How to Play Well in Non-zero Sum Games: Some Lessons from Generalized Traveler's Dilemma

Predrag T. Tošić and Philip Dasler

Department of Computer Science,
University of Houston, Houston, Texas, USA
{pedja.tosic,philip.dasler}@gmail.com

Abstract. We are interested in two-person games whose structure is far from zero-sum. We study the iterated *Traveler's Dilemma* (TD) which is a two-player, non-zero sum game that, depending on the exact values of its critical parameters, may offer plenty of incentives for cooperation. We first briefly summarize the results of a round-robin tournament with 36 competing strategies that was motivated by the work by Axelrod *et al.* on the iterated Prisoner's Dilemma. We then generalize the "default" version of Iterated TD with respect to two important game parameters, the bonus value and the "granularity" of the allowable bids. We analytically show the impact of the ratio of these two parameters on the game structure. Third, we re-run the 36-player round-robin tournament and investigate how varying the bonus-to-granularity ratio affects relative performances of various types of strategies in the tournament. We draw some conclusions based on those results and outline some promising ways forward in further investigating games whose structures seem to defy the prescriptions of classical game theory.

1 Introduction

Game theory is important to AI and multi-agent systems research because it provides mathematical foundations for modeling interactions among, in general, *self-interested rational agents* that may need to combine competition and cooperation with each other in order to meet their individual objectives [16,18,21]. An example of such interactions is the iterated *Prisoner's Dilemma* (PD) [1,2], a classical two-person non-zero-sum game that has been extensively studied by psychologists, sociologists, economists, political scientists, applied mathematicians and computer scientists. In this paper, we study an interesting 2-player game known as the (iterated) *Traveler's Dilemma* [6,7,11,15]. The Traveler's Dilemma (TD) is a non-zero sum two-person game in which each player has a large number of possible actions or moves. In the iterated TD, this means many possible actions per round and thus, for games of many rounds, an astronomic number of possible strategies overall. We are interested in the Iterated TD because its structure defies the usual prescriptions of classical game theory insofar as what constitutes "optimal" play.

We first define Traveler's Dilemma, briefly motivate its relevance and survey the prior art. Our study of Iterated TD begins with an experimental analysis of the "baseline" variant of the game using a round-robin tournament. We then analyze how the

N. Zhong et al. (Eds.): AMT 2011, LNCS 6890, pp. 300–311, 2011.

game structure changes as the game's critical parameters are varied, followed by a further experimental study of the relative performance of various strategies in our round-robin tournament with respect to alternative values of those parameters. The structural aspects of our primary interest are *Nash equilibria* and *Pareto-optimal* strategy pairs (e.g., [14]). We draw some conclusions based on our extensive experimentation and analyzes, and briefly discuss some promising ways forward on Iterated TD and other "far-from-zero-sum" iterated two-player games.

2 The Traveler's Dilemma

The Traveler's Dilemma was originally introduced in [5]. The motivation behind the game was to show the limitations of classical game theory [13], and in particular the notions of *individual rationality* that stem from game-theoretic notions of "optimal play" based on *Nash equilibria* [5,4,21]. The original version of TD, which we will treat as the "default" variant of this game, is described with the following parable:

An airline loses two suitcases belonging to two different travelers. Both suitcases happen to be identical and contain identical items. The airline is liable for a maximum of $100 per suitcase. The two travelers are separated so that they cannot communicate with each other, and asked to declare the value of their lost suitcase and write down (i.e., bid) a value between $2 and $100. If both claim the same value, the airline will reimburse each traveler the declared amount. However, if one traveler declares a smaller value than the other, this smaller number will be taken as the true dollar valuation, and each traveler will receive that amount along with a bonus/malus: $2 extra will be paid to the traveler who declared the lower value and a $2 deduction will be taken from the person who bid the higher amount. So, what value should a rational traveler (who wants to maximize the amount she is reimbursed) declare?

A tacit assumption in the default formulation of TD is that the bids, that is, the dollar amounts that the two players can write down, must be integers. That is, the *granularity parameter* (see below) is $1, as this amount is the smallest possible difference between two different bids. This default TD game has several interesting properties. Perhaps the most striking is that its unique Nash equilibrium, the action pair $(p, q) = (\$2, \$2)$, is actually rather bad for both players, assuming the level of players' well-being is proportional to the dollar amount they receive. This choice of actions results in a very low payoff for each player, only slightly above the absolute worst possible (which is $0); moreover, it minimizes *social welfare*, which we define to be simply the sum of the two players' individual payoffs. Yet, it has been argued [5,7,10] that a perfectly rational player, according to classical game theory, would "reason through" and converge to choosing the lowest possible value, $2. Given that the TD game is symmetric, each player would reason along the same lines and, once selecting $2, would not deviate from it (since unilaterally deviating from a Nash equilibrium is supposed to result in decreasing one's own payoff). However, the non-equilibrium pair of strategies $(\$100, \$100)$ results in each player earning $100, very near the best possible individual payoff for each player. Hence, the early studies of TD concluded that this game demonstrates a woeful inadequacy of classical, Nash (or other similar notion of) equilibrium based, game theory. However, it has been experimentally shown that humans (both game theory experts and laymen) tend to play far from the equilibrium, at or close to the maximum possible bid (\$100 in the default TD case), and therefore fare much better than if they followed the classical approach [6].

