the earth” (1993, p. 191).

Assuming that life has at least a trace of imperfection, with rare but inevitable miscues and faulty recollections, the problem of lowered immunity

---

[4] Curiously, TFT and WSLS do not get along as well as might be expected. If either side errantly defects, the result is an endless cycle of CD $\rightarrow$ DC $\rightarrow$ DD $\rightarrow$ CD.

may not arise. WSLS can use errors to test whether a cooperative partner can be safely exploited, a trick that TFT cannot use. In short, Axelrod’s failure to incorporate noise may have been unfair to WSLS, while Nowak and Sigmund’s use of minimally complex strategies may have given it an unfair edge.

I therefore conducted an evolutionary experiment very similar to these precursors, but with an allowance for both perceptual and behavioral mistakes (that is, errors on both the input side and the output), and the ability to overlook occasional transgressions. I designed a genetic algorithm (called GENET[5]) that contained 21 genes. Each gene carried a specific instruction to cooperate or defect, for each of 21 conditions to which the device was capable of responding.

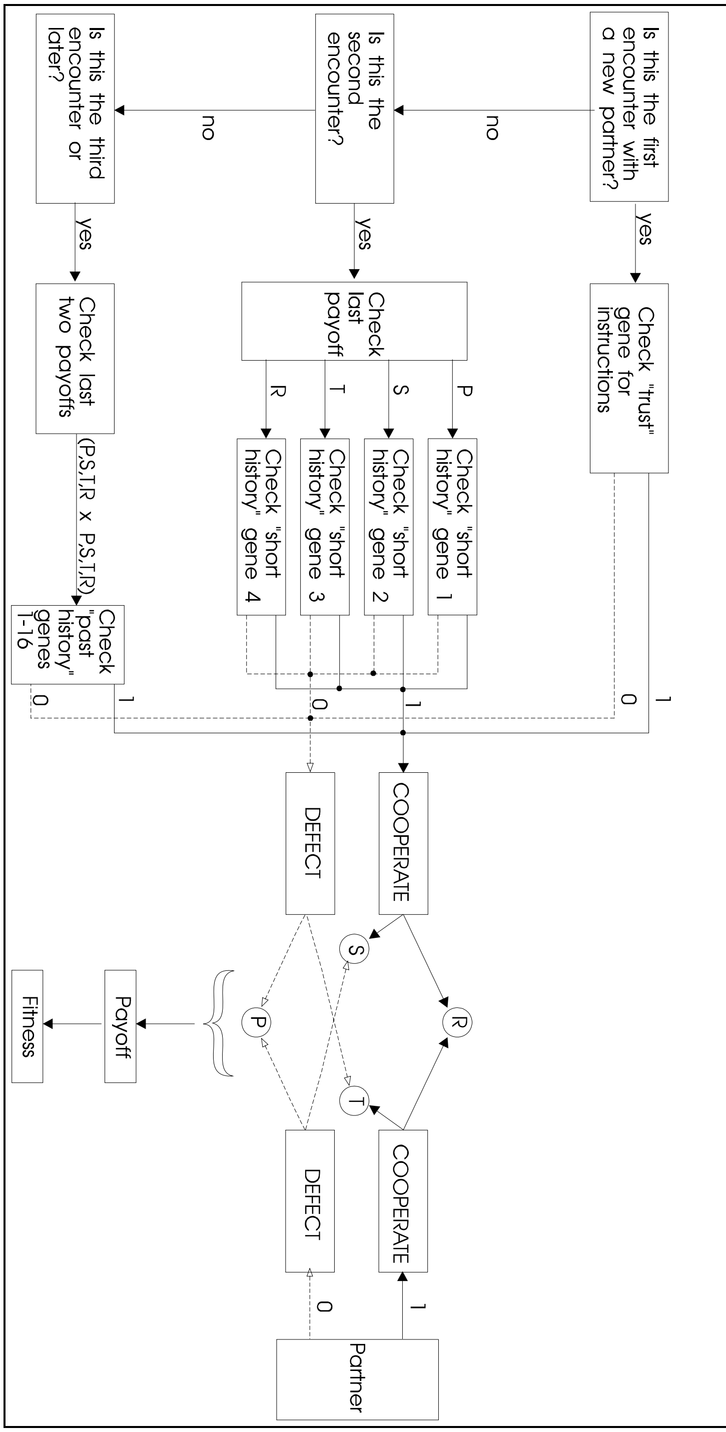
2.2 The Logic of GENET

Figure 1 diagrams the genetic code. The “trust gene” governs cooperation on the first move with a new partner (whose past behavior is unknown). Strategies that always cooperate on the first move are called "nice" (Axelrod 1984). Given the uncertainty with a partner whose past behavior is unknown, a nice strategy is one that is willing to trust a stranger. Trust gets the relationship off on the right foot and may be operationally defined as the propensity to cooperate in the absence of behavioral indicators.

---

[5] All simulation models were programmed in Turbo Pascal 7.0. The source code is available from the author on request.

Figure 1. Schematic of Genetic Algorithm for Iterated Prisoner's Dilemma



On the second move, four “short-history” genes instruct the response to the four possible outcomes of the previous move. After that, sixteen “past-history” genes take over for the remainder of the relationship. The past-history genes respond to the payoffs from the two previous rounds. Two rounds are the minimum needed to see if “Tit For Two Tats” can evolve, a strategy that forgives an intermittent defection. The four payoffs in each of two rounds generate sixteen possibilities, each represented by a different gene. Overall, then, the algorithm assumes a chromosome with 21-genes, a single trust gene, four short-history genes, and 16 past-history genes. The chromosome governs the decision to cooperate, depending on whether there is no history, a history of one move, or a history of two moves.

To illustrate the design, consider the genes for playing Tit For Tat. The genes must instruct the organism to cooperate on the first move and thereafter mimic the partner’s previous move. The trust gene for TFT must therefore contain a “1.” There are four short-history genes, one for each of the four game payoffs in the first encounter with a new partner. A simple way to structure this is to use the decimal equivalents of the binary numbers corresponding to the four combinations of 1’s and 0’s representing cooperation and defection by each side in the previous round. Starting at the first of the four “short history” genes, if both players defected (00), the locator moves forward zero, that is, it remains on the first of the four genes. If the player is suckered (01), the locator moves forward one, to the second gene in the chain. In the opposite case (10), the locator moves forward two. And if both cooperated (11), the locator moves forward three. The protocol for TFT thus requires a “1” in genes 3 and 4 of the short-history string, corresponding to the two outcomes in which the partner cooperated.

The past-history genes work the same way, except here there are sixteen combinations, decimal 0 (or binary 0000) through decimal 15 (or binary 1111). The string must therefore contain 16 genes, each with two alleles, 0 (defect) or 1 (cooperate). For example, if both sides cooperated last time, and neither did the time before, the decimal input value is 3 (or binary 0011), which means “move forward three genes,” to the fourth “past history” gene. If that gene contains a “1,” the player will then cooperate.

Table 1 shows how the sixteen “past history” genes respond to four binary inputs for cooperation (1) and defection (0) by GENET and its partner at times *t*-2 and *t*-1, so as to implement the protocol to cooperate if your partner previously cooperated and defect if your partner previously defected. GENET starts at the first gene in the sequence and skips forward to the location designated by the four inputs. The table shows that genes 3, 4, 7, 8, 11, 12, 15, and 16 must contain a “1” and all others must be “0” in order for GENET to mimic its partner’s previous move.

