Cooperative Behavior Strategies in Colored Trails

A thesis presented

by

André Veloso Moura

to
Computer Science

in partial fulfillment of the honors requirements
for the degree of
Bachelor of Arts

Harvard College
Cambridge, Massachusetts
April 8, 2003

Contents

Abstract

This work studies agent interactions in 2-player games of Colored Trails. We design and implement three styles of agent behavior: GIVER, TAKER, and RECIPROCATOR. We test these computer players of CT against each other in repeated rounds of the game, and we draw conclusions based on their performance. We find that the RECIPROCATOR agent is able to avoid exploitation by less generous players while at the same time achieving high levels of profitable trading when interacting with more cooperative agents. We also include a brief survey of related work in the field of computer science and beyond.

Chapter 1

Introduction

People are intrinsically social creatures, and our interactions are characterized by conversation and negotiation [4]. People, and even many animals, communicate with each other and join social groups because they have the potential to benefit from each other and to accomplish more as a group than as an assortment of individuals. Although in the long run, individuals may benefit from association with a group, often in the short run, members of the group will assist others at a cost to their own utility.

A collection of students, organized as a *study group* preparing for an exam, provides a good example to illustrate the potential costs and benefits of group membership for individuals. One reason students choose to join a study group is that the group provides an efficient way to have specific questions answered. However, the student in the group who is answering a particular question may gain little by discussing a topic with which she is already knowledgeable; a better use of that student's limited time would probably be to learn an unfamiliar topic. However, the student takes the time to answer the question because her membership in the group brings value to her; she expects that the other student, whose question she answered, might be able to answer own of her own questions in a future exchange.

In the study group example, the knowledge that students have and the limited time they have to share it are the resources which are being negotiated and traded. In other groups, other resources are at stake. Regardless of the types of resources, all groups and marketplaces work because the individuals who hold the resources may not necessarily be the ones who value them the most. This disparity between ownership and value can arise for a variety of reasons: the owner may have an abundance of a certain resource and therefore a decreasing marginal utility, or the owner may simply not require the resources he holds to accomplish the goals for which he aims.

However, this opportunity for synergistic gains and increased utility for the group as a whole, which drives group formation, does not guarantee that individual members of a group will necessarily derive individual benefits. In our study group example, a student who does not ask his own questions or talk about the subject area with which he is least familiar is not benefiting individually though the sum of the utilities of all the study group members may be higher as a result of the existence of the group.

Our work examines how agents act to try to ensure that in addition to increasing the entire group's welfare, they also benefit individually from group membership. We study the interactions between the smallest of all the possible groups: groups of only two agents. We model very specific behavior protocols, and we study the effects of specific negotiation strategies on both individual and group utilities. We also study the issue of an agent's *reputation* in a group and the ways *reputation-aware* agents might try to increase their own reputations and gain favor with other agents.

1.1 Sen and Reciprocity

Sen's 1996 paper entitled, "Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents," [7] served as a starting point for this work in terms of defining the language and basic behaviors of our cooperative agents. Within a package delivery domain, Sen experimented with four different agent types – philanthropic, selfish, individual, and reciprocative. Each of the agent types uses a different decision procedure to decide whether or not to cooperate with other agents.

Philanthropic agents go out of their way to help other agents, and they always accept cooperation requests; selfish agents request help from other agents but do not offer any assistance; individual agents perform their assigned tasks without asking for or giving any help; and reciprocative agents keep a balance of the costs and benefits of past exchanges with the other agents, and they make their decisions based on the balance of these past interactions.

Sen's hypothesis, confirmed through his experimental results, is that reciprocative agents are able to avoid being exploited by selfish agents while in the long run, achieving similar levels of utility gains through cooperation as the philanthropic agents are able to achieve. Sen's work emphasizes the value of using reputation as an important factor in making decisions about whether or not to cooperate with other agents.

1.2 Our Work

Our work confirms that Sen's hypothesis, regarding the importance of reciprocity to avoid exploitation, carries over to the Colored Trails domain. We design and implement three styles of agent behavior: GIVER, TAKER, and RECIPROCATOR. Though these agent strategies are loosely based on Sen's philanthropic, selfish, and reciprocative agents, the specifics of how the agents propose and deal with cooperation requests vary significantly. In our experiments, we test our CT computer players against each other in repeated rounds of the game, and we analyze the players' scores to gain insights about the success of their behavior strategies.

