

Chapter One. Introduction

Altruism: 1) unselfish interest in the welfare of others, 2) accruing a cost to self in order to raise the fitness of another. (from Webster Dictionary 5th Edition)

We live in a society that values altruistic behavior. From childhood on, we are taught that selfish behavior is undesirable and that sharing with others is "good". In literature, Ebenezer Scrooge from "A Christmas Carol" is portrayed as a villain because of his miserly ways, while Della and Jim Dellingham Young in O. Henry's "Gift of the Magi" are canonized for the sacrifice each makes for the other. From a recipient's standpoint, it is easy to understand why altruistic behavior is encouraged. Who doesn't like receiving gifts or benefiting from the actions of another? The question of what motivates people to behave altruistically, however, is more difficult to answer.

Sociobiology rests on the assumption that important aspects of human behavior are guided by our genetic inheritance (Wilson 1975). Evolution favors individuals whose genetic inheritance predisposes them to behave in ways that maximize their probability of survival and reproductive success (or "fitness"). In such a scenario, it is difficult to imagine why individuals would engage in behavior that is costly to them.

Yet altruism exists, and it is not a social phenomenon unique to modern human society. Although they lived in what has been termed our "environment of evolutionary adaptedness," our hominid ancestors utilized valuable resources to care for the elderly and crippled members of their tribes (Tooby and Cosmides 1990). More compellingly, a wide variety of non-human species also exhibit altruistic behavior. The evolutionary origin of altruism therefore poses a frustrating puzzle to sociobiologists and is the source of a great deal of debate.

A variety of possible explanations have been proposed. Inclusive fitness theory provides a convincing explanation for altruistic behavior between kin, but fails to account for altruistic behavior between unrelated individuals. Other theories, like reciprocal altruism and mutualism, suggest that altruism evolved from other forms of cooperative behavior.

In this introductory chapter I will begin by explaining the logic behind a comparative approach to this question and follow it with a brief review of the research on social exchange in humans to date. I will then discuss the sociobiological debate surrounding the evolution of cooperation and the important role of reciprocal altruism within it. Finally, I will summarize the empirical literature on cooperation in non-human animals and conclude with an explanation of the rationale behind my experiments.

the Logic Behind a Comparative Approach

Before progressing further into a discussion of cooperative behavior, it is important to first explain the logic behind the comparative approach utilized in my thesis.

Behavior can be understood as a combination of proximate and ultimate causation. Proximate explanations focus on the direct sources of behavior, such as physiological mechanisms or cognitive capacities, and the development that occurs within an individual's life span. Theories of ultimate causation attempt to ascertain the adaptive significance of a behavior and how the behavior evolved through time considering an organism's limited abilities (Daly and Wilson 1983). While many ultimate causation theories exist, they must be tempered with an understanding of an organism's proximate mechanisms, because an evolutionary scenario is useless if the organism lacks the capability of behaving as the theory predicts (Stamps 1991).

The empirical literature to date has taken two general approaches to investigating cooperative behavior. One line has studied existing cooperative behavior in humans. While these experiments reveal general behavioral trends and the influence of learned strategy upon them, very little is known about the initial state of cooperative behavior.

Another set of studies has addressed the evolutionary origins of cooperative behavior by examining its existence in other species. By analyzing the similarities and differences between living animal species, one can deduce when and in which species a trait originated. In making these comparisons, it is important to consider which similarities may arise from similar selection pressures (homoplasy), and which stem from shared ancestry (homology). To distinguish between the two, it is

necessary to investigate the existence of cooperative behavior in a wide variety of species, both distantly and closely related.

One criticism of comparative studies of this nature is the problem of designing a paradigm that can be applied across species. Variables like cognitive ability or cooperative tendencies are not easily quantified. The traditional method of testing these variables is to measure an individual's performance on a task designed to require that ability. If the ultimate purpose of a task is to compare abilities across species, it makes sense to devise a paradigm that can be applied across species (Macphail 1987). However, finding analogous methods between humans and other species is difficult. Experimental paradigms with animals tend to rely heavily on training procedures, which are rarely used in humans. Similarly, methods applied to humans may be beyond the cognitive capability of most species to comprehend and this precludes their application to animals.