So, TD has a unique Nash equilibrium, yet the corresponding strategies result in nearly as low a payoff as one can get. Adopting one of the alternative notions of game equilibrium found in the "mainstream" game theory literature does not appear to help, either. For example, it is argued in [11] that the action pair $(\$2, \$2)$ is also the game's only *evolutionary equilibrium*. Similarly, seeking *sub-game perfect equilibria* (SGPE) [14] of Iterated TD would not result in what would intuitively constitute optimal or close to optimal play, either: the set of a game's SGPEs is a subset of that game's full set of Nash equilibria in mixed strategies. We also note that Iterated TD is structurally rather different from the Centipede Game [14]; in particular, the latter has multiple pure strategy Nash equilibria (NE) and infinitely many NE in mixed strategies, whereas the game on which we have focused has a unique (pure strategy) NE and no additional mixed strategy equilibria. Furthermore, the game's only stable strategy pair is nowhere close to being *Pareto optimal*: there are many obvious ways of making both players much better off than if they play the equilibrium strategies. In particular, while neither stable nor an equilibrium in any sense of these terms, the action pair $(\$100, \$100)$ is the unique strategy pair that maximizes social welfare and is, in particular, Pareto optimal.

3 The Iterated TD Tournament

Our Iterated Traveler's Dilemma tournament is similar to Axelrod's Iterated Prisoner's Dilemma tournament [3]. It is a round-robin tournament where each strategy plays against every other strategy as follows: each agent plays N matches against each other agent and its own "twin". A match consists of T rounds. In order to have statistically significant results (esp. given that many of our strategies involve randomization in various ways), we have selected $N = 100$ and $T = 1000$. In each round, both agents must select a valid bid. Thus, the *action space* of an agent in the tournament is $A = \{2, 3, \ldots, 100\}$. The method in which an agent chooses its next action for all possible histories of previous rounds is known as a strategy. A *valid strategy* is a function S that maps some set of inputs to an action, $S : \cdot \to A$. In general, the input may include the entire history of prior play, or an appropriate summary of the past histories.

The participants in the tournament are the set of strategies that play one-against-one matches with each other. Let C denote the set of agents competing in the tournament. Agents' actions are defined as follows:

$x_t =$ the bid agent x makes on round t;
$x_{nt} =$ the bid agent x makes on round t of match n.

The *reward function* describes agent payoffs. *Reward per round*, $R : A \times A \to \mathbb{Z} \in [0, 101]$, for action α against action β, where $\alpha, \beta \in A$, is defined as $R(\alpha, \beta) = min(\alpha, \beta) + 2 \cdot sgn(\beta - \alpha)$, where $sgn(x)$ is the usual *sign* function. The total reward $M : S \times S \to \mathbb{R}$ received by agent x in a match against y is defined as

$$M(x, y) = \sum_{t=1}^{T} R(x_t,\, y_t)$$

In a sequence of matches, the reward received by agent x in the n^{th} match against y is denoted as $M_n(x, y)$.

We next describe the strategies participating in our round-robin tournament. In order to make apple-to-apple comparisons, we utilize the same strategies used in [8], ranging from rather simplistic to relatively complex. We outline the nine distinct classes of strategies. For a more in depth description see [8].

Randoms: The first, and simplest, class of strategies play a random value, uniformly distributed across a given interval. We have implemented two instances using the following intervals: $[2, 100]$ and $[99, 100]$.

Simpletons: The second very simple class of strategies are the ones that choose the exact same dollar value in every round. The values we used in the tournament were $x_t = 2$ (the lowest possible), $x_t = 51$ ("median"), $x_t = 99$ (slightly below maximal possible; would result in maximal individual payoff should the opponent consistently play the highest possible action, which is $100), and $x_t = 100$ (the highest possible).

