On Finding and Learning Effective Strategies for Complex Non-Zero-Sum Repeated Games

Predrag T. Tošić
University of Houston
Houston, Texas, USA

Philip C. Dasler
University of Maryland College Park
College Park, Maryland, USA

Carlos Ordonez
University of Houston
Houston, Texas, USA

***Abstract*—We study complex non-zero-sum iterated two-player games, more specifically, various strategies and their performances in *iterated traveler's dilemma* (ITD). We focus on the relative performances of several types of parameterized strategies, where each such strategy type corresponds to a particular "philosophy" on how to best predict opponent's future behavior and/or entice the opponent to alter its behavior. We are particularly interested in adaptable, learning and/or evolving strategies that try to predict the future behavior of the other player in order to optimize their own behavior in the long run. We also study strategies that strive to minimize risk, as risk minimization has been recently suggested to be the appropriate paradigm for ITD and other complex games that have posed difficulties to classical game theory. We share the key insights from an elaborate round-robin tournament that we have implemented and analyzed. We draw some conclusions on what kinds of adaptability and models of the other player's behavior seem to be most effective in the long run. Lastly, we indicate some promising ways forward toward a better understanding of learning how to play complex iterated games well.**

***Keywords*-game theory; iterated traveler's dilemma**

I. Introduction

Game theory provides mathematical foundations for modeling interactions among, in general, *self-interested* autonomous agents that may need to combine competition and cooperation in non-trivial ways in order to meet their objectives [19], [20], [24]. A classical example of such interactions is the *(iterated) Prisoner's Dilemma* [1], [6], [2], a two-person non-zero sum game that has been extensively studied by psychologists, sociologists, economists and political scientists, as well as mathematicians and computer scientists. We study an interesting and highly complex 2-player game known as the *(iterated) Traveler's Dilemma* [8], [9], [14], [18]. Traveler's Dilemma (TD) is a non-zero sum game in which each player has a large number of possible actions or moves. In the iterated context, this means many possible actions in each round and thus, for games with many rounds, an astronomic number of possible strategies overall. What makes Iterated TD particularly interesting is that its structure defies the usual prescriptions of the classic game theory insofar as what constitutes an "optimal" or a "good" strategy.

This paper is organized as follows. We first define the Traveler's Dilemma, briefly motivate its study, and survey the prior art. We then outline several types of strategies that have been studied in the Iterated TD context, with an emphasis on (i) selected classes of learning and adaptable strategies and (ii) some strategies that attempt to capture the idea of *risk minimization* [13]. We summarize our ITD round-robin tournament and analyze performances of various strategies. In that analysis, we place the main emphasis on comparison-and-contrast between various "greedy" strategies studied in the prior experimental work on ITD, especially (i) those that tend to greedily always bid high and (ii) those that try to "outsmart" the opponent [10], [11], against (iii) those strategies that are explicitly or implicitly striving to minimize the risk in the iterated play. We then draw the key lessons from our analysis, and propose conclusions of a broader significance for iterated non-zero sum games. Lastly, we outline some open problems and possible ways forward on gaining deeper understanding of complex strategic interactions among self-interested agents, especially when such interactions provide incentives for cooperative behavior.

II. Traveler's Dilemma

Traveler's Dilemma was originally introduced in [4]. The motivation behind the game was to show the limitations of classical game theory [16], and in particular the concepts of *individual rationality* that stem from game-theoretic notions of "optimal play" based on *Nash equilibria* [4], [5], [24]. The original version of TD (the "default" version), is described as follows:

An airline loses two suitcases belonging to two different travelers. Both suitcases are 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 (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 lower bid will be taken as the true valuation, and each traveler will receive that amount along with a bonus/penalty: $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 declare?