To be capable of reciprocal altruism, a species must first meet certain basic cognitive prerequisites (Cosmides and Tooby 1992). Some of these prerequisites, such as individual recognition or number discrimination, exist in most vertebrate species (though the latter is limited). Others, like the ability to track history of past exchanges, are more difficult to meet. The central question behind my thesis is whether any non-human animals are capable of reciprocal altruism.

In my thesis, I investigate reciprocal altruism in the cotton-top tamarin (*Sanguinus oedipus*), a species of New World monkey that diverged from the hominid line 40 to 60 million years ago. My experiments do not attempt to tease apart cross-species differences in cognitive ability; they seek to determine whether cotton-top tamarins meet the cognitive prerequisites for reciprocal altruism and if they exhibit it at all. Thus, finding an experimental paradigm applicable across species is not an issue.

Now that I have established the logic of my approach, it is important to review the history of research on reciprocal altruism in humans. In this next section I will discuss the rise of game theory and the development of the Prisoner's Dilemma, before presenting two proposed models for the maintenance of cooperative behavior in people.

Game Theory and the Prisoner's Dilemma

Game theory emerged in the mid-1940s as a means of determining the optimal course of action for rational players in situations involving a conflict of interest. A rational person is someone who acts to maximize personal benefit (Binmore 1992). Game theory predicts strategies based on the assumption that all players involved will act rationally. Thus, the optimal strategy profile (Nash equilibrium) that game theory presents is typically conservative; it does not represent the best possible outcome, but rather the best a rational player can expect when playing against another rational player. As will be discussed in more detail below, in scenarios involving irrational opponents, the Nash equilibrium changes and the rational player should modify his actions accordingly (Poundstone 1993).

To most people, acting to maximize personal benefit sounds fairly intuitive. However, choosing the best course of action is only straightforward in zero-sum games where the total winnings are fixed. In zero-sum games, Von Neumann's minimax theorem demonstrates that the rational course of action is to maximize the minimum amount one can win. For example, consider a situation where two people are deciding upon the best way to split a cake. In this scenario, one player has to divide the cake while the other gets to choose who gets what. Rational players will act to maximize the amount of cake they receive. Thus, the divider will split the cake as evenly as possible in order to maximize the minimum amount the chooser will leave him. In this scenario, choosing the optimal strategy is not difficult. It is only in non-zero sum games like Flood and Drescher's Prisoner's Dilemma that choosing the optimal course of action becomes more complicated (Poundstone 1993).

The Prisoner's Dilemma is a two-player game that explores the conflict between cooperating for mutual benefit or defecting for personal self-interest. The prototypical Prisoner's Dilemma scenario runs as follows: Two criminal suspects are placed in separate rooms. Currently, there is enough circumstantial evidence to put each away for one year. If both suspects confess to the crime, they both go to jail for three years. However, if one confesses but the other remains silent, the confessor (i.e. defector) walks away a free man while his partner goes to jail for five years. Thus, the payoff matrix looks as follows: [see Figure 1-1]

Figure 1-1.
Payoff Matrix for the Classic Prisoner's Dilemma.

		Player Two	
		Cooperate	Defect
Player One	Cooperate	1, 1	5, 0
	Defect	0, 5	3, 3

The optimal strategy for the rational actor depends on the strategy he thinks his opponent will choose. When played once, the Nash equilibrium in the Prisoner's Dilemma is mutual defection (Poundstone 1993; Axelrod 1984). Defecting is the best option whether your opponent defects or cooperates, even though a situation of mutual defection yields a worse outcome than one of mutual cooperation. The main barrier to cooperating is the commitment problem. Even if one is predisposed to cooperation, there is no way of making certain one's opponent does not defect. Thus, defection is the "safe" choice and there is little incentive to cooperate.

However, in practice, people do not always behave rationally. In experiments involving single plays of Prisoner's Dilemma scenarios, up to 50% of subjects attempt to achieve the cooperative outcome (Burnham 1998; Dickhaut, Hubbard, McCabe, and Smith 1995). Why do people behave irrationally and cooperate, even when it is not in their best interest to do so?