Tit-for-Tat-in-spirit: The next class of strategies are those that can be viewed as *Tit-for-Tat-in-spirit*, where Tit-for-Tat is the famous name for a very simple, yet very effective, strategy for the classical iterated prisoner's dilemma [1,2,3,17]. The idea behind *Tit-for-Tat* (TFT) is simple: cooperate on the first round, then do exactly what your opponent did on the previous round. In the iterated TD, each agent has many actions at his disposal. In general, playing high values can be reasonably considered as an approximate equivalent of "cooperating", whereas playing low values is an analogue of "defecting". Following this basic intuition, we have defined several Tit-for-Tat-like strategies for the Iterated TD. These strategies can be roughly grouped into two categories. The *simple* TFT strategies bid some value ϵ below the bid made by the opponent in the last round, where $\epsilon \in \{1, 2\}$. The *predictive* TFT strategies compare whether their last bid was lower than, equal to, or higher than that of their opponent. Then a bid is made similar to the simple TFT strategy, i.e. some value ϵ below the bid made by competitor c in the last round. The key distinction between simple and predictive TFTs is that, in case of the latter, a bid can be made relative to either the opponent's last bid or one's own previous bid. In essence, predictive TFTs try to predict the opponent's next bid based on the previous round(s) and, given that prediction, attempt to outsmart the opponent by bidding "one below" the opponent's expected next bid. Details can be found in [8,9].

Mixed: The mixed strategies combine up to three *pure strategies* probabilistically. In a mixed strategy, a pure strategy $\sigma \in C$ is selected from one of the other strategies defined in the competition in each round according to a probability distribution. Once a pure strategy for the given round has been selected, the value that σ would bid at time step t is bid. We chose to use only mixtures of the TFT, Simpleton, and Random strategies. This allowed for greater transparency when attempting to understand the causes of a particular strategy's performance.

Buckets - Deterministic: These strategies keep count of each bid by the opponent in an array of "buckets". The bucket that is most full (i.e., the value bid most often) is used as the predicted value, with ties being broken by one of the following methods: the highest valued bucket wins, the lowest valued bucket wins, a random bucket wins, and the newest tied bucket wins. The strategy then bids the next lowest value ("one under" [8]) below the predicted value. An instance of each tie breaking method competed in the tournament.

Buckets - Probability Mass Function: This strategy class counts instances of the opponent's bids and uses them to pick one's own next bid. Rather than always picking the value most often bid (see above), in this case the buckets are used to define a probability distribution according to which a prediction is randomly selected. Values in the buckets decay over time in order to emphasize newer data over old and we have set a retention rate ($0 \leq \gamma \leq 1$) to determine the rate at which this decay occurs. We have entered into our tournament several instances of this strategy using the following rate of retention values γ: 1.0, 0.8, 0.5, and 0.2. The agent then bids the next lowest value below the predicted value. Note that the "bucket" strategies based on probability mass buckets are quite similar to a learning model in [7].

Simple Trend: This strategy looks at the previous k time steps, creates a line of best fit on the rewards earned, and compares its slope to a threshold Θ. If the trend has a positive slope greater than Θ, then the agent will continue to play the same bid it has been as the rewards are increasing. If the slope is negative and $|slope| > \Theta$, then the system is trending toward the Nash equilibrium and, thus, the smaller rewards. In this case, the agent will attempt to maximize social welfare and play 100. Otherwise, the system of bidding and payouts is relatively stable and the agent will play the "one under" strategy. We have implemented instances of this strategy with an arbitrary Θ of 0.5 and the following values of k: 3, 10, and 25.

Q-learning: This strategy uses a learning rate α to emphasize new information and a discount rate γ to emphasize future gains. In particular, the learners in our tournament are simple implementations of *Q-learning* [20,19] as a way of predicting the best action at time $(t + 1)$ based on the action selections and payoffs at times $[1, ..., t]$. This is similar to the Friend-or-Foe Q-learning method [12] without the limitation of having to classify the allegiance of one's opponent. Details on our implementation of Q-learning can be found in [8,9].

Zeuthen Strategies: A Zeuthen Strategy [22] calculates the level of risk of each agent, and makes *concessions* accordingly. Risk is the ratio of loss from accepting the opponent's proposal to the loss of forcing the conflict deal (the deal made when no acceptable proposal can be found). While ITD is not a strict negotiation, we treat each bid (i.e. x_t and y_t) to be proposals. If $x_t = i$, then this can be viewed as x implicitly proposing (i, i) as the next action pair. We consider the *conflict deal* [22] to be the Nash Equilibrium at $(\$2, \$2)$. Given the proposals of each agent, a risk comparison is done. An agent will continue to make the same proposal while its risk is greater than or equal its opponent's. Otherwise, the agent will make the *minimal sufficient concession*, i.e. the agent adjusts its proposal so that (i) the agent's risk is higher than that of its opponent and (ii) the opponent's utility increases as little as possible. Due to the peculiar structure of TD, it is possible for a concession to actually lead to a *loss of utility* for the opponent. We have therefore implemented two Zeuthen-based strategies: one that allows negative "concessions" and one that does not.