A tacit assumption in the default formulation of TD is that the bids must be integers: the *granularity parameter* is $1, as this amount is the smallest possible difference between two different bids. Generalized version of (iterated) TD, where the granularity of bids is considered (together with the bonus value) one of the game's key parameters and the game structure examined as those parameters are varied, is studied in [21]. The default TD's *unique* Nash Equilibrium (NE) is the action pair $(p, q) = (\$2, \$2)$; however, this NE is clearly rather bad for both players, assuming that players' utilities are directly proportional to the dollar amounts they receive. Yet, it has been argued [4], [9], [12] that a perfectly rational player, according to the classical game theory, would "reason through" and converge to choosing to bid the lowest possible value. Given that TD is symmetric, each player would reason along the same lines and, once selecting \$2, would not deviate from it. In contrast, the non-equilibrium pair of strategies $(\$100, \$100)$ results in each player earning \$100. Adopting one of the alternative notions of game equilibria found in the classical literature does not seem to help. For instance, [14] argue that the action pair $(\$2, \$2)$ is also the game's only *evolutionary equilibrium*. Similarly, seeking *sub-game perfect equilibria* (SGPE) [17] of Iterated TD would not result in more favorable outcomes, either: the set of a game's SGPEs is a subset of that game's full set of Nash equilibria in mixed strategies. Hence, the early studies of TD concluded that this game demonstrates a woeful inadequacy of the classical game theory. However, it has been experimentally shown that humans (both game theory experts and laymen) tend to play far from the NE, and generally at or close to the maximum possible bid (\$100 in the default case), and therefore fare much better than if they followed the Nash equilibrium approach [8], [9].

It has been posited recently that *iterated risk minimization* [13], and not Nash or evolutionary or other classical notions of game equilibria, provide a satisfactory notion of a solution for many games that are challenging from the classical game theory stand-point, including but not limited to ITD. In particular, it has been argued that, for the default version of ITD (with the bonus value of $b = 2$ and the bid granularity $g = 1$ [21]), the unique risk minimizing pure strategy is to always bid \$97 [13]. An experimental investigation of ITD [8] indicated that \$97 is the most successful strategy "in practice". We therefore expand the ITD tournament originally studied in [10], [11] to include some risk-minimization strategies. Two such strategies are (i) to always bid \$97 and (ii) in each round to pick, uniformly at random, one of $\{96, ..., 100\}$; notice that this interval corresponds to the theoretically optimal (with respect to risk minimization) interval $\{100-2b, ..., 100\}$ as discussed in [13]. Other, more complex strategies (some of which combine the idea of risk minimization with different ways predicting the other player's next bid) are discussed in next section.

III. Iterated TD Tournament

Our Iterated Traveler's Dilemma tournament has been inspired by Axelrod's *Iterated Prisoner's Dilemma* tournament [1], [3]. It is a round-robin tournament where each strategy plays against every other strategy N matches, where a match consists of T rounds. The agents do not know T or N and cannot tweak their strategies with respect to the duration of the encounter. Similarly, the strategies are not allowed to use any other assumptions (such as, e.g., the nature of the opponent they are playing against in a given match). Indeed, the only data available to the learning and adaptable strategies in our "pool" of tournament participants is what they can learn and infer about the future rounds, against a given opponent, based on the bids and outcomes of the prior rounds of the current match against that same opponent; no other knowledge of meta-knowledge of any kind is available to the agents.

In order to make a reasonable baseline comparison, we use the same classes of strategies as in [10], ranging from rather simplistic to moderately complex. None of the models of the opponent's behavior is mathematically, cognitively or computationally hard to understand or implement. Therefore, our strategies and their relative performances can be easily re-validated by the research community. We briefly outline the strategies tournament below; more detailed descriptions can be found in [10].

The "Randoms": The first, and perhaps the simplest, class of strategies play a random value, uniformly distributed across a given interval. We have implemented three instances of such strategies using the following intervals: $\{2, 3, ..., 100\}$, $\{99, 100\}$ and $\{96, 97, 98, 99, 100\}$. The first two random strategies were originally introduced in [10]; the third is motivated by the risk minimization idea and approach found in [13].