The paper is broken up into the following chapters. In Chapter 2, we describe Colored Trails, the game developed by Grosz and Kraus that is the domain within which we experiment with the different behavior strategies. In Chapter 3, we introduce and detail the specifics of the strategies that are implemented in each of our three experimental agents: GIVER, TAKER, and RECIPROCATOR. Chapter 3 also explains our experimental setup. Chapter 4 presents and analyzes the results of our experiments, and Chapter 5 draws some conclusions and comments on many of the possible areas for future work. In Chapter 6, we provide a survey of some related work in the field of computer science and beyond.

Chapter 2

Colored Trails Domain

Within the Colored Trails game, we designed experiments to test different agent behavior strategies. CT is a flexible game which was developed by Grosz and Kraus to provide an abstract framework to study various topics, including agent interactions and decision making. The game can be played with many different options, and in addition to describing the essentials of the game, this chapter also describes the specific playing options that were used in our research. The chapter also details some of the different strategic considerations that CT players need to keep in mind as they play repeated rounds of the game.

2.1 Basics

The CT board is an NxM board of colored squares chosen from a palette of k colors. One or more of the squares on the board are designated as goal states. Each of the n players starts at a given location on the board with a set of colored chips chosen from the same palette of k colors as the board squares. Each players is also assigned a goal state, and multiple players may share the same goal state.[1]

The game is played in cycles of movement phases followed by communication phases. During movement phases, players are allowed to move to a bordering square provided that they have a chip of the same color as the square into which they want to move, and once the move is made, the player forfeits that chip. During the communication phase, players can send each other messages in which they propose and execute chip requests and exchanges. CT provides a good domain in which to study inter-agent interactions because of the negotiating and deal-making that goes on during the communication phases.

In the general CT game, the board visibility of the players can vary so that in some trials, players are permitted to see the entire board, and in others, they are not. Knowledge of the chips that other players hold is also a variable in the general CT game.

In our implementation, our agents play with the following parameters. Boards are of size 5x5; the palette is composed of 3 colors; and there is only one goal state on the board. The players have full visibility of the colors of all the squares on the board, and they can also see the locations of all the other players. However, players do not have information about the set of

[1]At this point, the confused reader may find it helpful to refer to Appendix A for a screen shot of the human computer interface to the CT game. The screen shot helps to visualize the game.

chips that each of the other players has. We limit the communication phase to 500 msec and the movement phase to 100 msec because we experiment with computer agents who do not need any more time than that to act.

2.2 Scoring

A round of a game ends once all the players have either reached the goal state or been idle for three consecutive movement phases. At the end of a round, each player receives a score for that round according to a scoring function which takes the following variables into account:

- *moves* - the number of moves made by the player
- *chips* - the number of chips the agent was left with at the end of the round
- *distance* - the Manhattan distance from the player's position at the end of the round to the goal state
- *bonus* - has a value of 1 if the player reached the goal state, otherwise is equal to 0.

The base score each player receives is:

$$base\_score(player_i) = bonus \cdot A - \frac{a \cdot moves}{M+N} + \frac{b \cdot chips}{M+N} - \frac{c \cdot distance}{M+N}$$

A player maximizes her base score by reaching the goal state in the shortest number of moves possible and by finishing the games with as many chips as possible. The A, a, b, and c parameters can be used by the player to assign a value to the chips that she holds because the player can compute the base_score that is associated with any given chip set and board location.

However, a player's total score weighs the base_scores of the other players according to a d parameter:

$$score(player_i) = base\_score(player_i) + \frac{d \cdot \sum_{j=1, j\neq i}^{numberofagents} base\_score(player_j)}{numberofagents - 1}$$

The fifth parameter adjusts the importance that other players' successes or failures have. A value of 0 for d indicates that a player's score in a round is completely dependent on her performance. Values of d greater than 0 mean that the players' interests are aligned while negative values of d signify that the different players in the game are antagonists who are not at all interested in seeing each other succeed.

In our experiments, we set the parameters of the scoring function as follows: $A = 100$, $a = 30$, $b = 20$, $c = 0$, and $d = 0$. These parameters are intended to primarily emphasize and reward reaching the goal state. We set $c = 0$ because our players start at random distances from the goal, and we are not interested in skewing the players' scores based on their random start positions, and we are also not interested in studying negotiations that will allow us to get closer to the goal state but not reach it. We run the experiments with the scoring parameter d set to 0 because we want the players to be scored purely based on their own individual performances so that we can more easily compare the relative successes and failures of the different behavior strategies.

2.3 Pathfinding

Since the players have full board visibility in our experiments, it would be possible for each player to compute his shortest path to the goal state from his current location on the board. The pathfinding computation is implemented in the CT game server using a dynamic programming algorithm.[2] The server sends path information to each of the players at the end of every movement phase.