Overall, then, to play Tit For Tat, the chromosome must contain a 1 in the trust gene, and 1's in all short-history and past-history genes that correspond to all possible events in which the partner previously cooperated. It must also contain 0's in genes corresponding to all possible events in which the partner previously defected. The complete 21-gene chromosome for TFT is then

1,0011,0011001100110011 (the commas are used to designate the starting points for reading each of the three gene sequences).

Table 1: Genes for Reciprocity.

t-2		*t*-1		*t*		
Partner	Self	Partner	Self	Skip Forward	Gene	Allele
0	0	0	0	0	1	0
0	0	0	1	1	2	0
0	1	0	0	4	5	0
0	1	0	1	5	6	0
1	0	0	0	8	9	0
1	0	0	1	9	10	0
1	1	0	0	12	13	0
1	1	0	1	13	14	0
0	0	1	0	2	3	1
0	0	1	1	3	4	1
0	1	1	0	6	7	1
0	1	1	1	7	8	1
1	0	1	0	10	11	1
1	0	1	1	11	12	1
1	1	1	0	14	15	1
1	1	1	1	15	16	1

2.3 Mating Algorithm

The mating algorithm was based on stochastic sampling (Goldberg 1989). At the end of each generation, each individual's probability of mating was a linear function of relative performance during that generation:

$$P_{ij} = \frac{F_i}{\sum_{n=1}^{N} F_n} \qquad \text{for } j = 1 \text{ to } N,\ j \neq i \qquad (1)$$

where P_{ij} is the probability that j is mated with i, F_i is i's "fitness" (or cumulative payoff over all previous rounds in that generation), and N is the size of the population. If the best strategy had only a small performance edge over the worst, it had only a small edge in the race to reproduce. In contrast, Axelrod replaced all below-average strategies with offspring of above-average parents, even if the winners had only a small margin of victory. This may bias evolution against mutants that take small risks, such as those that "test the waters" by cooperating on the first move with a new partner.

With stochastic sampling, each individual, even the least fit, selected a mate from the fitness-weighted pool of eligibles. In each pairing, the two parents combined their chromosomes to create a single offspring that replaced the less-fit parent. The two chromosomes were combined through crossover. First they were split after a randomly chosen gene. The “left” half of one parent’s chromosome was then combined with the “right” half of the other parent’s to create a new chromosome of the same length as the parents’ but with genetic material from each side. The amount contributed by each parent was random from 0 (all from 1 side) to 1 (all from the other). In addition, the strings were copied with an error rate of 0.02 (2 out of every 100 strings were copied with a single genetic mutation).

There were two additional sources of error, input and output. Input error means the partner’s past move was perceived or remembered incorrectly. Output error means the gene’s instructions were carried out incorrectly. However, GENET also had sufficient memory to overlook occasional transgressions, given the low probability of two consecutive errors.

2.4 Simulation Results

I applied GENET to an iterated Prisoner's Dilemma with the bilateral payoffs *R* = 3 for cooperation and *P* = 1 for defection, and the unilateral payoffs *S* = 0 for cooperation and *T* = 5 for defection (the payoffs used by Axelrod and by Nowak and Sigmund). I tested evolution from a random start (randomly programmed 21-bit strings), as well as from a “state of war” (all 21 genes were initialized at 0). The experiment then continued for 10^5 generations.

Each generation consisted of a 50-round game of iterated Prisoner's Dilemma, played with a single partner selected at random from a population of 100. (Multiple partners in a generation needlessly increases the search time.) Following Axelrod, as well as Nowak and Sigmund, I assumed zero population growth. Axelrod used 151 rounds, but 50 rounds were ample to provide conditionally cooperative rules with the necessary "shadow of the future.” Each generation began with a new partner, empty memory registers, and zero cumulative payoffs. Payoffs were then accumulated over the 50 rounds during each generation and represented the ability to compete in the race to reproduce.

Along with the phenotype of cooperation (based on observed behavior), the experiment also measured changes in genotypes for the “trust gene” and for four payoff responses, one for each of the four outcomes at *t*-1. I labeled these “conciliation,” “forgiveness,” “remorse,” and “prudence,” for the rules to cooperate after *P*, *S*, *T*, and *R*, respectively. Conciliation is the genotype for cooperation after repeated mutual defection (gene 1 in Table 1), the willingness to extend an olive branch after both sides have exchanged blows (but if the gesture is not reciprocated, there is no implication of appeasement or surrender). Its antithesis is resolve, the willingness to hold out, to persevere despite the costs. Forgiveness is the genotype for continued cooperation after being suckered; the antithesis is revenge (genes 2, 6, 10, and 14 in Table 1). Remorse is the complement of forgiveness, the genotype for switching to cooperation after making someone a sucker (genes 3, 7, 11, and 15). Finally, prudence is the response to repeated mutual cooperation, the willingness to resist the myopic temptation to score a quick hit (gene 16). Prudence was narrowly defined as a trait that, when echoed, sustains a stable local equilibrium at mutual cooperation. Hence, it was operationalized as the allele for a single gene. The same is true for resolve, which sustains mutual defection when echoed. In contrast, forgiveness and remorse were more broadly defined as traits that are applicable to unilateral outcomes that need not sustain a repeated pattern. Hence, forgiveness was operationalized as the aggregate of all four genes for cooperating after being suckered, and likewise for remorse.

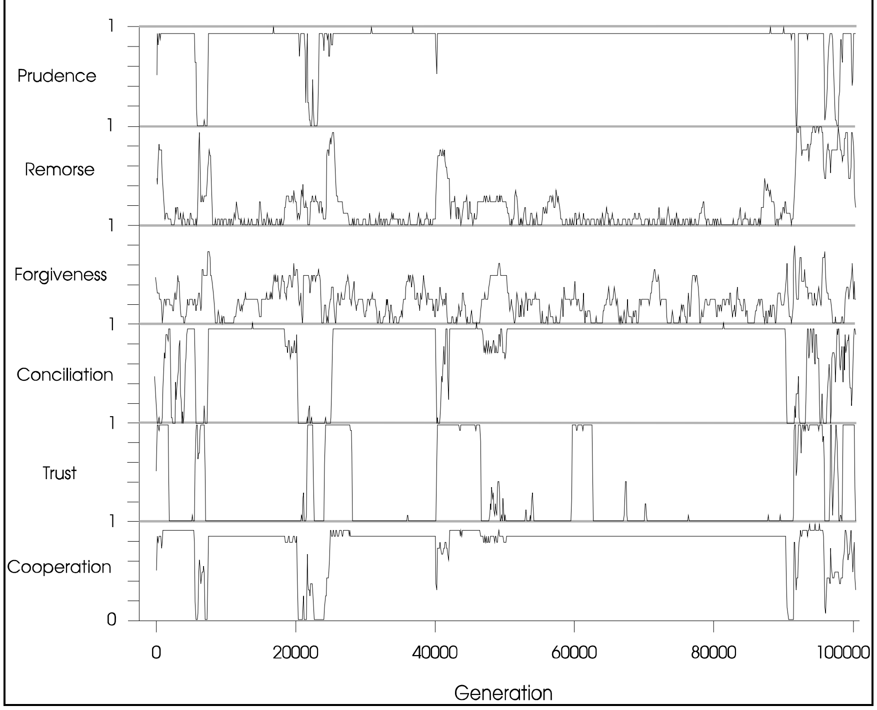
Figure 2 chronicles the evolutionary history of cooperation over 10^5 generations, from a random start, with a two percent mutation rate and a two percent error rate evenly divided between input (perception or memory error) and output (response error). The results confirmed earlier findings by Nowak and Sigmund (1993).[6] A highly cooperative equilibrium was punctuated by very rapid evolutionary transitions and quick recovery of cooperation. The population dynamics resonate with Sigmund's summary of previous research on the evolution of cooperation in iterated Prisoner's Dilemma: "All these computer simulations carry the same, highly uplifting message. Cooperation evolves even in a totally selfish population. Envy and greed appear futile" (1993).