The "Simpletons": The second extremely simple class of strategies which 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 iterated prisoner's dilemma [1], [2]. The idea behind *Tit-for-Tat* (TFT) is simple: cooperate on the first round, then "do to thy neighbor" (that is, opponent) exactly what he did to you on the previous round. The baseline PD can be viewed as a special case of our TD, when the action space of each agent in the latter game is reduced to just two actions: $\{BidLow, BidHigh\}$. We define two groups of Tit-for-Tat-like strategies for ITD. One group are the *simple* TFT strategies bid value ϵ below the bid made by the opponent in the last round, where we restricted $\epsilon \in \{1, 2\}$. The second group are the *predictive* TFT strategies that compare whether their last bid was lower than, equal to, or higher than that of the other agent. Then a bid is made similar to the simple TFT strategies, i.e., some value ϵ below the bid made by other player in the last round, where again $\epsilon \in \{1, 2\}$. The key distinction is that a bid can be made relative to either the opponent's last bid or the bid made by the agent himself.

"Mixed": The mixed strategies combine up to three pure strategies: for each mixed strategy, a pure strategy is selected from one of the other strategies in the competition independently for each round, according to a specified probability distribution. We choose to use only mixtures of the TFT, Simpleton, and Random strategies, to simplify analysis and ease understanding of the causes behind various strategies' performances. The notation in *Table 1* is *Mixed* followed by up to three $(Strategy, Probability)$ pairs, where each pair represents a strategy and the probability that that particular strategy is selected in a given round. Simpleton strategies are represented simply by their bid, e.g. $(100, 20\%)$. Random strategies are represented by the letter R followed by their range, e.g. $(R[99, 100], 20\%)$. TFT strategies come in two varieties: simple and complex. In *Mixed* strategies, a Simple TFT used in the "mix" is represented by $TFT(y-n)$, where n is the value to bid below the opponent's bid y. Complex TFTs used in a given "mix" are represented with L, E, and H indicators (denoting *Lower*, *Equal* and *Higher*), followed by the bid policy. Bid policies are based on either the opponent's previous bid (y) or this agent's own previous bid (x).

Buckets - Deterministic: These strategies keep a count of each bid by the opponent in an array of *buckets*. The fullest bucket (i.e., the value that has been 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 most recent tied-for-the-lead bucket wins. The strategy then bids the highest possible value strictly below (if possible) the predicted opponent's bid (else, it bids $2). An instance of each tie breaking method above competed as a different bucket-based strategy in the tournament.

Buckets - Probabilistic: As with deterministic buckets, this strategy class counts instances of the opponent's bids and uses them to predict opponent's next bid. Rather than picking the value most often bid, 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 assign greater weights to more recent data than to the older data; details can be found in [10], [22]. We have implemented two different "philosophies" based on deterministic and probabilistic buckets, based on whether an agent that uses the applicable bucket-based strategy tries to be adversarial and "outsmart" the other player by bidding "one under" the opponent's predicted bid, or the agent tries to be more amicable and cooperative, and bid exactly the same value as the predicted bid of the other player. Motivation behind the second approach is to try to "gently push" adaptable opponents toward higher bids in the long run.

Simple Trending: This strategy type 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. For the original *Simple Trend* strategies, several variants of which we also used in the tournament discussed in the present paper, we refer the reader to [10]. We introduce in this paper additional strategies based on the simple idea of opponent's bid prediction based on *linear extrapolation*. In one new type of simple trenders, the general philosophy behind *Simple Trending* is maintained, but with two "tweaks". One, if the opponent's bids have an upward trend, we distinguish whether the opponent's most recent bid is lower than ours or at least as high as ours. The second tweak pertains to the scenario when there is no clear-cut trending either in the upward or the downward direction; that is, the slope of the linear best fit of the opponent's recent bids is within the range between $-\theta$ and $+\theta$. In this case, rather than always bidding "one under" the opponent's predicted next bid, we apply the one-under strategy 80% of the time and we bid the highest possible value, $100, the remaining 20% of the time. The idea behind this tweak is to avoid the underbidding downward spiral that may eventually "sink" both agents down to the undesirable NE $(\$2, \$2)$. The second new type of simple trending based strategies keeps the two "tweaks" as above, and additionally changes the response to downward trend of the opponent's bids: instead of cajoling such an opponent to start increasing her bids by making the highest possible bid, we punish the opponent for decreasing his bids by bidding the lowest possible value, $2, thereby sending the message that we will not allow to be under-bid any more.

Q-learning: This type of strategies 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* [23] 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 [15], without the limitation of having to classify the allegiance of one's opponent. Details on our implementation of Q-learning strategies can be found in [10], [22].