From any location on the board, the pathfinder information specifies all the different chip sets that would allow a player to get to the goal state and the set of movements corresponding to each of those chip sets. The paths to the goal state are listed in increasing order by the number of moves required to reach the goal state. Chip sets that subsume other chip sets are not listed because it is assumed that a player would always prefer to take the shorter path if given the choice.

2.4 Negotiation Framework

As mentioned above, during the communication phases, the CT infrastructure gives players the opportunity to send each other various different types of messages. The computer agents in our experiments send chip request messages and chip exchange request messages. The distinction between the two is that in a chip request message, the proposing agent merely states the

[2]Implemented in conjunction with Marco Carbone.

chips she needs. The player receiving the request can use that information to compose a chip exchange request, or he can simply send the proposing agent the requested chips, or of course, he can also elect to do nothing.

The chip exchange request message outlines a deal proposal in which an actual exchange of chips between the two parties would take place. If the player receiving this exchange proposal, agrees with the deal, he will simply send the chips and then wait to receive his promised chips. In this research, we make the simplifying assumption that CT players are honest with each other, i.e. the player who proposed the chip exchange request will follow through with the deal if he receives the requested chips.

From this brief discussion, we can see that there are two distinct facets to a CT player's negotiation strategy: *proposer* behavior and *deliberator* behavior. Players must make decisions about the deals that they are going to propose, but they must also deliberate over the deals which are proposed to them. When describing negotiations in this paper, we generally use female pronouns to refer to the proposer, and we use male pronouns in reference to the deliberator. Also, our discussion makes the assumption that deal proposers never propose deals in which their scores in that round would decrease.

2.5 Strategic Considerations

CT is well suited for the study of inter-agent cooperation strategies because players can significantly increase their scores by exchanging chips with each other. Several decision scenarios arise in CT that simulate the types of interactions that happen among individuals in the real world or among computer agents in practical applications. In this section, we describe the choices that CT players must make as they negotiate deals and build up reputation histories with other players, and we explain how these choices parallel the types of decisions that are made in other scenarios.

In the following sections, we discuss the fact that in some deals, only one party gains whereas in others, both parties might benefit, but it is important to note that when the proposer proposes her deal, she can not know whether execution of the deal will increase or decrease the deliberator's base_score because she has no information about the chips that deliberator holds. However, over the course of the game, the proposer is able to learn whether deliberator granted her a favor or in fact, also benefited from the deal. Proposer learns this by examining the path that deliberator followed to the goal, and then computing whether deliberator would have been able to follow a shorter path if he had not agreed to the deal.

2.5.1 Even Deals

We define an *even* deal to be an exchange in which the score of the deliberator neither goes up nor down. The deliberator has to go through the small inconvenience of executing the exchange, but other than that his final score will remain unchanged as a consequence of the deal. What the deliberator gains from an even deal is the good favor of the deal proposer whose score may go up dramatically as a result of the successful execution of the deal.

An example of an even deal in CT: the proposer realizes that one more blue chip would allow her to reach the goal and receive the associated score bonus. The proposer suggests a blue-for-red chip exchange to the deliberator. If the deliberator has an extra red chip that is not used along her shortest path to the goal, then he faces an even deal situation in which his score in that round is independent of whether or not he accepts the proposed deal.

Even deal situations come up often in human interactions. One real world example took place recently over an email distribution list in a college residential community. One member of the community requested a one-for-one exchange of firewood logs so that she could start a fire on a cold day. The proposer's logs were slightly damp, and she was hoping to find someone with enough dry logs that it would make no difference to have a couple of damp ones that would eventually dry out.

2.5.2 Favor Request Deals

We define a *favor request* deal to be a deal in which the proposer asks the deliberator for a favor. The proposer suggests a deal in which the deliberator's score will actually decrease if he accepts the proposal. The typical CT example of a favor request deal: proposer needs one blue chip to reach the goal state. If the deliberator gives away his blue chip, he will still be able to reach the goal state, but he will have to take a longer path and thereby have a lower score in that round.

A player's response to a favor request deal is one of the major defining characteristics of his behavior strategy. Agents who are *reputation-aware* will grant favors when they are deliberating because they may be looking to build up favors with other agents so that their own future requests will be looked upon kindly. Also, if the d parameter of the scoring function is sufficiently high, the deliberator's final score in the round may actually be higher even if he has a lower individual base_score. However, it is important to note that the deliberator has no information about the deal's scoring effect on the proposer because the deliberator has no information about the proposer's chip set.

The proposer must also consider how often she requests favors. If the favor will only grant the proposer a small score increase in the form of a shorter path to the goal, it may not be in the proposer's best long term interests to ask for a favor of that kind because the deliberator may be less inclined to grant future requests that may have a more significant score impact for proposer.