Moreover, as predicted by Nowak and Sigmund, the clear evolutionary winner was WSLS, not TFT. By definition, both TFT and WSLS are prudent but unforgiving. However, TFT is remorseful but resolute, while WSLS is conciliatory but without remorse. Looking at the first few generations following a random start, we see that TFT emerged immediately. Prudence, remorse, and trust spread quickly, while forgiveness and conciliation plummeted. However, TFT's initial success was short-lived. Within 8000 generations, conciliation and distrust of strangers had displaced remorse. From that point on, prudence and conciliation were highly robust and closely correlated with cooperation, while remorse flourished only in brief periods when cooperation had collapsed.

[6] The term "confirm" may seem misplaced when referring to data from computer simulations. Despite their formality, these are nothing more than "thought experiments" that test only the internal validity of a theory and have no bearing on the external validity. Nevertheless, the experiments test hypotheses by manipulating variables under controlled conditions. I use terms like "confirm," "findings," and "results" to refer to the outcomes. Simulations also pose reliability problems. All experiments were repeated ten times to check reliability. Reported figures are representative illustrations.

2

Figure 2. Evolution of Cooperation based on WSLS
(N = 100, e = 0.02)



I also tested whether cooperative rules could invade a population of ruthless predators. The experiment was based on conditions identical to those in Figure 2 except all genes were initialized at zero. Again, WSLS eventually triumphed, but it did not lead the assault. WSLS cannot invade a population of aggressive strategies that prey on its foolish attempts to extend the olive branch. The eventual propagation of conciliation was a consequence, not a cause, of widespread cooperation. WSLS must rely on more resolute colleagues to give it the necessary foothold.

Over repeated simulations, the invasion of the predator population was usually led by a highly defensive reciprocator called Suspicious Tit For Tat, a protocol that rewards but never initiates cooperation. Its resolve (or unwillingness to offer conciliatory gestures) made it phenotypically indistinguishable from predators, allowing its genes for reciprocity to spread as dormant traits until there were sufficient copies to safely emerge from hiding. Eventually, a mutation produced a strain of prudent cooperators that used conciliatory gestures to search for benevolent partners. Paired with ample dormant reciprocators, the new rule (WSLS) easily outperformed the predators and quickly spread. The farther it spread, the better WSLS performed. Its performance suggests that the evolution of cooperation is based not on reciprocal altruism, as previously believed, but on its more adaptive cousin, "pragmatic altruism."

2.5 The Uncertainty Paradox

The success of pragmatic altruism can be attributed to the ability to learn from mistakes. Once established, TFT is vulnerable to becoming trapped in ongoing feuds triggered by random miscues. WSLS, on the other hand, is not only immune to these "echo effects" but actually thrives on occasional errors, especially its own. Blurry vision and a trembling hand paradoxically help pragmatic altruists spot naively forgiving cooperators. This helps weed out strategies that invite an invasion by more aggressive rules. Along with mistakes, WSLS also uses distrust of strangers to test whether a cooperative partner can be safely exploited, as reflected in the lack of trust evident in Figure 2.

Experiments that manipulated the probability of miscues confirmed the effect of uncertainty on the viability of WSLS. Table 2 reports the effect of uncertainty on the level, stability, and genetic basis of aggregate cooperation. The level was measured as the grand mean, based on the proportion of the population choosing to cooperate during each generation, averaged over 10^5 generations. Stability was based on deviation of generational population proportions from the grand mean, that is, as variance over time, not across individuals. Genotype means and variance were similarly computed, except that the proportions were based on the number of genes for each trait that contained a "1." Correlations were based on covariance between population means for cooperation and each of the genotypes, without time lag.

The experiment tested four error rates, ranging from zero to six percent. Conditions were otherwise identical to those in Figure 2. The main finding was that cooperation became increasingly stable as uncertainty increased. In a world without error, cooperation was much less stable and WSLS failed to secure a firm grip on the ecology.

This is a remarkable and counter-intuitive result. Adding a little noise at the individual level dramatically reduced uncertainty at the ecological level! With flawless information, a population of perfectly predictable individuals cooperated at about the rate that would be expected by random behavior (0.513), with random fluctuations between punctuated equilibria. However, as the behavior of the individuals became less certain, the behavior of the population became increasingly predictable, as indicated by the dramatic decline in the standard deviation, from 0.4 to 0.088.. With six percent error, WSLS quickly seized control and never lost its grip on the population. Aggregate cooperation barely wavered.

Table 2. Effect of Uncertainty on Level, Stability, and Sources of Cooperation.

	Mean[a]				Standard Deviation			
By Error Rate	0%	2%	4%	6%	0%	2%	4%	6%
Cooperation	.513	.807	.775	.678	.400	.215	.129	.088
Trust	.335	.246	.153	.090	.413	.415	.348	.259
Conciliation	.253	.783	.848	.909	.313	.316	.265	.146
Forgiveness	.363	.218	.118	.039	.193	.166	.130	.084
Remorse	.406	.179	.081	.040	.238	.256	.180	.128
Prudence	.621	.892	.928	.906	.399	.190	.101	.172

	Correlation with Cooperation			
By Error Rate	0%	2%	4%	6%
Trust	0.570*	-0.035	-0.264*	0.031
Conciliation	0.411*	0.695*	0.738*	0.788*
Forgiveness	-0.189*	-0.154*	-0.187*	-0.219*
Remorse	0.393*	-0.195*	-0.201*	-0.552*
Prudence	0.584*	0.609*	0.618*	0.789*

* Significant at the 0.01 level.
[a] All differences in means were significant at the 0.01 level.

Unfortunately, ecological stability was purchased at the expense of cooperation, as miscues took an ever greater toll on population welfare. Each accidental defection by WSLS triggers at least one retaliation before the echo dies out. Two percent uncertainty turned out to be close to the optimal tradeoff. Below two percent, WSLS could not get established, and cooperation suffered accordingly. Above two percent, WSLS approached evolutionary stability. By six percent, forgiveness and remorse were driven close to extinction (about 0.04), crowded out by conciliation and prudence (each over 0.9). However, the level of cooperation had declined, from 0.807 to 0.678.

The correlations revealed a similar pattern. As uncertainty increased, trust and remorse became less useful or even counter-productive, while conciliation and prudence became more closely associated with cooperation. In short, WSLS kept getting stronger as uncertainty increased, even though the effect on cooperation was non-monotonic.

To conclude, these experiments confirmed the evolutionary robustness of WSLS reported by Nowak and Sigmund. Reciprocators appeared to play a decisive role in the evolution of cooperation, securing a safe beachhead for more pragmatic altruists. In the long run, however, learning prevailed over teaching. Pragmatists rode reciprocators to power and then dispensed with them.

There is a strong temptation to draw out the moral implications of the demise of remorse and the triumph of conciliation. Beginning with Axelrod's evolutionary tournament over a decade ago, social scientists have used analogs of natural selection to explain the evolution of moral rules and cultural norms.

For example, Darwinian logic informed Allison’s (1992) model of the evolution of “beneficent norms” and Frank’s (1988) explanation for “moral sentiments” grounded in emotion. More pointedly, the robustness of TFT (and its close cousins) has been used to explain diverse instances of reciprocity in social exchange, from tacit collusion in the Great War (Axelrod 1984) to loose reciprocity among the Inuit (Kollock 1993). The clear dominance of WSLS does not disprove the optimistic moral lessons previous theorists have drawn from TFT’s success. Nice guys still finish first, they just turn out to be more conciliatory and less remorseful than previously believed.