Zeuthen Strategies [25]: A Zeuthen-based strategy calculates the risk level of each agent, and makes *concessions* accordingly. Risk is the ratio of loss from accepting the opponent's proposal vs. the loss of forcing the *conflict deal* (the deal made when no acceptable proposal can be found). While ITD is strictly speaking not a negotiation, one can still treat each player's bid on each round, x_t and y_t, to be a proposal: if $x_t = i$, then agent x is proposing to agent y any pair (i, j) with $j \geq i$ as the next round's action pair. We consider the conflict deal to be the NE at $(\$2, \$2)$. Given agents' proposals, a risk comparison is done. An agent continues making the same bid as long as its risk is greater than or equal to her opponent's. Otherwise, the agent makes the *minimal sufficient concession*: she adjusts her proposal so that (i) her risk is higher than opponent's risk and (ii) the opponent's utility increases as little as possible. Due to the peculiar structure of TD, it is possible that a "concession" actually leads to a loss of utility for the opponent. We have implemented two Zeuthen strategies: one that allows counter-intuitive negative concessions and one that does not.

The metric that we use to evaluate relative performances of various strategies is essentially "the bottom line", that is, appropriately normalized dollar amount that a player would win if she engaged in the prescribed number of plays against a particular (fixed) opponent.

IV. Tournament Results

The Traveler's Dilemma Tournament that we have implemented involves a total of 42 competitors. Each competitor plays each other competitor (including its own "twin") $N = 100$ times. Each match is played for $T = 1000$ rounds. The relatively large number of rounds is to ensure that various adaptable, evolving and/or learning strategies "converge" to a stationary behavior (against a fixed opponent). We note that "learning" or"adaptation" of agents strictly takes place within a given match (i.e., across those 1000 rounds) and does not carry over from one match to the next. The tournament settings in this paper as well as our prior work [10], [11], [21], [22] assume *perfect information* in the sense that, at the end of each round, each agent sees not only the payoff it receives, but also what bid the other agent made in that round.

The results of the tournament, with respect to the "bottom line" performance metric, are summarized in *Table 1*. Some clarifications on the notation are due. In Mixed and TFT strategies, the granularity parameter g is equal to 1 throughout our experiments, but we use the generic notation to be consistent with the literature on *Generalized* ITD where g is allowed to vary [21]. For simple trending strategies where the prediction of the opponent's next bid is based on linear extrapolation over some time window, K denotes the time window and Eps denotes the threshold parameter θ discussed in the previous section. Strategies denotes just as *Simple Trend* are the unmodified, original simple trending strategies as found in [10]. The *Simple Trend Tweak* strategies introduced in the present paper, share the philosophical approach of the original simple trenders when it comes to how they react to a downward trend of the opponent's recent bids; they differ from *Simple Trend* strategies in terms of the two "tweaks" previously discussed. In contrast, the *Simple Trend New* strategies denote those whose underlying philosophy is to punish the downward trend of the other agent's bids by beginning to bid $m =$ \$2; in that sense, *Simple Trend New* strategies have a Tit-For-Tat, vengeance aspect to them. Again, we were keen to investigate whether such adversarial or vengeful approach would tend to get punished or rewarded in the long run – and we suspected it would tend to get punished, as confirmed by the results in *Table 1.* For the bucket-based strategies of both deterministic and probabilistic varieties, the *Under Buckets* notation refers to the old version where, whatever the prediction of the opponent's next bid may be, the strategy strives to bid one under the opponent [10], [11]. In contrast, the strategies denoted simply as *Buckets* are for the first time introduced in the present paper; they are the ones that strive to match the opponent's next bid. In particular, the new *Bucket* strategies are the more amicable and less adversarial versions of the original *Under Bucket* strategies.