Elaborate experimentation and detailed analysis of the default version of Iterated TD can be found in [8,9]. Analyzes in those two papers are performed with respect to four distinct utility metrics. The first, U1, treats the actual dollar amount as the payoff to the agent. Most of the prior literature on TD and the Iterated TD generally considers only some variant of this, "bottom line" metric. In contrast, the second metric in [8,9], U2, is a "pairwise victory" metric: an agent strives to beat its opponent, regardless of the actual dollar amount she receives. Finally, [8,9] consider two additional metrics, U3 and U'3, that attempt to capture both the actual payoff (dollar amount) that an agent has achieved, and the "opportunity lost" for not acting differently (in particular, due to not knowing what the other agent would do). Both U3 and U'3 attempt to quantify the difference between how much an agent wins vs. how much an *omniscient* agent (one that always correctly predicts the other agent's bid) would be able to win.

Due to space constraints, in this paper we focus our performance analyzes of the default version of TD (Section 4) and its generalizations (Sections 5 and 6) solely based on metric U1. For details on individual strategy performances w.r.t. the other three metrics, see [9]. Metric U1 is essentially the sum of all payoffs (dollar amounts) gained by an agent over all rounds and all matches, normalized by the total number of rounds played and the maximum allowable reward:

$$U_1(x) = \frac{1}{|C|} \sum_{j \in C} \left[\frac{1}{max(R)NT} \sum_{n=1}^{N} M_n(x,j) \right]$$

where $max(R)$ is the maximum possible reward given in one round, N is the number of matches played between each pair of competitors, T is the number of rounds per each match, and $|C|$ is the number of competitors in the tournament.

4 Results and Analysis for Default Iterated TD

The Traveler's Dilemma Tournament with which we have experimented involves a total of 36 competitors (i.e., distinct strategies). Each competitor plays each other competitor 100 times. Each match is played for 1000 rounds. The following summarizes our main findings; details can be found in [8].

First, the top three performers in our tournament turn out to be three "dumb" strategies that always bid high values; interestingly enough, the strategy which always bids the maximum value and the strategy which always bids one under the maximum value are both outperformed by the strategy which randomly alternates between the two. Second, performances of Tit-for-Tat-based strategies, as it turns out, vary greatly depending on the details of bid prediction method. So, while some of the relatively complex TFT-based strategies that, in particular, (i) make a nontrivial model of the other agent's behavior and (ii) "mix in" some randomization, are close to the top, other TFT-based strategies show either fairly mediocre performances or, in some cases, are among the very worst performers. Third, "simple trenders" and negotiation-inspired Zeuthen that disallows negative concessions (which we call Zeuthen-Positive for short) turn out to be clearly the best performers among adaptable strategies, simple as they (esp. the trenders) are. Some of the bucket-based strategies perform quite well, as well.

Another major finding, fairly surprising, is relative mediocrity of the Q-learning based strategies: none of them excels. On the other hand, the adaptability of Q-learning based strategies ensures that they do not do too badly overall, either. It is also worth noting that the choice of the learning rate seems to make very little difference: all three Q-learning based strategies show similar performance, and, hence, end up ranked relatively near each other (see the ranking tables in [8,9]). Due to space constraints, we leave out details and move on to *Generalized* Iterated TD.

5 Generalized Iterated TD

We now generalize the game with respect to the two most important parameters: the bonus, denoted by δ, and the granularity of the bids, g. Specifically, we are interested in (i) how the game structure changes and (ii) how the relative performances of strategies in our round-robin tournament change as a function of the relationship between these two parameters. The focus of this section is on (i) and the next section summarizes our experimental findings on (ii).

The most important aspects of the game structure for us are Nash equilibria and Pareto optimal pairs of strategies, as a function of δ and g. We denote the lower bound of allowable bids by m, and the upper bound by M. Therefore, the "default" version of Iterated TD has these parameter values set to $\delta = 2$, $g = 1$, $m = 2$ and $M = 100$.

We make two additional assumptions. One, much of the prior work on the TD assumes that $m - \delta \geq 0$, i.e., that there is no possibility of a player ever receiving a negative payoff (e.g., [15]). We also adopt that assumption. However, the analysis in this section holds whether negative payoffs are possible or not. While the possibility of "losing money" may have a practical psychological impact on the decision making of human agents who get involved in TD-like scenarios, insofar as the game-theoretical structural aspects of our interest are concerned, this possibility of negative payoffs is immaterial. Two, we assume that *bonus* equals *malus*, i.e., values of the "honesty reward" and the overbidding penalty are the same, and both equal δ. Due to space constraints, we just summarize our analysis with respect to the basic relationship between δ and g. For the bid pair (u, v), the corresponding payoff pair is

$$\begin{array}{ll} (u-\delta, u+\delta), & \text{if} u < v; \\ (v+\delta, v-\delta), & \text{if} u > v; \\ (x, x), & \text{if} u = v = x. \end{array}$$

If two bids u and v are not equal, then one bid is greater than the other by an integer multiple of granularity g.