One argument presented by Robert Frank in his book *Passions within Reason* (1989) is that irrational behavior is an adaptive technique that may actually serve people better in the long run. His theory is that emotions overcome the commitment problem by serving as commitment devices. According to Frank, being known to experience certain emotions enables people to guarantee commitments. For example, a person who is known to dislike unfair bargains can credibly threaten to

walk away from one, even if it is in his narrow interest to accept it. By virtue of being known to have this preference, that person can become a more effective negotiator. Thus, the failure to pursue self-interest can confer advantages to the irrational individual.

As Frank notes, however, the success of this emotional model is contingent upon people's ability to successfully detect and punish defectors (those who "fake" emotion without following through on the commitment).

Research suggests that people are capable of doing so, and that they possess cognitive mechanisms specifically designed for reasoning about social exchange. The Wason selection task is an experimental paradigm that can be used to explore people's ability to detect violations of conditional rules. Unless it involves violations that can be interpreted as cheating on a social contract people perform poorly on the Wason selection task (Cosmides and Tooby 1993; Gigerenzer and Hug 1992). These findings suggest that although people lack a general-purpose ability to detect violations of conditional rules they are capable of detecting defectors in social situations.

One mechanism that has been proposed as an adaptive method of keeping track of "defectors" and quickly assessing whether to cooperate or defect is the friend-or-foe (FOF) module (Burnham 1998). Under the FOF module, people possess a variety of behavioral strategies that they employ depending on how they perceive their opponent. For example, if I perceive my opponent as a friend, I will perform behavior A. I do so because I anticipate higher probability of positive reciprocity if I perform behavior A, and therefore larger gain on my part. If, however, I perceive my opponent as a foe, I will perform behavior B, because I expect a lack of reciprocity on his part. If at any time I invoke a behavioral strategy but do not receive the expected response, I re-evaluate my perception of my opponent and shift my behavioral strategy accordingly.

Because it eliminates the necessity of conscious decision-making in choosing behavioral strategy, the friend-or-foe module is an adaptation that probably served our ancestors well in uncertain, potentially hostile situations and made the maintenance of cooperative behavior possible.

While Frank and Burnham's models provide possible explanations for how cooperative behavior is maintained in our species, they do not offer insight into how it may have evolved in the first place.

The first plausible explanation as to the origins of cooperative behavior developed from the application of game theory to the iterated Prisoner's Dilemma. In the iterated Prisoner's Dilemma, choices made can influence both present and future outcomes (Axelrod 1984). The temptation to defect still exists because all animals, human and non-human, have a tendency to preferentially value immediate payoffs over future ones (Rachlin, in press; Green and Rachlin 1996; Platt 1973). However, in scenarios involving an indefinite number of interactions, it becomes possible for cooperation to emerge and be profitable.

In a series of computer simulations involving the Prisoner's Dilemma run in the early 1980s, Axelrod demonstrated that a strategy of defection could be "defeated" (i.e. was not the optimum) by a strategy of conditional cooperation, tit-for-tat. Tit-for-tat, or TFT, is a strategy that cooperates on the first move and then mimics whatever the other player did on the previous move. The success of TFT sparked the development of a whole new subset within game theory that dealt with variations on the TFT strategy. Nor was its popularity restricted to the field of economics. As I will explain shortly, the principles of game theory and the TFT strategy were quickly adopted by biologists as a means of explaining the evolution of cooperative behavior in animals.

Reciprocal Altruism: the Sociobiological Debate

In biology, natural selection is accepted as the primary non-random mechanism behind evolution. However, as Darwin acknowledged in *Origin of Species* (1859), one problem with the theory of natural selection is it doesn't provide an adequate explanation for the prevalence of cooperative and altruistic tendencies in social animals. How did cooperative behavior evolve?

In the middle of the twentieth century, kin selection and group selection were presented as potential solutions to the problem. Kin selection, the key mechanism in the theory of inclusive fitness, considers the effect of selection not on the individual but on the gene, of which replicas exist in those of

common descent (i.e. kin). If two individuals are sufficiently related, altruism can benefit the gene, depending on the cost of the act. Thus, relationship between individuals is an important parameter in predicting how selection will operate (Hamilton 1964; Brown 1975; Dawkins 1976). Kin selection provides a persuasive explanation for eusocial behavior and communal breeding. However, it fails to account for cooperative relationships between unrelated individuals or between species. For example, inclusive fitness theory cannot explain the development of the mutually advantageous symbiosis between the fungus and alga that compose a lichen, nor the cooperative alliances exhibited among dolphins and chimpanzees.