0.896494	Always 100
0.893108	Zeuthen Strategy - Positive
0.892622	Simple Trend Tweak - K = 3, Eps = 0.5
0.890135	Random [99, 100]
0.889497	Always 97
0.881519	Always 99
0.880039	Random [96, 100]
0.879055	Mixed - L(y-g) E(x-g) H(x-g), 80%); (100, 20%)
0.875824	Simple Trend Tweak - K = 10, Eps = 0.5
0.861113	Simple Trend - K = 3, Eps = 0.5
0.851513	Simple Trend Tweak - K = 25, Eps = 0.5
0.819885	Mixed - TFT (y-g), 80%); (R[99, 100], 20%)
0.806206	Simple Trend - K = 10, Eps = 0.5
0.770808	Mixed - L(x) E(x) H(y-g), 80%); (100, 20%)
0.730023	Simple Trend - K = 25, Eps = 0.5
0.719380	Simple Trend New - K = 25, Eps = 0.5
0.661882	Q Learn - alpha= 0.5, discount= 0.9
0.659993	Q Learn - alpha= 0.2, discount= 0.0
0.659053	Buckets - (Fullest, Highest)
0.658911	Q Learn - alpha= 0.8, discount= 0.9
0.657025	Q Learn - alpha= 0.2, discount= 0.9
0.655324	Q Learn - alpha= 0.8, discount= 0.0
0.654302	Q Learn - alpha= 0.5, discount= 0.0
0.636117	Mixed - L(y-g) E(x-g) H(x-g), 80%); (100, 10%); (2, 10%)
0.618563	Under Buckets - (Fullest, Highest)
0.610442	Simple Trend New - K = 10, Eps = 0.5
0.609484	TFT - Low(y-g) Equal(x-g) High(x-g)
0.571560	Buckets - (Fullest, Random)
0.566314	Buckets - PD, Retention = 0.8
0.562517	Buckets - PD, Retention = 0.2
0.562493	Buckets - PD, Retention = 0.5
0.525703	Under Buckets - (Fullest, Random)
0.525482	Under Buckets - PD, Retention = 0.5
0.524108	Under Buckets - PD, Retention = 0.8
0.517205	Under Buckets - PD, Retention = 0.2
0.507776	Under Buckets - (Fullest, Newest)
0.494994	TFT - Simple (y-1)
0.435022	Under Buckets - (Fullest, Lowest)
0.416794	Zeuthen Strategy - Negative
0.373517	Random [2, 100]
0.293930	Simple Trend New - K = 3, Eps = 0.5
0.025141	Always 2

Table I
Results w.r.t. Metric U_1

We now summarize the main findings from our round-robin tournament. First, when it comes to simplistic, non-adaptable strategies, we observe the same general pattern previously reported in [10], [11]. In particular, agents that always bid very high (at $M =$ \$100 or close to it) do very well overall. Unlike the results in [10] where the "50-50" random alternation between bidding \$100 and bidding \$99 was the top performer, in the present tournament the pure strategy "Always bid \$100" is the overall winner. This is due to a somewhat different pool of opponents in comparison to the tournament in [10]. The other simple strategies which always bid high and are oblivious to the opponents' bids also generally do well. Two of these strategies are inspired by [13] and the desire to minimize risk in the iterated play. Those risk minimization inspired strategies are "Always bid \$97" and "Randomly choose among $\{96, ..., 100\}$ with equal probabilities". The remaining "Always bid high" strategies are directly taken from the original Iterated TD tournament in [10]. In our view, there is no particular significance of the fact that some of the "bid high" strategies perform (very slightly) better than other such strategies; who comes out on top appears to be primarily due to the exact choice of opponents in the tournament.

We focus on performances of various adaptable and learning strategies. The best adaptable strategy overall turns out to be Zeuthen-positive. That result coincides with the outcome of earlier ITD tournaments (where some, but not all, of the strategies were the same as in our tournament); see [11], [21]. The consistently highly successful performance of the Zeuthen-Positive strategy in ITD provides some interesting lessons. Specifically, the success of this non-greedy, long-term focused, cajoling-the-opponent-to-increase-her-bids strategy indicates that (i) highly collaborative behavior in general tends to get rewarded (at least against a reasonable mix of opponents; if most or all of other participants in the tournament were adversarial, the results would be rather different) and (ii) non-greedy, altruistic behavior in the early rounds, generally turns out to get rewarded in the long run.

We next discuss "Simple Trenders", a set of strategies that predict the opponent's next bid based on a pre-specified "window" of the other agent's K most recent bids. We have showed in [22] that, among all classes of closely related adaptable strategies, simple trenders are the most consistent and successful overall. The new insights on simple trenders are summarized below.