Case 1: $\delta > g$

This is the general scenario of which the "default" TD with $\delta = 2$ and $g = 1$ is a special case. The main properties of the default TD carry over to the general case:

Lemma 1. *When bonus is greater than granularity of the bids, the only Nash equilibrium of TD is the bid pair* (m, m).

Proof: When $u = v = x > m$, then the pair $(u, v) = (x, x)$ is not a Nash equilibrium, since if either player unilaterally changes to $x - g$ (and the other still bids x), then the player who changed his bid will receive a higher payoff, $x-g+\delta > x$. Clearly, the only strategy pair of the form (x, x) where one cannot bid lower is (m, m), and it is easy to verify that, just like in the default case with $m = 2$, this pair is a Nash equilibrium (NE) for any choice of parameter values δ, g, m and M (as long as $\delta > g$). When $u < v$, the second player would fare better if unilaterally changing to $v = u$ (or, if possible, to $u - g$). By game's symmetry, the case $u > v$ similarly cannot result in a NE. Hence, $(u, v) = (m, m)$ is the unique pure strategy NE.

Let's consider Pareto optimal pairs of bids and social welfare. Since bonus is assumed to always equal malus, the sum of payoffs always equals $2 \cdot min\{u, v\}$ where u, v are bids. Consequently, the unique social welfare maximizer is the bid pair (M, M), leading to a cumulative payoff of $2M$. Maximum individual payoffs are obtained by the lower bidders in bid pairs $(M - g, M)$ and $(M, M - g)$, since $M - g + \delta > M$, and hence these bid pairs are also Pareto optimal.

Case 2: $\delta < g$
In this case, $x - \delta + g < x$ where $x = min\{u, v\}$, hence now there is no incentive to bid "one (or, more generally, g) under" one's opponent, which in Case 1 was the individually optimal strategy for an omniscient player who knows what the other player would bid (this strategy works as long as the other agent bids $v > m$). One immediate implication is that

Lemma 2. *When bonus is strictly smaller than granularity ($\delta < g$), then (i) each strategy pair of the form $(u, v) = (x, x)$ is a Nash equilibrium and (ii) the Generalized TD has* the unique *Pareto optimal pair of bids, namely* $(u, v) = (M, M)$.

Proof sketch: If the first player deviates and bids any value u such that $u > v = x$, she will receive $x - \delta < x$, where x is what she'd get had she stayed with $u = x$. If the first player underbids $u < v = x$, then she receives $u + \delta$. She cannot bid less than x yet higher than $x - g$, so by bidding $x - g$ her payoff is $x - g + \delta < x$. Hence, it's impossible to unilaterally deviate from (x, x) and benefit. It can be readily established that no pair (u, v) with $u \neq v$ can be a NE. Hence the set of Nash equilibria is precisely the set of bid pairs (x, x) with $x \in \{m, m + g, m + 2g, \ldots, M - g, M\}$. Each of these pairs is a *strict* (i.e., strong) NE.

What is the Pareto optimality structure of Generalized TD with $g > \delta$? It turns out that strategy pairs $(M - g, M)$ and $(M, M - g)$ are no longer Pareto optimal, as the player who is bidding lower now gets only $M-g+\delta < M$. In particular, the maximum individual payoff in this case is M, and the pair (M, M) is the unique Pareto optimal pair that also maximizes social welfare. No other pair (u, v) is Pareto optimal, as one (or both) player(s) can be made better off (e.g., by adopting $u = v = M$) without making anyone worse off, since no agent can possibly win more than \$M regardless of what he or the other agent does. These observations establish that, when $\delta < g$, Generalized TD has a unique Pareto-optimal pair of bids, namely $(u, v) = (M, M)$.

Case 3: $\delta = g$
In this borderline case, $x - g + \delta = x$. Hence, betting "one (that is, g) under" the other agent is just as good for the lower bid agent as bidding exactly the same as the other agent; of course, the other agent is worse off in the first case where he's been underbid than in the second, where both agents make equal bids. Consequently,

Lemma 3. *When $\delta = g$, each $(u, v) = (x, x)$ is a weak Nash equilibrium, and the unique Pareto optimal pair (which also maximizes social welfare) is $(u, v) = (M, M)$.*

The proof (omitted) is along the lines of arguments establishing Lemmas 1 and 2.