Wade and Wilson's theory of group selection, trait group selection, attempts to provide a solution to the problem of cooperation between unrelated individuals (Wilson and Sober 1994). This theory also extends beyond the classic individual selection model to consider the evolution of both individual and group beneficial traits. According to the theory, cooperation can evolve even when it has a relative cost to the individual performing it, if the within-group cost is offset by a between group-benefit such that cooperative groups are more productive than selfish groups (Wilson and Sober 1994). While it is true that evolutionarily altruistic groups will outperform selfish groups, group selection has been the target of a great deal of criticism as unstable at the individual level. For example, within a cooperative group, the selfish members enjoy greater reproductive success than the altruists because they reap the benefits of cooperation without paying the price. Thus in time, the individual selfish strategy will take over, regardless of the benefits of group cooperation. Mathematical models demonstrate that except under very special conditions, group selection loses out to individual selection because selection at the individual level is more stable (Williams 1966; Dawkins 1976). One response to this criticism is that within-group differences are washed out when the groups are averaged, and that, because of its high reward value, altruistic behavior within a group is not incompatible with larger individual selfishness (Rachlin in press; Wilson and Sober 1998).

Despite the potential validity of this theory, it has been displaced in recent years by two other explanations: mutualism and reciprocal altruism. In mutualism, two or more individuals work together on

a task. In doing so, each derives an immediate benefit that could not have been obtained solitarily (Packer 1988; Clements and Stephens 1995). The only requirement for such a scenario is that the benefits of participating in the mutualistic act outweigh the cost. There is no expectation of future benefits or interaction; it could occur randomly and still succeed (Boyd 1988). Critics of mutualism argue that because there is no temptation to cheat, mutualism should not be counted as cooperative behavior; under this category, non-cooperators do worse than cooperators (Dugatkin 2000). The criticism is semantic. No-cost cooperation emerges as a likely solution precisely because natural selection does not act against it. However, the popularity of reciprocal altruism as an explanation for the evolution of cooperative behavior between unrelated individuals means that mutualism has traditionally been ignored.

Robert Trivers introduced the theory of reciprocal altruism in 1971. He was the first to mesh game theory and evolution by considering altruism in the context of the iterated Prisoner's Dilemma. Although the theory quickly gained popularity among biologists, a great deal of controversy over the definition of reciprocal altruism soon emerged. [For a list of authors, dates, and definitions, please refer to Table 1-2].

For the purposes of my thesis, reciprocal altruism will be defined as follows: First, the behavior must be altruistic (accruing cost to self to give benefit to recipient) and not attributable to kin selection. Second, the performance of the behavior must not depend on the receipt of immediate or mutual benefit. Third, a large but indefinite number of opportunities to exchange benefits must exist. Fourth, there must be a time delay between favors given and received. Finally, a mechanism must exist for detecting and punishing defectors, individuals who receive benefits without paying the altruist costs (Axelrod and Hamilton 1984; Wilkinson 1988; Rothstein and Pierotti 1988; Trivers 1971). Rapaport's tit-for-tat (TFT) strategy and all variations of it can be considered forms of reciprocal altruism.

As mentioned previously, Axelrod's computer simulations demonstrate that a tit-for-tat strategy can outperform an always-defect strategy (Axelrod 1984; Axelrod and Hamilton 1981). For TFT to emerge, all that is required is an "initial clustering" of TFT to overcome the threshold problem (Axelrod

and Hamilton 1981). TFT therefore provides an important theoretical model for how cooperative behavior can emerge in a world where the maximization of self-interest is the driving selective force (Rachlin, in press; Dugatkin 2000; Nowak and Sigmund 1992; Axelrod and Hamilton 1981).