However, the purpose of these experiments was not to challenge reciprocity as the basis of cooperation in social exchange. Quite the opposite. My chief concern is the tendency for social scientists to draw cultural lessons from the evolutionary robustness of any hardwired rule, TFT or WSLS. The problem is that cultural rules are softwired, procured through learning, not lineage. Cosmides and Tooby (1987, pp. 278-9) have warned researchers not "to apply evolutionary theory directly to the level of manifest behavior,” but to use it “as a heuristic guide for the discovery of innate psychological mechanisms.” Human biological adaptation has produced social organisms, not with an instinct to cooperate, but with an ability to *learn*. Once that faculty is flourishing, adaptation can proceed in an entirely new way. Hence, we need to study how cooperation might spread among organisms capable of rewiring their rules on the fly.

3 The Co-Adaptation Of Softwired Strategies

As a genetically hardwired strategy, WSLS can change its behavior, based on associated outcomes, but it cannot change the rule that tells it to choose pragmatically. Simply put, while WSLS is guided by trial-and-error to cooperate or defect, it cannot learn to abandon a trial-and-error strategy in favor of reciprocity or some other protocol. The rule can only change when its carrier is crowded out by neighbors with better genes.

That is not the case, however, in the cultural evolution of cooperation. Cultural evolution can be modeled as an adaptive process among self-programmable devices capable of learning rules through interaction. This process differs from sociobiological evolution in that the rules are softwired, that is, the rules can be modified without replacing their carriers. If the logic of softwired adaptation differs in important ways, then previous studies, based on analogs of natural selection, should be restricted to the problem of instinctual cooperation and not generalized to the emergence of normative solidarity.

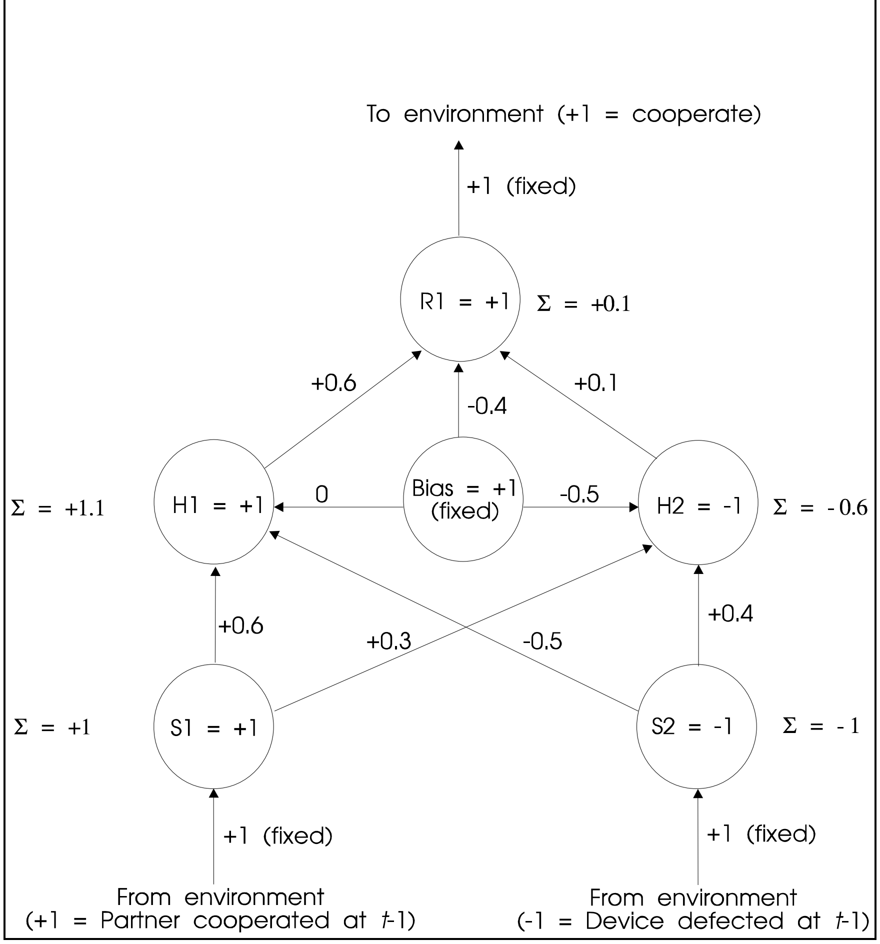
An artificial neural pathway is a simple type of self-programmable learning device based on parallel distributed processing (Rummelhart and McClelland 1988). Like genetic algorithms, neural nets have a biological analog, in this case, the nerve systems of living organisms. In elementary form, the device consists of a web of neuron-like units (or neurodes) that fire when triggered by impulses of sufficient strength, and in turn stimulate other units when fired. The magnitude of an impulse depends on the strength of the connection (or "synapses") between the two neurodes. The network learns by modifying these path coefficients in response to environmental feedback about its performance.

Neurodes consist of four types of units, as illustrated in Figure 3. The most straightforward are sensory and response units. Sensory units are triggered by input from the environment, such as the payoff from the previous round of an iterated Prisoner's Dilemma. In Figure 3, there are two sensory units. S1 has been activated by the environment (+1), while S2 has not (-1). For example, this might correspond to a condition in which the partner had cooperated on the previous play while the neural net had not. Response units, in turn, trigger action by the organism on the environment, such as the decision to cooperate or defect. In Figure 3, response unit R1 has been triggered, as indicated by the output value of +1.

The other two neurodes are less intuitive. Intermediate (or "hidden") units link sensory and response units so as to increase the combinations of multiple stimuli that can be differentiated. Figure 3 shows two hidden units, H1 and H2. Unlike sensory and response units, hidden units have no direct contact with the environment, hence their name. Finally, "bias" units are similar to hidden units but they have no inputs. Instead, they continuously fire, creating a predisposition toward excitation or inhibition in the units they stimulate, depending on the size and sign of the path coefficients. If the path weight is positive, the bias is toward excitation, and if the path is negative, the bias is toward inhibition.

A neurode fires when the sum of its weighted inputs exceeds zero. For example, in Figure 3, R1 fires because its total input is 0.1. Similarly, S1 fires because it is excited by a stimulus from the environment. When it fires, the neurode's output changes from -1 to 1, which means that its effect on later neurodes in the chain reverses. If the neurode was an inhibitor, it now becomes an excitor, and vice versa. In Figure 3, S1 excites both H1 and H2. The pathways that link the neurodes are weighted with values that determine the strength of the signals moving along the path. A low absolute value means that an input has little influence on the output. A large positive weight makes the input operate as an excitor. When it fires, the input excites an otherwise inhibited output. A large negative path weight makes the input operate as an inhibitor. When it fires, the input inhibits an otherwise excited output.

Figure 3. Schematic of Artificial Neural Network



The neural device in Figure 3 has learned to play TFT. The device cooperates if S1 = 1 (partner cooperated last time) and S2 = -l (device defected), as illustrated. The inputs to H1 are then 0.6*1, -0.5*-1, and 0, for a total of 1.1. Since 1.1>0, H1 fires, sending a value of 0.6 to R1. The inputs to H2 are 0.3 (from S1), -0.4 (from S2), and -0.5 (from bias), for a total of -0.6. Hence, H2 does not fire, sending a value of -0.1 to R1. R1 receives 0.6 (from H1), -0.1 (from H2), and -0.4 (from bias), for a total of 0.1, causing R1 to fire and the device to cooperate.

The device will also cooperate if S1 = S2 = 1. Now both H1 (total of 0.1) and H2 (total of 0.2) fire, causing R1 to fire (total of 0.3). However, the device will defect if S1 = S2 = -1 (both defect), or if S1 = -1 and S2 = 1 (the device is suckered).