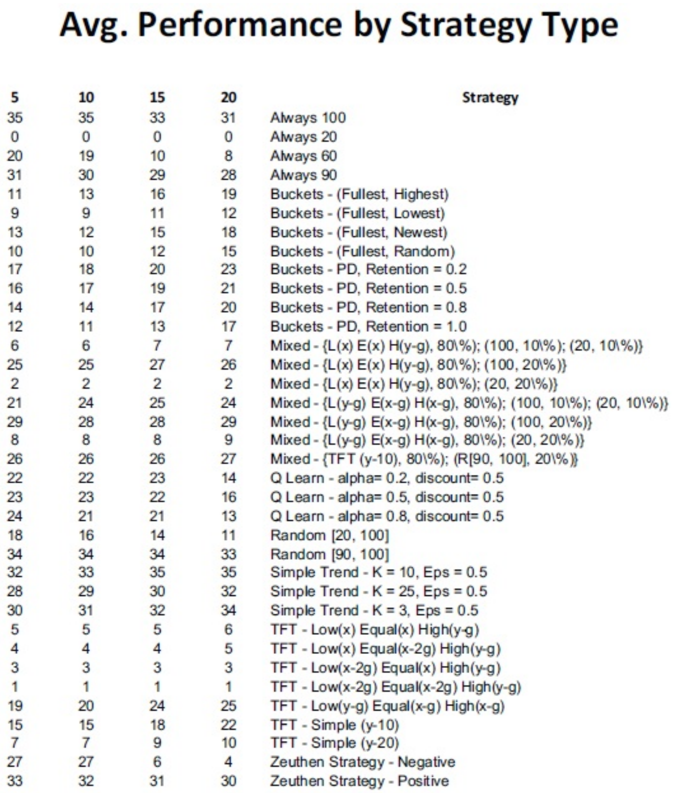
6 Experimental Study of Generalized ITD

In order to explore the performance change of strategies with respect to the relative size of δ (and its relationship to g) we have run the same tournament as before but with some important changes to the structure of the game itself. Rather than use the original definition of the Traveler's Dilemma (i.e., $g = 1$, $\delta = 2$, $m = 2$, and $M = 100$), we have run our tournament with the following attributes: $g = 10$, $\delta = \{5, 10, 15, 20\}$, $m = 20$ and $M = 100$.

By selecting a somewhat arbitrary value of $g = 10$, we have sufficient room to fully explore the dynamics of $\delta < g$, $\delta = g$, and $\delta > g$. In particular, the tournament was run four separate times, once for each value of δ listed above. While there is no difference between the δ's of 15 and 20 with respect to the $g : \delta$ relationship, it does allow us to investigate the effect of a greater penalty on strategies that are less careful (such as, e.g., the simpleton always bidding M) while keeping a uniform distribution w.r.t. δ among our test cases. We use $m = max(\delta)$ in order to constrain our tournaments to the previously stated assumption that $m - \delta \geq 0$ (see discussion in [8]). Also, we have kept M at the original value for the TD as there seemed to be little to be gained from altering it.

All strategies entered into the default tournament competed in this generalized tournament as well, the only difference being that the strategies were generalized to accommodate the new action space $A = \{m, m+g, m+2g, \ldots, M-g, M\}$. Our expectation was that the performance of those strategies which are more careful not to overbid their opponent will increase as δ does. Conversely, with the penalty becoming harsher for overbidding, strategies that wantonly bid high should show a decrease in their performance. We use the ranking of the strategy after each tournament (for $\delta = 5, 10, 15$, and 20) as a basis to measure its change in performance. However, as it seems more intuitive that a higher "score" equates to better performance, we are actually using the complement of the ranking, i.e., the number of competitors that did worse than this strategy. *Inverse rank* means, how many other strategies a given strategy beats w.r.t. metric U1; hence, the top performer has the inverse rank of $n - 1 = 35$. We then create a line of best fit across the four data points (one for each value of δ tested) for each strategy. If the slope of this line is positive, then the strategy's performance trends toward improvement as the bonus is increased. On the other hand, if this slope is negative, this is an indication that a higher bonus leads to a lower payout for this strategy. Finally, since our interest lies with whether or not a general classification of strategies (i.e., random, simple trenders, Q-learners, etc.) does better or worse with a change in δ, we average the rank complement across all strategies in a given group rather than looking at the individual performance. This gives us a better picture of how the given class of strategies does independently of whether or not the parameters have been optimally set.

Using these slope magnitudes, we find that our predictions generally appear to be correct – with a couple of exceptions (i.e., surprises). We find that the Random and