There are many practical examples of human beings requesting favors of each other and usually people will go out of our way to help individuals who have built up a reputation of trust and mutual cooperation. For example, a college student who is too busy to return his library books is likely to have more luck if he asks his roommate to do him the favor of returning the books than if he asks a casual acquaintance for the favor.

2.5.3 Win-win Deals

The last deal type that we define is the *win-win* deal. This is an exchange in which both the proposer and the deliberator directly gain from the successful execution of the deal. In most cases, it seems obvious that the deliberator will gladly accept the proposed deal in a non-antagonistic environment, i.e. the d parameter of the scoring function is not less than 0.

Chapter 3

Experimental Setup

We have designed and implemented three different classes of computer players for the CT game: GIVER, TAKER, and RECIPROCATOR. This chapter describes these agents on a high-level and also details the specific strategies that each of the agents follows as it plays the game. GIVER and TAKER are simple strategies that do not adjust their behavior depending on what the other players do. RECIPROCATOR is a more interesting agent because it has an adaptive strategy, meaning that the agent's decision-making takes into account its previous interactions with the other players of the game, and it adapts its behavior accordingly. The chapter also describes the experimental setup in which the computer players, and their variations, face off against each other.

3.1 GIVER

GIVER's deliberator strategy is to generally grant other player's requests unless granting the request would make GIVER unable to reach the goal state. The rationale is that by helping others whenever possible the agent tries to build up a strong reputation with the other players in the game and make as many "friends" as possible, GIVER hopes that its reputation will cause other players to respond positively to its own chip request messages in future rounds.

As a deal proposer, GIVER only requests chips when the goal state is unreachable with its current chip set. Like a friend who does not ask for help on a trivial matter, GIVER does not request chips that would allow it to get to the goal along a shorter path because it "saves up" its favor requests for the times when they would have the most impact.

We experiment with two variations of GIVER's behavior in proposing deals. In one version, the agent simply requests the chips it would need without offering any in exchange. In the second version, GIVER gives away all the chips that would not be used in reaching the goal if it were to receive the chips it requested. In the first variation of the GIVER's proposer strategy, the deals that are proposed are more likely to be favor request deals than in the second version. In the second version, we hypothesize that the deal proposals are more likely to be well received and accepted because they are more likely to be even deal or win-win deal proposals since the deliberating agent has more chips. The question we have is whether GIVER's score loss in giving away all its extra chips will be offset by the gain it receives from having its requests granted with a higher likelihood.

We can see that in CT even a simple-minded agent has a lot of decisions to make and issues to consider as it contemplates which chips to ask for and which to give away.

3.2 TAKER

TAKER pursues a deliberator strategy that is diametrically opposed to GIVER's. TAKER does not concern itself with its reputation, and it only grants other player's requests if TAKER's base_score in the round would increase from the exchange, i.e. if the proposal would be a win-win deal. TAKER is modeled after the hard-driving, if potentially short-sighted, business person who only considers the short term costs and benefits of any particular deal; TAKER does not consider the potential benefits of building up a "working relationship."

But TAKER does pursue an active proposer strategy, and unlike GIVER, it will request chips even if it already has a path to the goal available to it and the requested chips will only help TAKER reach the goal along a shorter path. With the TAKER agent, we do not experiment with the two variations in proposing that are described in the GIVER section. TAKER's proposing module only requests chips; it never offers its own chips to other agents.

3.3 RECIPROCATOR

RECIPROCATOR's basic strategy is to act towards others as others have acted towards it in previous rounds. If RECIPROCATOR thinks that another player has granted it favors in the past, then as it deliberates a deal proposal from that player, it will be more likely to grant a favor request deal and take a loss in its base_score for that round.

RECIPROCATOR keeps track of its "favor balance" with all players with whom it has interacted in the past. The favor balance reflects how many score points RECIPROCATOR has gained or lost from its interactions with other players. An example: RECIPROCATOR sends a chip request message to PLAYER for a chip that would increase its score by 100 points because it would allow RECIPROCATOR to reach the goal. If PLAYER agrees to the request and gives away the chip, then RECIPROCATOR will register an increase of 100 points in its favor balance with PLAYER. Conversely, if RECIPROCATOR agrees to a deal with PLAYER which lowers its base_score for the round, then a decrease would be registered in the favor balance.