The network learns by modifying the path weights linking its neurodes. Learning only occurs when the response to a given sensory input pattern is unsatisfactory. The paths are then adjusted so as to reduce the probability of repeating the mistake the next time this pattern is encountered.[7]

In many applications, neural nets are trained to recognize certain patterns or combinations of inputs. For example, suppose we want to train a neural net to predict stock prices from a set of market indicators. We first train the net to correctly "predict" known prices. The net begins with random path coefficients. These generate a prediction (or "wild guess"). The error is then used to adjust the coefficients in an iterative process that improves the predictions. These coefficients are analogous to those in a linear regression, and like a regression, the weights can then be applied to new data to predict the unknown. An important advantage of the net over least squares estimation is that it requires fewer assumptions about the functional form of the solution. On the other hand, this can also cause the net to over-fit the data. Hence, while neural nets can be very sharp prognosticators of the future, they can also generate badly convoluted interpretations of the past (Bainbridge et al. 1994, p. 428).

3.1 ANNET: A Prisoner's Dilemma Application

The Prisoner's Dilemma application in this study is a self-organizing community of co-adaptive devices that train one another on the fly, in response to the training they receive. Each individual consists of an artificial neural network, called ANNET, that is an elaboration of the Bush-Mosteller stochastic learning model I have used in several previous social dilemma simulation experiments (1990, 1991a, 1991b). The Bush-Mosteller model is equivalent to a neural net with only a single bias unit and an output, but with no sensory inputs or hidden units. Such a device is capable of learning *how often* to cooperate, but not *when* to cooperate, that is, it is incapable of learning conditional strategies. In contrast, ANNET can learn to differentiate environmental cues and respond using more sophisticated protocols for contingent cooperation.

ANNET's predecessor is Bainbridge's MIND program (1995b), a neural network used to study the dynamics of social interaction.[8] It is a very simple device by the standards of the 1990's. ANNET has only 10 neurodes. The five sensory inputs are identical to the inputs to GENET (new partner, partner

---

[7] In principle, if a response unit fires at the wrong time, the weights that excited the unit would be reduced and the weights that inhibited the unit would be increased. In actuality, things are a bit more complicated; the effect of a weight also depends on the state of the neurode behind it, and so on back up the chain.

[8] Like Bainbridge's MIND (for "Minimally Intelligent Neural Device"), ANNET has only one layer of hidden units and does not require back propagation. MIND only modifies the path weights between input and hidden units. All other weights (between bias, hidden, and output units) are fixed. The key difference with MIND is that ANNET also modifies these weights, but only one set at a time, alternating between the sensory input side of the hidden layer and the response side (see note 7).

cooperated, ANNET cooperated, at times *t*-1 and *t*-2). When the partner is new, the first input value is 1 and the other four inputs are 0 (which means the input values are not known). If the partner has not changed since the previous interaction, the first input value is -1 and the other inputs are 1 for cooperation, -1 for defection, and 0 if unknown. (The *t*-2 inputs are still not known on the second interaction with a new partner.) There are also three hidden units, one bias, and one response unit.

ANNET functions the same way as the simple two-input device illustrated in Figure 3. All path coefficients were initially random and were then adjusted only if outcomes were unsatisfactory. The simulations assumed that the object of the game was to get one's partner to cooperate. Hence, an outcome was unsatisfactory if the partner defected. I implemented this assumption by subtracting a constant from the payoff matrix used with GENET, such that $T > R > 0 > P > S$. When the outcome was unsatisfactory, the paths were modified so as to reduce the probability of repeating the mistake. The correction was simply the product of the input, the output, and the negative payoff (either *S* or *P*). If an input and output both fired (+1) and the outcome was unsatisfactory, the correction was negatively signed and the path weight was reduced, thereby decreasing the excitation of the output. If the output fired but not the input, the weight was increased, so as to inhibit the output. If the input fired but not the output, the weight increased, so as to excite the output. And if neither fired, the weight was reduced (which decreased the inhibition). The more aversive the outcome (given that $S < P$), the stronger the correction. Finally, the adjustment included an error term that allowed for idiosyncrasy, and a constant that adjusted the learning rate. More formally,

$$W_{oi} = W_{oi} + s_i s_o \pi l e \qquad (2)$$

where W_{oi} is the coefficient for the path from *i* into *o*, s_i is the input stimulus to the path ($s_i = \pm 1$), s_o is the output stimulus ($s_o = \pm 1$), π is the payoff ($\pi \in$ [S,P], where $S < P < 0$), *l* is the learning rate ($0 < l < 1$), and *e* is an error term with a uniform distribution ($0 < e < 1$). ANNET's self-correcting logic keeps W_{oi} from wandering more than a few payoffs from zero.

3.2 The Unit Of Co-Adaptation And The Destabilizing Effects Of Uncertainty

Iterated Prisoner's Dilemma involves an ongoing relationship between two players. The dynamics of the relationship depend on whether their rules are softwired. Two hardwired interactants can influence each other's behavior, but they cannot change the rules that govern their behavior, in response to one another. The rules only change when their carriers are replaced, in response to competition not only from one another but from other rules in the ecology.

In contrast, two softwired interactants continually influence each other's rules. Here, the unit of co-adaptation is neither the rule nor the ecology but the dyad. In effect, two interacting neural nets become a single entity (the dyad) that searches through combinations of path coefficients until it finds a pair of strategies whose interaction yields two satisfactory outcomes (*R*). Although *T* is also satisfactory, it is paired with *S*, which is not, so the dyadic net keeps looking. Once *R* is found, learning stops and the mutually cooperative responses are repeated until a mistake is made or the relationship is terminated at the end of the game. The greater the uncertainty, the more mistakes and the more prolonged the search.

The easiest strategy pair for locking in cooperation is mutual prudence, the protocol to cooperate after *R*. Prudence can also be combined with mutual reconciliation, which induces self-sustaining joint cooperation after both sides have defected. With prudent reconciliation, non-remorse and vengefulness contribute to cooperation indirectly by inducing mutual defection, as a stepping stone to joint cooperation. These traits reduce to the protocol "defect after *T* and *S*; cooperate after *P* and *R*," or WSLS, the pragmatic strategy to keep doing what seems to work. The sequence moves from unilateral to bilateral defection and from there to a stable equilibrium at bilateral cooperation.

Other combinations are also possible. If one player is prudent, remorseful, and resolute (a strategy based on reciprocity), then the other needs to be prudent, forgiving, and conciliatory (a strategy that is naively cooperative). Now the sequence moves from bilateral to unilateral defection and then to bilateral cooperation. This shows that ANNET need not settle upon pragmatic altruism. ANNET is hardwired to change *rules* if the outcome is unsatisfactory, but because ANNET learns rules, not choices, ANNET can select a rule-pair based on reciprocity and naiveté rather than pragmatism.

Indeed, there is no guarantee that ANNET will find any of the possible cooperative solutions. The dyad must search through 21^2 possible rule-pairs, compared to only 2^2 choice-pairs when WSLS is hardwired. The coordination problem faced by the dyad may require more iterations than are available to the relationship. This suggests that in relationships of limited duration, the viability of cooperation will decrease with strategic complexity.

The search is also complicated by the possibility of mistakes. When WSLS is hardwired, it is trivially easy for a cooperative dyad to recover from occasional errors. The sequence is simply CC $\rightarrow$ DC $\rightarrow$ DD $\rightarrow$ CC. However, the search space for a stable rule pair is vastly larger, and this badly compromises recovery. Hence, the hypothesized effect of uncertainty is precisely the opposite of the effect observed with hardwired rules. With softwiring, rather than stabilizing the domination of WSLS at the ecological level, uncertainty may instead destabilize mutual cooperation at the dyadic level.