Avg. Performance by Strategy Type

5	10	15	20	Strategy
35	35	33	31	Always 100
0	0	0	0	Always 20
20	19	10	8	Always 60
31	30	29	28	Always 90
11	13	16	19	Buckets - (Fullest, Highest)
9	9	11	12	Buckets - (Fullest, Lowest)
13	12	15	18	Buckets - (Fullest, Newest)
10	10	12	15	Buckets - (Fullest, Random)
17	18	20	23	Buckets - PD, Retention = 0.2
16	17	19	21	Buckets - PD, Retention = 0.5
14	14	17	20	Buckets - PD, Retention = 0.8
12	11	13	17	Buckets - PD, Retention = 1.0
6	6	7	7	Mixed - {L(x) E(x) H(y-g), 80\%); (100, 10\%); (20, 10\%)}
25	25	27	26	Mixed - {L(x) E(x) H(y-g), 80\%); (100, 20\%)}
2	2	2	2	Mixed - {L(x) E(x) H(y-g), 80\%); (20, 20\%)}
21	24	25	24	Mixed - {L(y-g) E(x-g) H(x-g), 80\%); (100, 10\%); (20, 10\%)}
29	28	28	29	Mixed - {L(y-g) E(x-g) H(x-g), 80\%); (100, 20\%)}
8	8	8	9	Mixed - {L(y-g) E(x-g) H(x-g), 80\%); (20, 20\%)}
26	26	26	27	Mixed - {TFT (y-10), 80\%); (R[90, 100], 20\%)}
22	22	23	14	Q Learn - alpha= 0.2, discount= 0.5
23	23	22	16	Q Learn - alpha= 0.5, discount= 0.5
24	21	21	13	Q Learn - alpha= 0.8, discount= 0.5
18	16	14	11	Random [20, 100]
34	34	34	33	Random [90, 100]
32	33	35	35	Simple Trend - K = 10, Eps = 0.5
28	29	30	32	Simple Trend - K = 25, Eps = 0.5
30	31	32	34	Simple Trend - K = 3, Eps = 0.5
5	5	5	6	TFT - Low(x) Equal(x) High(y-g)
4	4	4	5	TFT - Low(x) Equal(x-2g) High(y-g)
3	3	3	3	TFT - Low(x-2g) Equal(x) High(y-g)
1	1	1	1	TFT - Low(x-2g) Equal(x-2g) High(y-g)
19	20	24	25	TFT - Low(y-g) Equal(x-g) High(x-g)
15	15	18	22	TFT - Simple (y-10)
7	7	9	10	TFT - Simple (y-20)
27	27	6	4	Zeuthen Strategy - Negative
33	32	31	30	Zeuthen Strategy - Positive

Fig. 1. Summary of performances of each of the 36 strategies as the bonus δ is varied (for fixed granularity g = 10)

Simpleton strategies both exhibit a significant decrease in performance as δ is increased. In contrast, the Bucket strategies show relative performance improvements as δ is increased. Though the buckets are quite simplistic in the way that they predict their opponents' future bids, it appears that this is "good enough" to take advantage of the increasing δ, as we generally expected.

However, not all results are entirely in-line with our intuition and theory-based predictions. In particular, we again observe surprisingly mediocre performance from the Q-Learning strategies; moreover, Q-learners' performance can be seen to generally decrease as δ increases. A possible explanation is that this is due to an over-coarsening of the state/action space to accommodate resource constraints (see [8] for details on our implementation of Q-learning). Assuming this basic intuition is correct, we still need to fully understand why this over-coarsening has considerably stronger negative impact on Q-learners' overall performance for high bonus values as compared to their performance for low values of δ.

Interestingly, some of the TFT-based strategies tend to perform significantly better with higher values of δ, despite doing little to model their opponents. In contrast, other TFT strategies – especially those that perform poorly in the tournament – show the identical or almost identical relative performance (consistently poor) for all four values of δ. The Positive Zeuthen strategy does very well overall, and exhibits only a modest deterioration in performance as δ is increased. In contrast, once δ becomes greater than g, the performance of Negative Zeuthen drops catastrophically. Last but not least, the Simple Trend and Mixed strategies show little change as a function of $\delta : g$ ratio. We expected simple trenders to be able to track the opponent's bids well regardless of the $\delta : g$ ratio, and this theoretical prediction has generally been validated by our experiments. On the other hand, the catastrophic performance drop of Negative Zeuthen once δ becomes greater than g (by far the most dramatic such change as a function of $\delta : g$ observed to date) is a phenomenon we are still trying to fully understand and provide a comprehensive account for.

The plot in *Figure 2* summarizes how given strategies are affected, insofar as their performance w.r.t. the metric U1, as a function of varying bonus δ (while the granularity $g = 10$ is held fixed throughout).