However, TFT has been criticized for its evolutionarily instability. An evolutionarily stable strategy (ESS) is one that if adopted by all members of a population, cannot be invaded by a

Table 1-2: Evolving Definitions of Reciprocal Altruism over Time

Name	Definition of Reciprocal Altruism and Criticisms
Trivers (1971)	<ul style="list-style-type: none"> • Following conditions must be met: 1) behavior must be altruistic; 2) benefit to recipient must be greater than cost to altruist; and 3) significant time delay between favors given and received. • <i>Criticism?</i> difficulty in defining "significant" time delay (Packer 1986)
Wilkinson (1988)	<ul style="list-style-type: none"> • Same as Trivers except also requires presence of a mechanism for identifying freeloaders (individuals who give back less than they receive) and expectation of large, non-finite number of interactions during one's lifetime. • <i>Criticism?</i> Debate over whether these are factors that contribute to evolution of RA and should not be considered part of the definition itself. I disagree with this criticism. If reciprocal altruism is to be presented as an ESS, these conditions must be included within the definition.
Koenig (1988)	<ul style="list-style-type: none"> • Same as Trivers, except does not require a time delay, merely the anticipation that the recipient will return the favor. • <i>Criticism?</i> If animals live in a stable long-lived population with frequent interactions between altruists, cooperation can be achieved without the cognitive ability to anticipate others' long-term behavior (Brosnan and DeWaal 1999)
Brosnan and DeWaal (1999); DeWaal and Luttrell (1988)	<ul style="list-style-type: none"> • General definition of reciprocal altruism identical to Trivers. However, also subdivides reciprocal altruism into two subcategories: <ol style="list-style-type: none"> a. symmetry-based reciprocity (members of a species preferentially direct favors to close associates; the distribution of favors will automatically be reciprocal owing to the symmetrical nature of association) b. calculated reciprocity (exchange of favors based on conscious scorekeeping of given and received favors) • <i>Criticism?</i> First subcategory can be more appropriately described as kin selection or group selection.

mutant strategy (Maynard Smith 1982). Although TFT can resist invasion from defecting strategies, TFT is only collectively stable compared to other cooperative strategies. This means an always-cooperate strategy could “drift” into TFT populations and eventually reach high enough frequencies to replace it (Dugatkin 2000). This is problematic because an always-cooperate strategy is vulnerable to exploitation by defecting strategies.

In response to this argument, evolutionary game theorists developed a number of “reactive” variations of TFT that might be evolutionarily stable enough to maintain themselves within a population even after they have served their function as a “teaching strategy” for cooperative behavior. [For a brief list of popular evolutionary strategies refer to Table 1-3]

Table 1-3:
Popular Variations on Tit-for-tat

Name	How it Differs from TFT
Tit-for-tat <i>Axelrod and Hamilton (1981)</i>	Starts by cooperating then mirrors what opponent did on last move.
Tit-for-two-tat	Similar to TFT except forgives all single defections.
Generous TFT <i>Nowak and Sigmund (1992)</i>	Similar to TFT except overlooks single defections about 1/3 of the time.
Simpleton (Pavlov) <i>Rapaport; Nowak and Sigmund (1993)</i>	Instead of just reacting to opponent's last play, remembers own last play as well and acts accordingly. Thus, it only changes strategies when confronted with defection.
Two-tit-for-tat	Similar to TFT except punishes all single defections by defecting twice in return.
Massive Retaliatory Strike	Cooperates until defected upon and then never cooperates again.

Although these variations of TFT are theoretically sound, criticism of reciprocal altruism continues. One major argument is that the cognitive requirements involved in the tracking of individuals and favors owed are beyond the capabilities of many animal species (Stevens and Stephens, in press). This criticism is potentially valid, but only regarding the existence of a tracking mechanism.

While the ability to reason about the goal-oriented behavior of others means that more advanced exchange systems can develop, it is not essential to reciprocal altruism. If animals live in a stable long-lived population with frequent interactions between altruists, cooperative behavior can be achieved without the cognitive ability to anticipate others' long-term behavior (Brosnan and DeWaal 1999). The high frequency of interactions means that as long as a tracking mechanism exists, defectors will be detected and a theory of mind is unnecessary.

The cognitive requirements for the maintenance of tracking mechanisms like the FOF-detector are not beyond the ability of many animals; most vertebrate animals are capable of individual recognition, memory, and limited number discrimination (Cheney and Seyfarth 1990). Whether any non-human animals actually possess such a tracking mechanism is a trickier issue to resolve. Memory-based mechanisms are notoriously difficult to differentiate from simpler alternatives. Until such a mechanism is found, Stephens and Stevens' criticism cannot be dismissed.