- The single best performer among all nine simple trending strategies (three choices of memory windows, and three "philosophies" insofar as how to respond to different trends by the other player) is our modification of the "cajoling" from [10] for the shortest time window we have experimented with, namely, K = 3.
- The two relatively minor "tweaks" to the default Simple Trending strategies as found in [10], [11] generally work well. In particular, the combination of (i) increasing one's bids once the opponent's bids reach or exceed one's own and (ii) "gently pushing" the opponent to higher bids by periodically bidding \$100 when there is no clear upward or downward trend tends to lead most adaptable opponents to actually bid higher, thereby resulting in greater payoffs to both players in the long run.
- Choosing to punish the opponent with a clear downward bid tendency (by beginning to bid the lowest possible value, $m = \$2$) instead of encouraging such opponents toward high(er) bids, in general, does not work well; in fact, one of the punishing simple trender strategies turns out to be the worst performer among all adaptable strategies. Such adversarial behavior results in poor long-term performance against a wide range of opponents.
- The impact of the width of the "history window" (i.e., memory) apparently depends greatly on the overall philosophy (i.e., cajoling vs. punishing a downward-heading opponent) of a simple trending strategy. In particular, those simple trenders that try to entice downward-heading opponents toward higher bids tend to do better with a shorter memory window. In contrast, simple trenders that punish downward-heading opponents perform decently for relatively long memory windows ($K = 25$), but their performance drastically deteriorates as the memory window shortens, and is abysmal for $K = 3$.

Insofar as the adaptable strategies which use a "bucket" based prediction of the opponent are concerned, we were primarily interested in comparing the bucket-based strategies (both deterministic and probabilistic) as originally defined in [10], with the modified bucket-based strategies as defined in this paper. The prediction model of what the opponent is anticipated to bid in the next round is the same in both types of bucket-based strategies. The difference is in our response to the anticipated bid of the opponent. In particular, the old bucket-based strategies that try to bid "one under" the opponent's predicted next bid can be viewed as somewhat adversarial – they basically try to outsmart the opponent. In contrast, our proposal of the new bucket strategies, where we bid exactly the same value that we anticipate the opponent will bid, are cooperative in a sense that we don't try to outsmart the opponent, but rather indicate that we are happy to earn just as much as they earn, thereby, hopefully, pushing the opponent toward higher bids in general.

Our tournament results experimentally validate that ITD, in general, tends to award cooperative behavior, at least to the extent that we can draw general conclusions based on such a tournament-based study, whose results necessarily are dependent on the choice of the pool of competitors. In particular, *all* amicable, "let's bid exactly as much as we predict the opponent to bid" strategies outperform *all* "let's outsmart and bid one under the opponent's next bid" bucket strategies.

V. Summary and Future Work

We study the Iterated Traveler's Dilemma as an example of a highly complex two-player non-zero sum game. Our method of study is primarily experimental – via simulating a round-robin tournament with a broad variety of competing strategies. The analysis summarized herewith has three main objectives. One, we want to investigate to what extent adaptation and non-trivial models of one's opponent really help an agent, and in particular, how do cognitively and computationally more sophisticated strategies fare in comparison to the simple, "always bid high" based strategies. The simple bid high strategies have been previously experimentally found to tend to do well both in the cognitive psychology context (i.e., with the human subjects engaging in the game) [8], [9] and in the prior computer simulation based studies [10], [11], [22]. Two, given a relatively broad class of strategies (such as Tit-For-Tats or Simple Trenders or Bucket-based), we explore the impact of "tweaking" some key parameters. One motivation behind this inquiry is to gain some insights into possible ways of evolving a parameterized strategy type toward the optimal variation (or at least, the optimal parameter values given the fixed pool of opponents). Second motivation behind comparing-and-contrasting different variants within the same underlying "philosophy" of how play ITD is to, in essence, complement our earlier study [21] which compares and contrasts different classes or "teams" of strategies against each other. Last but not least, we want to experimentally investigate to what extent is striving for *risk minimization*, as defined in important work [13], successful in the Iterated Traveler's Dilemma context, at least with respect to the pool of competing strategies that we consider. While we argue that our pool of strategies covers a fairly broad ground (in particular, it builds on, and expands upon, the shoulder of giants [1], [2]), we are also aware that there is no such a thing as a fully general tournament, hence any tournament-based study unavoidably has its limitations insofar as the generality of its findings.