As a deliberator, RECIPROCATOR decides whether or not to grant favors based on the balance of the agent proposing the deal. The probability that RECIPROCATOR will grant a favor request is:

$$\frac{1}{\pi} \cdot \arctan\left[\alpha \cdot (favor\_balance - loss\_in\_round) + generosity\_factor\right] + \frac{1}{2}$$

where α is a constant, $loss\_in\_round$ refers to the decrease in RECIPROCATOR's base_score in the round, and $generosity\_factor$ is a variable that can

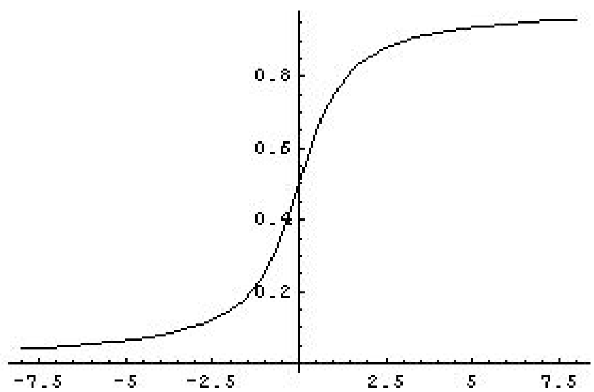

Figure 3.1: Graph of the shape of RECIPROCATOR's probability function

change in different versions of the RECIPROCATOR agent. The larger the value of the *generosity_factor*, the more likely it is that RECIPROCATOR will grant the favor request.

When playing against only one other player, RECIPROCATOR's proposer strategy is the same as GIVER's if $favor\_balance \geq 0$. Otherwise, the proposer strategy is the same as TAKER's.

3.4 Setup Details

In our experimental setup, two players face off against each other and play 100 rounds of the CT game on 100 different boards and initial chip distributions. The data in Chapter 4 are the cumulative scores that the agents accumulate over the course of 100 rounds of play. The same set of 100 boards and chip distributions is used in all the experiments in Chapter 4.

3.4.1 Random Boards

As mentioned above, experiments are run on 5x5 size boards with a palette of three possible chip colors. The color of each square on the board is chosen randomly from the palette. The starting locations of the players on the board are also chosen randomly. However, in our experiments, the goal state is always designated as the bottom, right corner of the board so that players will more often have longer paths to the goal and therefore more path options with any given chip set; experience showed us that goal states in this corner produced the most noticeable results.

In our experiments, we save the boards, and after the agents play all 100 rounds against each other, we run the experiments again with the starting locations and chip distributions of the agents reversed. This allows us to better compare the performance of the different agent strategies when they are played against each other because we can correct for the effect that

3.4.2 Chip Distributions

Each of the players is given a starting set of chips with which he may or may not be able to reach the goal state. The number of chips that a player receives is equal to the player's Manhattan distance to the goal state plus n extra chips. We experiment with different values of the variable n. These chips are distributed randomly over the palette colors, and players are not guaranteed to receive a minimum number of any of the chip colors.

Chapter 4

Experimental Results and Analysis

We divide our experiments into two categories: homogeneous experiments in which the same agent types play with each other, and heterogeneous experiments in which different agent types face off against each other. We use the homogeneous experiments to draw some conclusions about CT in general and also to make comparisons between the combined scores that the different agent types receive. Our emphasis in the heterogeneous experiments is to look more at the individual scores that the agents receive and examine the degree of success that the different strategies enjoy when they encounter other strategies different than theirs.

4.1 Homogeneous Experiments

On every set of random boards in a game setup, we first play our control agents against each other. The control agent neither requests nor gives out any of its chips; it merely serves as a baseline so that we know the score that players of the game would receive if they did not make any negotiation attempts and merely followed the shortest available path to the goal if such a path were even possible with the initial chip set.

We hypothesize that two GIVER agents playing against each other will have a higher combined score than either the two TAKER agents or the two RECIPROCATOR agents because we expect that the two GIVER agents will be able to reach the goal more often than the other agents as a result of the higher sharing of chips between the agents. We also hypothesize that the RECIPROCATOR agents will reach a combined score close to that of the GIVER agents because we expect them to settle into a "groove" of helping each other.

Agents Playing	*Values of* n				
	0	1	2	3	6
CONTROL VS. CONTROL	6450	9925	12675	14895	18840
TAKER VS. TAKER	6450	9995	13015	14985	19245
RECIPROCATOR VS. RECIPROCATOR	8010	11640	14245	16120	19290
GIVER-1 VS. GIVER-1	8195	11630	14845	16730	19325
GIVER-2 VS. GIVER-2	8845	12000	14915	16605	19320

Table 4.1: Cumulative combined scores over 100 rounds

Tables 4.1 and 4.2 show data that confirm these initial hypotheses. The variable n in the column heading refers to the number of additional chips that agents receive at the beginning of the game beyond the number they receive depending on their Manhattan distance from the goal state.