Clearly, the only effective method of responding to this criticism is to present a model of reciprocal altruism in a non-human animal species. If it is responsible for the emergence of cooperative behavior, reciprocal altruism should be widely prevalent in nature.

Reciprocal Altruism in Nature

Many animal examples of reciprocal altruism have been presented. Unfortunately, most of these examples have proved either unclear/inconclusive or better explained by other theories like kin selection or mutualism. [Refer to Table 1-4 for a list of well-known models of cooperative behavior and why they do not qualify as reciprocal altruism].

One of the earliest and most popular models of reciprocal altruism was blood sharing in vampire bats (Wilkinson 1984). Female vampire bats regurgitate blood to share with roostmates who have failed to obtain blood in the recent past. The cost of sharing blood is not very high,

Table 1-4: Examples of Cooperative Behavior in Animals and Why They Don't Qualify as Reciprocal Altruism (RA)¹

Author	Example	Why Not RA?
Calhoun 1967	Rats demonstrate cooperation in a task that requires two to stand together on a lever in order for both of them to receive water.	Mutualism
DeWaal 2000; Rose 1997	Food sharing in chimpanzees and capuchins	Tolerated begging or coercion (Dugatkin 1997; Stevens & Stephens, in press)
Dugatkin 1991; Milinski 1987; George 1960	When examined in mirror experiments, patterns of predator inspection in shoaling fish appear to resemble TFT.	Predator orientation and social cohesion (Stephens 1995)
Hart & Hart 1992; DeWaal 1982	Grooming behavior in animals like impala and chimpanzees	Possibly byproduct mutualism. No punishment of cheaters and cost involved in grooming is questionable. Kin selection
Wilkinson 1984	Blood sharing in vampire bats	Kin selection, group selection
Cheney & Seyfarth 1990; Koenig 1988; Sherman 1985, 1977; Trivers 1971	Group alarm calling in non-human primates, ground squirrels, and birds	Kin selection and possibly byproduct mutualism (Dugatkin 1997; Dublin 1983)
McGrew 1992; Packer 1988, 1991; Douglas-Hamilton 1975	Alloparenting in New World monkeys, dwarf mongooses, and elephants	Mutualism or byproduct mutualism (Scheel & Packer 1991)
Telecki 1973; Boesch 1994, 1989; Packer and Ruttan 1988	Cooperative hunting in lions, wolves, hyenas, and chimpanzees	

but it is of critical importance to the recipient, because the bats will die if they do not have a blood meal every sixty hours. Moreover, bats only share blood with other roostmates, those who are likely to have shared with them in the past. Thus, upon initial observation, blood sharing appears to satisfy all the requirements for reciprocal altruism.

However, blood sharing is inconclusive and cannot be described as a definitive example of reciprocal altruism for two reasons. First, female vampire bats roost together in social groups composed of several matriline. A high degree of relatedness exists between bats from the same matriline. Studies have demonstrated that bats preferentially share blood with related group members (McNab 1973). Therefore, kin selection provides an equally likely explanation for blood sharing and cannot be eliminated.

Second, a mechanism for detection of "defectors" has not been definitively proven to exist. In his experiment, Wilkinson (1984) determined that females would only share with roostmates, and not with strange bats. While it is true that roostmates are more likely to have shared blood in the past, this comparison alone is insufficient proof that a mechanism for detecting defectors exists. Thus, while promising, it is unclear whether blood sharing in vampire bats can actually be explained as reciprocal altruism.

Another initially promising animal example of reciprocal altruism that has not stood up to rigorous testing is predator inspection behavior in shoaling fish (Dugatkin 1991, Milinski 1987, and George 1960). Predator inspection behavior describes a fish's slow, saltatory movements away from a school and towards a potential predator. Its functional purpose is apparently to determine whether a predator is in "hunting" mode (Magurran and Pitcher 1987). Predator inspection puts the inspector at greater risk of being eaten, but also confers benefits; once inspectors return to their school, noninspectors who have not seen the predator decrease their rate of foraging and increase their rate of

skittering, an anti-predator behavioral response (Magurran and Higham 1988). Inspection also has the potential to deter predator attacks. Thus, although inspection can be costly, a group with no inspectors suffers higher predation rates.