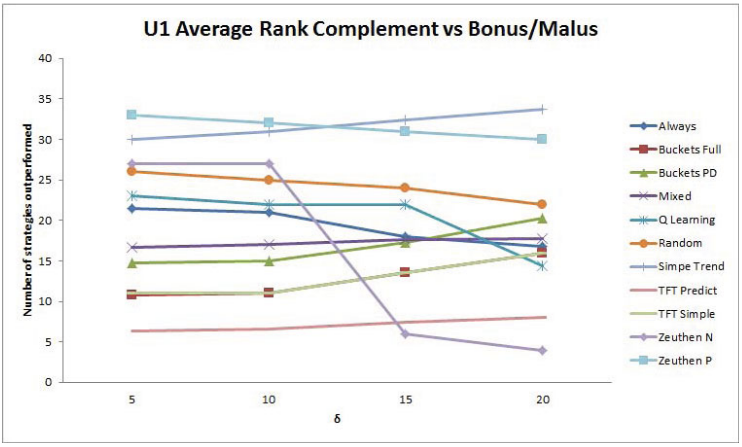


Fig. 2. Relative performances of different strategies as a function of δ

7 Conclusions and Future Work

We study the generalized Iterated Traveler's Dilemma two-player non-zero sum game. We analyze several variants of this game by designing, implementing and analyzing a round robin tournament with 36 participating strategies. Our detailed performance analysis of various strategies with respect to the "bottom-line" metric (directly proportional to the dollar amount earned) has corroborated that, for a game whose structure is far from zero-sum, the traditional game-theoretic notions of rationality and optimality based on the concept of Nash (or similar kinds of) equilibria turn out to be rather unsatisfactory. Our analysis also raises several interesting questions, among which we are particularly keen to further investigate (i) to what extent simple models of learning can be expected to help performance; and (ii) to what extent complex models of the other agent really help an agent increase its payoff in the iterated play. We hope to address these and several other open questions and report new results based on larger, more complex strategy sets in the near future.

References

1. Axelrod, R.: Effective choice in the prisoner's dilemma. Journal of Conflict Resolution 24(1), 3–25 (1980)
2. Axelrod, R.: The evolution of cooperation. Science 211(4489), 1390–1396 (1981)
3. Axelrod, R.: The evolution of cooperation. Basic Books (2006)
4. Basu, K.: The traveler's dilemma. Scientific American Magazine (June 2007)
5. Basu, K.: The traveler's dilemma: Paradoxes of rationality in game theory. The American Economic Review 84(2), 391–395 (1994)
6. Becker, T., Carter, M., Naeve, J.: Experts playing the traveler's dilemma, Department of Economics, University of Hohenheim, Germany (January 2005)
7. Capra, C.M., Goeree, J.K., Gmez, R., Holt, C.A.: Anomalous behavior in a traveler's dilemma? The American Economic Review 89(3), 678–690 (1999)
8. Dasler, P., Tosic, P.: The iterated traveler's dilemma: Finding good strategies in games with bad structure: Preliminary results and analysis. In: Proc of the 8th Euro. Workshop on Multi-Agent Systems, EUMAS 2010 (December 2010)
9. Dasler, P., Tosic, P.: Playing challenging iterated two-person games well: A case study on iterated travelers dilemma. In: Proc. of WorldComp. Foundations of Computer Science FCS 2011 (to appear, July 2011)
10. Goeree, J.K., Holt, C.A.: Ten little treasures of game theory and ten intuitive contradictions. The American Economic Review 91(5), 1402–1422 (2001)
11. Land, S., van Neerbos, J., Havinga, T.: Analyzing the traveler's dilemma Multi-Agent systems project (2008), `http://www.ai.rug.nl/mas/finishedprojects/2008/JoelSanderTim/index.html`
12. Littman, M.L.: Friend-or-Foe q-learning in General-Sum games. In: Proc. of the 18th Int'l Conf. on Machine Learning, pp. 322–328. Morgan Kaufmann Publishers Inc., San Francisco (2001)
13. Neumann, J.V., Morgenstern, O.: Theory of games and economic behavior. Princeton University Press, Princeton (1944)
14. Osborne, M.: An introduction to game theory. Oxford University Press, New York (2004)
15. Pace, M.: How a genetic algorithm learns to play traveler's dilemma by choosing dominated strategies to achieve greater payoffs. In: Proc. of the 5th International Conference on Computational Intelligence and Games, pp. 194–200 (2009)
16. Parsons, S., Wooldridge, M.: Game theory and decision theory in Multi-Agent systems. Autonomous Agents and Multi-Agent Systems 5, 243–254 (2002)
17. Rapoport, A., Chammah, A.M.: Prisoner's Dilemma. Univ. of Michigan Press (December 1965)
18. Rosenschein, J.S., Zlotkin, G.: Rules of encounter: designing conventions for automated negotiation among computers. MIT Press, Cambridge (1994)
19. Watkins, C.: Learning from delayed rewards. Ph.D. thesis, University of London, King's College (United Kingdom), England (1989)
20. Watkins, C., Dayan, P.: Q-learning. Machine Learning 8(3-4), 279–292 (1992)
21. Wooldridge, M.: An Introduction to MultiAgent Systems. John Wiley and Sons, Chichester (2009)
22. Zeuthen, F.F.: Problems of monopoly and economic warfare / by F. Zeuthen ; with a preface by Joseph A. Schumpeter. Routledge and K. Paul, London (1967); first published 1930 by George Routledge & Sons Ltd