The first thing that is clear from this data is that increasing values of n have a major impact on the scores of the agents playing the game. In the control experiments in Table 4.1, the first extra chip increases the combined scores of the agents by approximately 50%, and the second chip increases their scores by over 25%. We see that in this setup, the addition of only one additional chip greatly increases a player's chances of reaching the goal. However, by the $n = 6$ case, we see that the scores essentially converge and the agents' behavior strategies have little effect on their scores because little trading is needed for the players to reach the goal state. These observations serve as indicators of the high potential for increased scores from trading chips.

We also note that the second variation of the GIVER agent achieves slightly higher combined scores. This is because when an agent offers its extra chips to the other player, more possibilities for paths to the goal open up for the player receiving the chips, and consequently, the player has a higher likelihood of being able to grant the chip request while continuing to reach the goal.

Agents Playing	*Player 1*	*Player 2*	*Combined score*
CONTROL VS. CONTROL	6408	6267	12675
TAKER VS. TAKER	6484	6531	13015
GIVER-1 VS. GIVER-1	7565	7280	14845
GIVER-2 VS. GIVER-2	7629	7286	14915

Table 4.2: Individual scores over 100 rounds with $n = 2$

Table 4.2 breaks down the cumulative scores and shows their individual components in the $n = 2$ case, the middle column in Table 4.1. We can see that the score differences between the two players are relatively small, and they are a reflection of the randomness present in the boards since the behavior strategies are exactly the same.

Values of n	RECIPROCATOR	RECIPROCATOR	*Combined score*
0	3961	4049	8010
1	5928	5712	11640
2	7275	6970	14245
3	7906	8214	16120

Table 4.3: Individual scores over 100 rounds, averaged over 3 runs

Table 4.3 shows the individual scores of the RECIPROCATOR agents. One thing that is important to note about the RECIPROCATOR agent is that since its decision-making has a probabilistic element, in different experimental runs on the same boards, RECIPROCATOR will register different scores. The data that we present for RECIPROCATOR is averaged over 3 runs. We experimented with different values of the *generosity factor*, which we mention in Chapter 3, but our data did not show any significant score differences unless we used extreme positive or negative values of "generosity," thereby essentially turning RECIPROCATOR into either GIVER or TAKER.

4.2 Heterogeneous Experiments

The next set of experiments pits the different agent types against each other. As mentioned in the section on experimental setup, the different agents play both "sides" of the setup, even though as we saw in Table 4.2, the data indicates that the scores from each side are similar to each other. In the rest of the tables in this chapter, we average together the players' scores from the different sides of the setup.

4.2.1 Non-reciprocative interactions

Values of n	GIVER-1	GIVER-2	*Combined score*
0	4416	4197	8613
1	6359	5531	11890
2	7997	6998	14995
3	8628	7887	16515

Table 4.4: Individual scores over 100 rounds

Table 4.4 shows the results of the experiments between the two GIVER agent variations. We see GIVER-1 receiving slightly better scores than GIVER-2. The fact that GIVER-2 offers up all its extra chips to the other player whenever it makes a chip request comes at a high cost because of the significant value that the chips have in our scoring function. If the parameters of the scoring function valued chips less, then GIVER-2's scores would be closer to those of its co-player.

GIVER-2's strategy of freely giving its extra chips away helps its co-player significantly because not only do the extra chips have value in the scoring function, but they also sometimes allow the other player to find a path to a previously unreachable goal or find a shorter path. In some situations in which the scoring parameter d is significantly greater than 0, GIVER-2's strategy may even have a concrete short-term benefit.

Values of n	GIVER-2	TAKER	*Combined score*
0	3023	5567	8590
1	5019	6696	11715
2	6224	7861	14082
3	7119	8696	15815

Table 4.5: Individual scores over 100 rounds

Table 4.5 presents data that shows the TAKER agent achieving significantly higher individual scores than the second variation of the GIVER agent. We do not include the scores that the first variation of GIVER received against TAKER because they are almost essentially the same. The point to be made is the same for both GIVER variations: TAKER takes full advantage of his co-player's generosity and rarely grants GIVER's requests, and since the GIVER agents are not reciprocative, they continue to happily agree to TAKER's favor request deals and to give TAKER all the chips he asks for unless giving those chips would cause the GIVER agent to no longer be able to reach the goal.

4.2.2 Reciprocative interactions

The fundamental design principle behind the RECIPROCATOR agent is that it is able to avoid being "exploited" by players like the TAKER agent while at the same time being able to interact cooperatively with more generous players like the GIVER agent. The next two tables detail the actual performance of the RECIPROCATOR agent against the TAKER and GIVER agents respectively.