3.3 Simulation Results

The first experiment tested the hypothesized effects of uncertainty. Figure 4 reports ANNET's ability to find a cooperative solution to iterated Prisoner's Dilemma, from a random start, with and without the possibility of miscues. The graph traces the mean rate of cooperation at each iteration duration 100 games, each lasting 50 rounds, for a total of 5000 iterations. Learning occurred only when outcomes were unsatisfactory, that is, when the partner defected, with a learning rate[9] of 0.2. Otherwise, the experiment assumed conditions similar to those for GENET. Each game lasted for 50 rounds with a partner randomly chosen from a population of 100 contestants. The level of uncertainty was either 0 or 0.02, evenly divided between input and output errors (memory/perception and response, respectively).

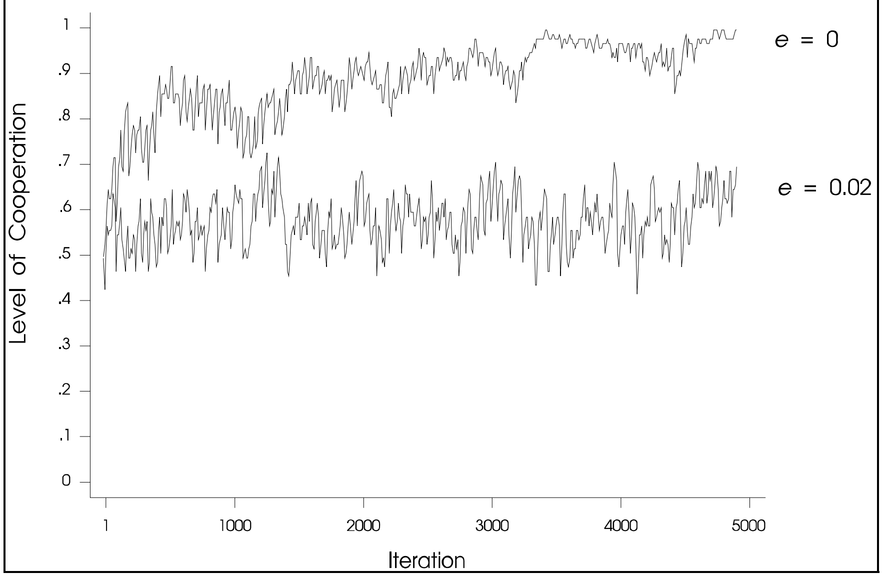
Figure 4 shows that the effect of uncertainty on the evolution of cooperation was the opposite of the effect observed for hardwired rules: The equilibrium rate of cooperation was depressed by uncertainty. From a random start and no miscues, the population eventually attained near-universal cooperation, and averaged 0.89 over all 5000 iterations. In contrast, with an error rate of two percent, mean cooperation fell to 0.58. The change is the mirror image of what happened with hardwired rules: With no uncertainty, cooperation averaged about 0.51 but climbed to near-universal cooperation (and 0.81 overall) when the error rate was two percent (as indicated in Table 2).

This finding shows the danger in generalizing evolutionary logic from natural to social ecologies. If we assume uncertainty, natural selection badly overstates the prospects for the cultural evolution of cooperation. If we assume no uncertainty, natural selection badly understates the prospects. Either way, models of cultural evolution based on analogs of natural selection are likely to be misleading when applied to individuals capable of learning to cooperate.

---

[9] This rate means that one-fifth of the payoff was applied to the alteration of the path coefficient The learning rate was optimized by testing ANNET’s ability to recognize all possible combinations of sensory inputs. Like a fox chasing a clever rabbit, ANNET must balance raw speed against the risk of over-correction.

Figure 4. Effect of Uncertainty on Cooperation
(N = 100, 1 game = 50i)



Comparison with Figure 2 reveals other differences between social learning and competitive selection. In Figure 4, there was no punctuated equilibrium. The disturbances were smaller and more frequent. Figure 4 also shows the main reason for disturbances: the level of trust lagged behind the level of cooperation. Note the cyclical drop in cooperation that occurred with the same period as each new partner. Each relationship started out with ambivalence about the new partner that was then overcome with experience. In the absence of mistakes, the population could eventually learn to trust strangers, but the process was clearly much more difficult than learning to trust a familiar face. Nevertheless, the overall level of trust was much higher than observed in Figure 2.

Inspection of the emergent rules at cooperative equilibrium reveals another difference with hardwired adaptation. At the end of the experiment, I forced latent rules to surface by inputting all possible combinations of stimulus events to each device and observing the response. Paradoxically, in a population of organisms that adapt through learning, WSLS did not emerge as the most viable strategy. Prudence and forgiveness were most prevalent, followed by remorse and conciliation.

3.4 Strategic Complexity

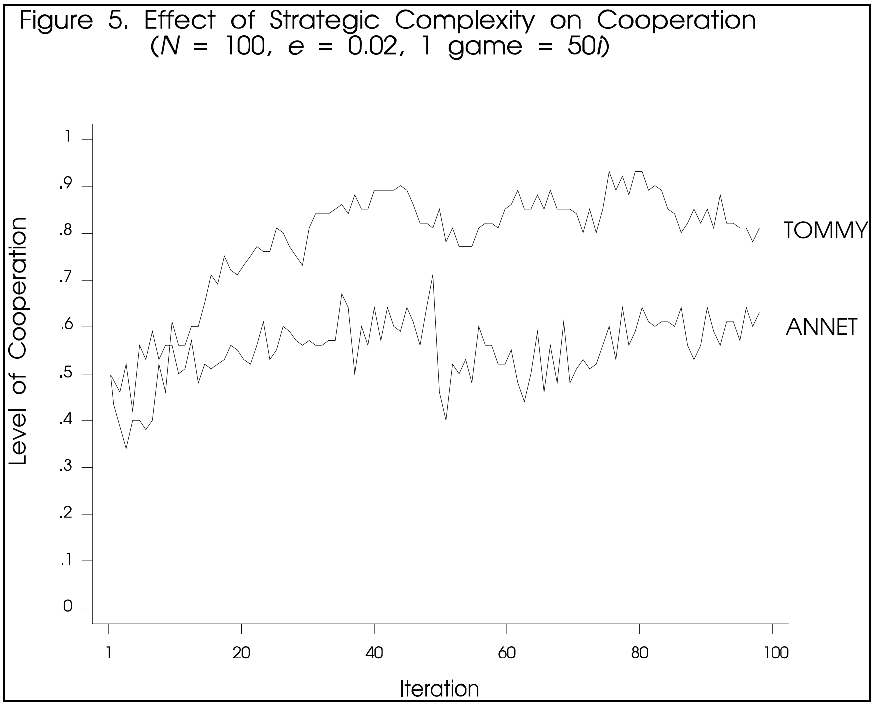
When two neural nets are paired and become strategically interdependent, they merge to form a co-adaptive entity that searches for mutual cooperation. With any other outcome, path weights on one or both sides of the net will be modified. Holding constant the level of uncertainty, the search time will depend on the number of possible rule combinations which in turn depends on the complexity of the rules.

I tested the effect of strategic complexity by comparing ANNET's performance against TOMMY, a neural net named after a mythical pinball wizard who relied on an intuitive feel for the game and consistently outperformed other players. TOMMY was identical to ANNET except that all its sensory units were blocked (set to 0). In effect, the neural device was denied sensory input from the environment and could therefore not use conditional strategies. Instead, TOMMY had to rely entirely on "intuition," defined as internally generated cues and operationalized as the state of its hidden and bias units. The payoffs could tell TOMMY how often to cooperate, but not when. This made the device functionally comparable to the Bush-Mosteller learning model that I have used in a series of computer simulation experiments on cooperation in social dilemmas.