Because the payoffs of predator inspection appear to approximate a Prisoner's Dilemma, Milinski and Dugatkin (1987, 1991) conducted a series of experiments on stickleback fish that sought to determine whether fish inspecting a predator in pairs exhibited reciprocal altruism by using a TFT strategy. They used mirrors to simulate a cooperating inspection partner and a defecting partner. In the cooperating task, the mirror ran the length of the tank so the perceived inspection partner always stayed by the side of the subject. In the defector task, the mirror only went halfway, and was placed at an angle so that it simulated a co-inspector who stayed on the "safe" side of the tank. The results of these experiments suggested that fish do follow a TFT strategy. Fish in the cooperating mirror treatment were more likely to be found in sections of the tank closest to the predator, while fish in the detecting mirror treatment were more likely to be found on the "safe" side of the tank. Moreover, inspectors appeared to "forgive" prior moves of defection if followed by moves of cooperation.

These experiments are, however, open to two major criticisms. First, reciprocal altruism requires the ability to recognize co-inspectors and remember their tendency to inspect or not inspect in the past. The experiments were conducted with mirrors; thus, evidence of individual recognition or tracking of past behavior could not be tested (Lazarus and Metcalfe 1990, Masters and Waite 1990). Secondly, more parsimonious explanations for the predator inspection behavior have not been eliminated. Stephens et al (1997) repeated the initial Milinski and Dugatkin experiments in mosquitofish and also added another experiment designed to separate the effects of social cohesion and predator cohesion. Stephens' experiment demonstrated that predator inspection behavior was actually the result of social cohesion and predator orientation effects, and could therefore be more accurately categorized as byproduct mutualism instead of reciprocal altruism.

The current lack of examples of reciprocal altruism in the animal kingdom is troublesome. If none exist, then the validity of the theory as an explanation for the evolution of cooperation must be re-

examined. Thus, finding a concrete model of reciprocal altruism in a non-human animal species is important.

Why Cotton-top Tamarins?

In my thesis, I investigate the existence of reciprocal altruism in the cotton-top tamarin (*Sanguinus oedipus*), a species of New World monkey that has been diverging from the hominid line for a period four times longer than that in which hominids have been separated from the rest of the primate order (Fleagle and Kay 1997). There are several important reasons for this choice.

First, the cotton-top tamarin is a primate species relatively distant from humans. This distance is important because it means they can serve as an important comparison group to humans and their more closely related primate relatives, upon which numerous cooperation studies have already been conducted. As mentioned previously, making these comparisons is important because it helps us separate out cognitive similarities that arise as a result of homology and those that arise from homoplasy.

Secondly, cotton-top tamarins are a good choice for an experiment on reciprocal altruism because previous experiments have already established that they exhibit cooperative behavior. Cotton-top tamarins demonstrate group monitoring for alarm calls, actively share food (McGrew 1992), exhibit altruistic behavior towards cagemates (Moskovitz, Santos, and Hauser, in submission), and like humans, breed cooperatively (Hrdy 1999; Cleveland and Snowdon 1984). As natural cooperators they may be more likely to exhibit reciprocal altruism.

Finally, they are a good choice because of their cognitive abilities. Cotton-top tamarins are capable of individual recognition (Jordan 2001), they can learn from experience, they have the ability to reason about the goal-directed behavior of other individuals (Santos 1997), and they can discriminate between small numbers (Flombaum 2002). This means they are likely to possess the cognitive

mechanisms such as detection of defectors or tracking history of past exchanges involved in reciprocal altruism.

Summary and Aims

In the following chapters I will present the results from three experiments. Chapter Three reviews the relevant empirical literature on cooperative behavior in non-human animals and discusses the results of my first experiment, which investigated whether cotton-top tamarins would participate in a game of reciprocity. Chapter Four considers the question of punishment in animal societies and presents the results from my second experiment, which examined whether the tamarins could distinguish between a cooperator and a defector and modify their behavior accordingly. Chapter Five reviews the results from the previous two experiments and presents the results of my final experiment, which explores the effect of experience/learning upon the tamarins' reciprocal exchanges and tested their understanding of the experimental paradigm.

The general aim of these experiments is to explore the validity of reciprocal altruism as an evolutionary theory of cooperation by testing its existence in a non-human primate species, the cotton-top tamarin.