Values of n	RECIPROCATOR	TAKER	*Combined score*
0	3069	3471	6540
1	5247	5078	10325
2	6265	6500	12765
3	7349	7606	14955

Table 4.6: Individual scores over 100 rounds

If we compare Table 4.6 to Table 4.5, we in fact see that the TAKER agent is unable to take advantage of the RECIPROCATOR agent in the same way that it is able to when it plays against the GIVER agents. For the four values of n, TAKER's mean score is only 5664 when it plays against RECIPROCATOR; whereas, against GIVER, TAKER was able to register a mean score of 7205.

However, there is another interesting point to be made here: though the scores that RECIPROCATOR receives in Table 4.6 are higher than GIVER's in Table 4.5, they are not substantially higher. This is because the biggest contributor to higher scores is being able to reach the goal state more often;

the scoring function is engineered to reward reaching the goal much more than holding on to extra chips.

In fact, with certain values of the d parameter in the scoring function, the GIVER agent would actually receive a higher individual score than the RECIPROCATOR. Table 4.7 shows the minimum d values that for these particular runs would have resulted in a higher individual score for the GIVER agent than for the RECIPROCATOR.

n values	*d values*
0	0.022
1	0.141
2	0.030
3	0.211

Table 4.7: Values of the scoring parameter d that would result in GIVER receiving a higher score than RECIPROCATOR in these runs against TAKER

We begin to see that if there is a high enough d, the best strategy for a given player may actually be to keep giving chips away even if the other player does not give any chips back.

Values of n	RECIPROCATOR	GIVER-2	*Combined score*
0	5412	3798	9210
1	6521	5469	11990
2	7981	6904	14885
3	8473	7952	16425

Table 4.8: Individual scores over 100 rounds

The data in Table 4.8 confirms our expectation that when playing against GIVER, the RECIPROCATOR agent achieves joint scores nearly as high as the two variations of GIVER playing together. We also see that the RECIPROCATOR agent actually has slightly higher individual scores than the GIVER it plays against because during their first interactions, the GIVER agent has not yet built up a reputation with RECIPROCATOR so with a certain probability, RECIPROCATOR may deny some of the GIVER agent's early chip requests.

If we compare the data in Tables 4.7 and 4.8, we see that GIVER's concern with its reputation pays off when playing against a RECIPROCATOR agent: GIVER noticeably outperforms TAKER.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Several conclusions come as a result of our agent design and experimentation. First, we should state our more obvious finding that in CT, for the group, there is a definite advantage associated with trading chips and being generous, and to the extent that the d parameter of the scoring function rewards the individual for the success of the group, generosity can also lead to increased individual rewards.

We conclude that a player's success in 2-player CT is very much determined by her co-player's generosity, and we find that by keeping a balance of favors done and favors received, the RECIPROCATOR agent is able to avoid exploitation by less generous players while at the same time achieving high levels of profitable trading when interacting with more cooperative agents.

Within a controlled, experimental environment, our work verifies and quantifies basic negotiation intuitions that over repeated encounters it is critical to act reciprocally, and we also conclude that if it is known that the co-player pursues a reciprocal strategy, it is better to act like a GIVER than a TAKER.

5.2 Future Work

There are many promising directions in which this work could be extended, and the CT game infrastructure provides a flexible platform for a wide variety of provocative experiments. In this section, we list and describe some research areas in CT that would be natural extensions of our work.

Our experiments were all run with groups of two agents. In these types of interactions, an agent has no choice in his negotiating partner so strategy is limited to evaluating deal proposals and making decisions on whether or not to deal with the other agent. Games with more than two players introduce the variable of deciding *with whom* to deal. In these games, it would seem that an agent's reputation would be even more important because the other players of the game would have choice in their negotiation partners; conceivably, a selfish TAKER style agent could be "ostracized" from all negotiations.

In our work, computer agents play against other computer agents exclusively. Setting up our computer players to play against human test subjects might provide interesting insights about how people make decisions on whether or not to help each other over repeated encounters. Study of human behavior in repeated CT games could also lead to the design of other more sophisticated models of computer play.

Chapter 6

Related Work

Sen and Dutta [8] introduces variants on the basic agent strategies outlined in [7] and described in the introduction. These variants include believing reciprocative, earned-trust based reciprocative, individual lying selfish, and collaborative lying selfish agents. The believing reciprocative agent considers not only its own cost-savings balance, but also the balances that other agents report to it. The earned-trust based reciprocative agents only consider balances that are reported by agents who have "earned trust," i.e. agents who have registered positive balances.

The two selfish agent variants introduced are even more aggressive than the 1996 selfish agent. The individual lying agent reports negative information about other helpful agents to ruin their reputations, and the collaborative lying agent not only attempts to ruin the reputations of other helpful agents but also actively attempts to bolster the reputations of other selfish agents. These variant behavior strategies are important because they highlight the fact that an agent must consider its reputation in the "community" as a whole as opposed to its reputation in just bilateral relationships, especially when agent behavior is constantly changing and adapting.