One advantage of ignoring environmental cues is that one need not worry about making perceptual or memory errors, although response errors are still possible. I therefore gave TOMMY the same handicap as ANNET on the output side (a one percent error rate), but no input error. Figure 5 compares the performance of the two models over 100 iterations, which was all that TOMMY needed. The results are strikingly clear: Strategic complexity clearly compromised the search for a solution to Prisoner's Dilemma. Simply put, sensory deprivation prevented TOMMY from looking for social cues for when it was safe to cooperate. But when everyone overlooked the cues, the danger evaporated. TOMMY's success rested, quite literally, on blind faith.[10]

---

[10]Kollock's (1993) study of reciprocity in iterated Prisoner's Dilemma quoted Ghandi's observation that "An eye for an eye leaves everyone blind." It now appears that blindness in Prisoner's Dilemma may not necessarily be a liability.



3.5 Imitation of the Fittest

According to social learning theory (Bandura 1977), the efficiency of adaptation can be enhanced through mimetic learning, or role modeling. “Imitation of the fittest” is the cultural analog of competitive selection pressures in natural ecologies. Imitating a more successful partner parallels the "mixing" of one's own genes with those of a fitter mate in the sociobiological specification. However, there is a decisive difference. Imitation operates only on rules that are behaviorally manifest. Hence, rules cannot spread unless they are directly observed. This precludes evolutionary drift--the spread of latent programs that would be snuffed out if they were immediately tested.

This limitation can be avoided if cultural transmission operates through instruction rather than imitation. An instructor can reveal protocols that apply to unobserved conditions. Instruction can provide access to latent rules but may also be distorted, by intentional deception or cognitive error (analogous to genetic mutation).

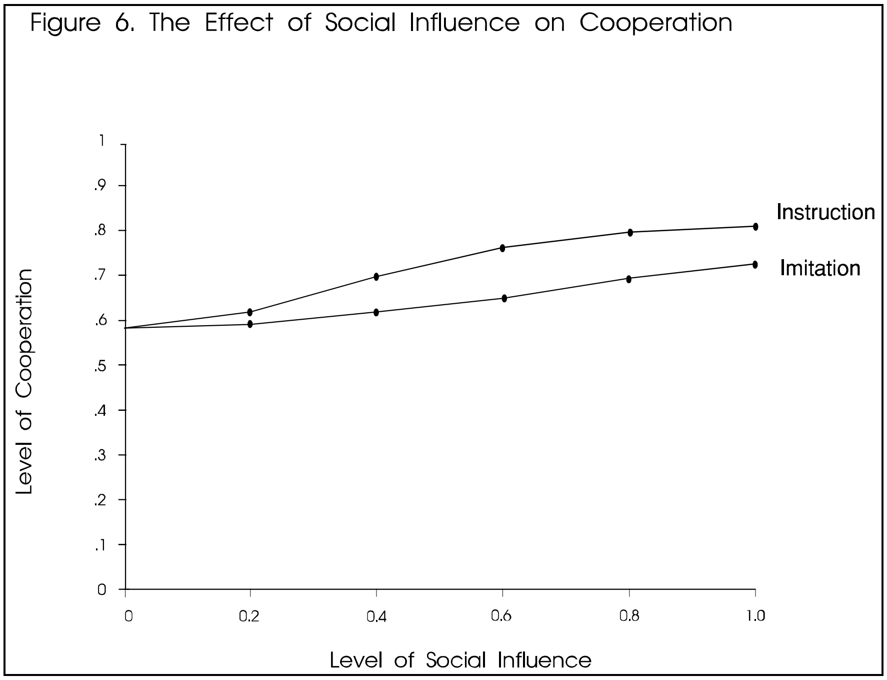
Both imitation and instruction differ from natural selection in two other ways. First, neither imitation nor instruction is necessarily triggered by a performance evaluation. The mimic may choose to copy those who are nice, nasty, widely imitated, or simply willing to share their secrets. Second, social influence does not involve replacement of the carrier of the displaced rule. This means that influential rules can spread without disrupting existing relationships.

I tested the effects of social influence by equipping ANNET with the capacity for vicarious learning. For comparability with GENET, I limited the selection criteria to fitness, or cumulative payoffs. However, without generational population turnover, fitness could accumulate across multiple games. To keep the rate of change in relative fitness constant over the entire simulation, I used a moving average over the last 100 payoffs, the minimum needed for success to influence the upcoming partner. (Fitness was randomly distributed at the starting line.) If a partner differed in behavior and was more successful, ANNET imitated by adjusting the weight on the path from the bias unit to the response output. If the role model cooperated, the weight increased, which then biased the response toward firing (cooperating). If the role model defected, the weight decreased. The learning rate was identical to that for direct reinforcement. The lower the rate of influence, the more repetitions that were required to alter behavior.

With instruction, if a partner was observed to be more successful, ANNET incorporated the partner’s path matrix as a weighted average. The learning rate represented the relative weight of the two matrices and varied from 0 (no influence) to 1 (complete replacement of the weight matrix by that of the partner). An important difference with imitation was that the path coefficients included much more information than was behaviorally manifest, insofar as many combinations of stimuli may never occur. However, the inability to verify the information entails a high probability of copying error. I therefore assumed a ten percent rate of mutation, much higher than GENET’s two percent rate.

Figure 6 shows the effect of social influence on mean cooperation over 5000 iterations, under two conditions, imitation and instruction. The level of influence ranged from zero to 1 in five steps. At 0.2, the rate of imitation was equal to the learning rate for direct reinforcement. At maximum influence, imitation had five times the impact of direct experience. For instruction, influence was measured as the weight of the mentor’s path coefficients relative to the native values. At 0.5, the two rules were equally weighted. All other conditions were identical to those in Figure 4, with an error rate of 0.02, evenly divided between input and output errors.

The results show that social influence substantially increased the rate of cooperation. For instruction, the proportion of cooperative responses increased from 0.58 at no influence to 0.81 with a fully influential mentor. The effect of imitation was somewhat less pronounced, boosting cooperation to 0.72.



4 Conclusion

These simulation experiments tested the convergence of the sociobiological and social learning models when applied to the most widely studied problem in evolutionary game theory, the viability of cooperation in an iterated Prisoner's Dilemma. Natural selection and social learning can be differentiated based on whether rules are hardwired or softwired in the organisms that carry them. If rules are hardwired, adaptation can occur only at the ecological level; the individual organisms can be replaced but they cannot be modified. However, if rules are softwired, adaptation can proceed through reinforcement instead of reproduction. Reproduction alters the frequency distribution of strategies within a population of actors, while reinforcement alters the probability distribution of strategies within the repertoire of each individual.

I operationalized this distinction by modeling natural selection with a genetic algorithm and social learning with an artificial neural network. The models were applied to the problem of cooperation in an iterated Prisoner's Dilemma game in which the contestants knew the previous two moves and whether their partner was a stranger. Computer simulations showed how an emergent, self-organizing system of interdependent, co-adaptive agents might evolve very differently, depending on whether the rules were hardwired.

Findings based on a genetic algorithm corroborated the evolutionary viability of WSLS, a strategy of pragmatic altruism based on prudence and conciliation. The simulations also confirmed the existence of punctuated equilibria, in which long periods of mutual cooperation were interrupted by swift and devastating collapses. However, recovery from collapse was usually swift.