Other methods have been developed to understand and model the complex social relationships that exist between agents in the real world. Glass and Grosz discuss the concept of "brownie points" as a way to measure social consciousness [3]. The basic idea is that agents who sacrifice their utility in the short-term gain brownie points, and the quantity of brownie points that an agent builds up over a historical time period reflects that agent's contributions to the group as a whole. Brownie points capture what the authors call an agent's "good guy stature in the community," i.e. its reputation, and agents gain utility from the satisfaction of having brownie points.

Extensive work has been done in the fields of economics and political science to study how agents cooperate and interact with each other. For instance, Fudenberg has studied issues of reputation and cooperation in repeated games [2]. He identifies the "fear of retaliation" as a motivator that can have a strong effect on an agent's behavior in a scenario where there is a strong possibility of a long term relationship.

Axelrod's *The Evolution of Cooperation* [1] is a study of strategies that are used by agents as they make decisions about whether to cooperate or not with other agents. Axelrod cites examples from the real world that motivate his research, e.g. senators deciding whether or not to go out of their way to help another senator on a piece of legislation.

Axelrod invited scientists to submit strategies to compete in a prisoner's dilemma tournament. Basically, the game involved rounds of prisoner's dilemma scenarios in which if both players cooperate, each player gains three points, if one player cooperates while the other defects, then the defecting player gains five points while the cooperating "sucker" takes no points, and if both players defect, then they each get only one point.

Axelrod identifies a very important factor which determines the behavior that agents should follow. This factor is how many future rounds the agents are expecting to play. If there is only one round of play, then it will always be in the agent's best interest to defect because no matter what the other player does, the reward for the defecting agent will always be higher if it defects than if it had cooperated. But if there are many rounds of games, i.e. if the senator is expecting to have many future interactions with the other senator, then it is worthwhile to try to build up a relationship which can be mutually beneficial.

The entrant that performed best in the repeated PD tournament was tit-for-tat, one of the simplest of all the strategies. Tit-for-tat's algorithm is that in the first round the agent will cooperate, and then in each subsequent round, it will do exactly what the opposing player did in the previous round.

Many other scientists have identified the tit-for-tat strategy in nature. Milinski (1987) describes tit-for-tat in the behavior of sticklebacks [6]. Milinski studies the behavior of the stickleback fish as it approaches a potential predator. The fish is in the most danger if it approaches alone, but if it is with another "cooperating" fish than it is in less danger. The ideal situation though for a fish is to "defect" and not have to approach the predator at all because the partner fish will have approached and undergone the risks alone. Milinski's research shows that the strategy that the stickleback fish follow in deciding whether or not to cooperate with other fish is exactly tit-for-tat.

Bibliography

[1] Robert Axelrod. *The Evolution of Cooperation.* New York: Basic Books, 1984.

[2] Drew Fudenberg and Eric Maskin. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica*, Vol. 354, No. 3, pages 533-554. May 1986.

[3] Alyssa Glass and Barbara Grosz. "Socially Conscious Decision-Making." In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 217-224. New York: ACM Press, 2000.

[4] Morton Hunt. *Profiles of Social Research: The Scientific Study of Human Interactions.* New York: Russell Sage Foundation, 1985.

[5] James Kahan and Amnon Rapoport. *Theories of Coalition Formation.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1984.

[6] Manfred Milinski. "TIT FOR TAT in sticklebacks and the evolution of cooperation." *Nature*, Vol. 325, pages 433-435. January 29, 1987.

[7] Sandip Sen. "Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents." In *Proceedings of the Second International Conference on Multiagent Systems*, pages 322-329. Menlo Park, CA: AAAI Press, 1996.

[8] Sandip Sen and Partha Dutta. "The Evolution and Stability of Cooperative Traits." In *Proceedings of the First International Joint Conference on Autonomous Agents & Multiagent Systems*, pages 1114-1120. New York: ACM Press, 2002.

Appendix A

Screen Shot of HCI

On the next page, we include a screen shot of the interface through which people can play the CT game. The screen shot helps to visualize the CT board and more easily understand the basics of the game. In this particular screen shot, we see that the goal state is the green square in the lower right corner that has a yellow border. This is the interface for the player represented by the smiley face who has the 121 ID. In the upper left corner of the screen shot, we can see that the player has 5 blue chips, 2 green chips, and 2 red chips. Beneath the chip information, we see that a chip request message has been sent from the smiley face player to the longhorn player even though with its current chip set, the smiley face player can reach the goal state.