Punctuated equilibria were not apparent with softwired rules because the unit of co-adaptation was the dyad, not the ecology. Dyads couple two neural nets into a single adaptive entity that searches for mutual cooperation. Any other outcome entails continued search. In contrast, in a genetically hardwired population, the adaptive entity is the ecology, searching for evolutionary stability (a distribution of strategies that cannot be invaded). Any other outcome entails further evolution. Mutual cooperation at the dyadic level is much easier to locate than is evolutionary stability at the ecological level. Moreover, the dispersal of the search process precludes sudden system-wide shifts. Cooperation is thus much more stable when the unit of co-adaptation is isomorphic with the unit of strategic interdependence.

In social learning, complex rules for contingent cooperation increase the time required to coordinate a pair of co-adapting strategies. The effect of strategic complexity on learning contrasts markedly with the effect on natural selection. Complex hardwired rules create spaces where latent cooperative genes can be protected from selection pressures until the rules are sufficiently widespread that they can safely emerge.

The effect of uncertainty was also very different for social learning and natural selection. Mistakes badly complicated the dyadic coordination problem but helped hardwired pragmatists spot naive cooperators before they could lure in more aggressive rules.

As expected, social influence greatly improved the prospects for cooperation. Social learning allows new rules to spread through the population, through imitation and instruction. Influence provides an effective antidote to the effects of uncertainty by increasing the probability of bilateral choices relative to unilateral. This simplifies the coordination problem. The risk is that bilateral defectors will then trap one another in a hall of mirrors, with no escape. However, the costs of mutual defection bias imitation toward coordination of a mutually gratifying solution.

Further research is needed to test the generality of these results. These simulations assumed small, stable, and homogenous populations. The Prisoner's Dilemma game can also be extended to multilateral social dilemmas like "tragedy of the commons." Although this study did not include the formation of commitments, ANNET can be complicated with a second decision, whether to remain with the current partner or exit. A game with an option to exit would also challenge contestants to learn the benefits of commitment, and how to find a suitable mate. Future research should explore how co-adaptive agents might learn to build long-term relationships. (For a preliminary study, see Macy, Skvoretz, and Bainbridge 1995.)

Researchers might also want to test the effects of network structures as conduits for the spread of influence in social learning. Influence processes and cascade effects may be particularly important for understanding the viability of cooperation in multilateral games where synchronized cooperation is exponentially more difficult to attain.

In conclusion, the specific findings reported here should not be taken as definitive but as invitations to pursue a highly promising theory-building technology. Artificial neural networks have not been previously used to study social dilemmas. These experiments demonstrate the rich possibilities for promising sociological applications.

References

Allison, Paul (1992). "The Cultural Evolution of Beneficent Norms." Social Forces 71, 279-301

Axelrod, Robert (1984). The Evolution of Cooperation. New York: Basic.

_____ (1987), "The Evolution of Strategies in the Iterated Prisoner's Dilemma." In L. Davis (Ed.) *Genetic Algorithms and Simulated Annealing.* London: Pitman.

Axelrod, Robert and Douglas Dion. 1988. "The Further Evolution of Cooperation." *Science* 242:1385-1390.

Bainbridge, William (1995), "Neural Network Models of Religious Belief." *Sociological Perspectives.* 38:483-495.

_____ (1995), "Minimum Intelligent Neural Device: A Tool for Social Simulation." *Journal of Mathematical Sociology* 20:179-192.

Bainbridge, William, E Brent, K Carley, D Heise, M Macy, B Markovsky, and J Skvoretz (1994), "Artificial Social Intelligence." *Annual Review of Sociology 20*, 407-436.

Bandura, Albert (1977), *Social Learning Theory*. Englewood Cliffs, NJ: Prentice Hall.

Bendor, Jonathan (1993), “Uncertainty and the Evolution of Cooperation.” *Journal of Conflict Resolution 37*, 709-734.

Boyd, Robert and Peter Richerson (1985), *Culture and the Evolutionary Process.* Chicago: University of Chicago Press.

Cosmides, Leda and John Tooby (1987). "From Evolution to Behavior: Evolutionary Psychology as the Missing Link?" In J. Dupré (Ed.) *The Latest on the Best.* Cambridge, MA: MIT Press.

Dawkins, Richard (1989), *The Selfish Gene*. Oxford: Oxford University Press.

Durham, W. (1992), "Applications of Evolutionary Culture Theory." *Annual Review of Anthropology 21*, 331-355.

Frank, R. 1988. *Passions Within Reason: The Strategic Role of the Emotions.* New York: Norton.

Gallant, Stephen (1993), *Neural Network Learning.* Cambridge, MA: MIT Press.

Goldberg, David (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*. New York: Addison-Wesley.

Gould, Roger (1993), “Collective Action and Network Structure.” *American Sociological Review 58*, 182-196.

Heckathorn, Douglas. Forthcoming. "The Dynamics and Dilemmas of Collective Action.” *American Sociological Review*, April, 1996.

Hinton, Geoffrey (1992), “How Neural Networks Learn from Experience.” *Scientific American 9*, 145-172.

Hirshleifer, David and Eric Rasmusen (1989), "Cooperation in a Repeated Prisoner's Dilemma with Ostracism." *Journal of Economic Behavior and Organization 12*, 87-106.

Holland, John (1975), *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

_____ (1992), “Genetic Algorithms.” *Scientific American 7*, 66-72.

Kollock, Peter (1993), "An Eye For an Eye Leaves Everyone Blind: Cooperation and Accounting Systems." *American Sociological Review 58*, 768-786

Lopreato, Joseph (1990), "From Social Evolutionism to Biocultural Evolutionism." *Sociological Forum 5,* 187-212.

Macy, Michael (1990), "Learning Theory and the Logic of Critical Mass." American Sociological Review 55, 809-826.

_____ (1991a), "Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange." American Journal *of Sociology 97,* 808-843.

_____ (1991b), "Chains of Cooperation: Threshold Effects in Collective Action." *American Sociological Review 56*, 730-747

_____ (1993). "Backward Looking Social Control." *American Sociological Review 58*, 819-36

Macy, Michael and Andreas Flache (1995), "Beyond Rationality in Theories of Choice." *Annual Review of Sociology 21*, 73-91.

Macy, Michael, John Skvoretz, and William Bainbridge (1995), "Telltale Signs: Learning to Cooperate with Strangers." Paper prepared for 1995 annual meeting, American Sociological Association, Washington DC.

Nielsen, François (1994), "Sociobiology and Sociology." *Annual Review of Sociology 20,* 267-303.

Nowak, Martin and Karl Sigmund (1993), "A Strategy of Win-Stay, Lose-Shift that Outperforms Tit-for-Tat in the Prisoner's Dilemma Game." *Nature 364*, 56-58.

Nowak, Andrzej and Bibb Latané (1994), "Simulating the Emergence of Social Order from Individual Behavior." In N. Gilbert and J. Doran (Eds.) *Simulating Societies.* London: UCL Press.

Prata, Stephen. 1993. Artificial Life Playhouse. Corte Madera, CA: Waite Group Press.

Rummelhart, David and James McClelland (1988), *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. Cambridge, MA: MIT Press.

Schuessler, Rudolf (1989), "Exit Threats and Cooperation Under Anonymity. *Journal of Conflict Resolution 33,* 728-749.

Selten, Reinhard (1975), "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory 4*, 25-55.

Sigmund, Karl (1993), *Games of Life*. Oxford: Oxford University Press.

Thorndike, Robert (1911), *Animal Intelligence: Experimental Studies.* New York: MacMillan.

Vanberg Viktor and Roger Congleton (1992), "Rationality, Morality, and Exit." American Political Science Review 86, 418-431.