In future work, we intend to address the inherent limitations of tournament-based studies when the pool of participating strategies is fixed, and to dynamically evolve the set of competitors. Such evolutionary approach to "weeding out" less successful strategies has been found very promising in, e.g., [7]. Evolving a set of strategies so that, in the long run, only the most successful ones are still around would, in case of Iterated TD, also provide novel insights into (i) what are the good parameter values for the parameterized classes of strategies such as those discussed in this paper, as well as (ii) shed more light on the interesting phenomenon of *mutual reinforcement* among pairs of possibly very different adaptable strategies, which was initially addressed in [21] but, in our opinion, warrants a more systematic further study.

Acknowledgments: This work was partially supported by NSF grant IIS-0914861 (first and third authors) and NSF grant CCF-1117259 (the second author).

References

[1] R. Axelrod, Effective choice in the prisoner's dilemma, *Journal of Conflict Resolution*, **vol. 24**(1), pp. 3–25, 1980.

[2] R. Axelrod, The evolution of cooperation, *Science*, **211**(4489), 1390–1396, 1981.

[3] R. Axelrod, *The evolution of cooperation*, Basic Books, 2006.

[4] K. Basu, The traveler's dilemma: Paradoxes of rationality in game theory, *The American Economic Review*, **vol. 84**(2), pp. 391–395, 1994.

[5] K. Basu, The traveler's dilemma, *Scientific American Magazine*, 2007.

[6] B. Beaufils, J.P. Delahaye, P. Mathieu, Complete classes of strategies for the classical iterated prisoner's dilemma, *Evolutionary Programming*, pp. 33–41, 1998.

[7] B. Beaufils, J.P. Delahaye, P. Mathieu, Adaptive Behaviour in the Classical Iterated Prisoner's Dilemma, *Proc. Artificial Intelligence & Simul. Behaviour Symp. on Adaptive Agents & Multi-Agent Systems* (AISB'01), York, UK, 2001.

[8] T. Becker, M. Carter, J. Naeve, Experts playing the traveler's dilemma, Technical report, Department of Economics, University of Hohenheim, Germany, 2005.

[9] C.M. Capra, J.K. Goeree, R. Gmez, C.A. Holt, Anomalous behavior in a traveler's dilemma?, *The American Economic Review*, **vol. 89**(3), pp. 678–690, 1999.

[10] P. Dasler, P. Tosic, The iterated traveler's dilemma: Finding good strategies in games with "bad" structure: Preliminary results and analysis, *Proc of the 8th Euro. Workshop on Multi-Agent Systems, (EUMAS'10)*, 2010.

[11] P. Dasler, P. Tosic, Playing challenging iterated two-person games well: A case study on iterated travelers dilemma, *Proc. WorldComp Foundations of Computer Science (FCS'11), pp. 219–225*, 2011.

[12] J.K. Goeree, C.A. Holt, Ten little treasures of game theory and ten intuitive contradictions, *The American Economic Review*, **vol. 91**(5), pp. 1402–1422, 2001.

[13] J.Y. Halpern, R. Pass, Iterated regret minimization: a new solution concept, *Proc. 21st Int'l Joint Conf. on Artificial Intelligence*, IJCAI'09, pp. 153–158, San Francisco, CA, USA, Morgan Kaufmann Publishers, 2009

[14] S. Land, J. van Neerbos, T. Havinga. Analyzing the traveler's dilemma, Multi-Agent Systems project, 2008.

[15] M.L. Littman, 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, 2001.

[16] J. von Neumann, O. Morgenstern, *Theory of games and economic behavior*, Princeton Univ. Press, Princeton, NJ, 1944.

[17] M. Osborne, *An introduction to game theory*, Oxford Univ. Press, New York, NY, 2004.

[18] M. Pace, How a genetic algorithm learns to play traveler's dilemma by choosing dominated strategies to achieve greater payoffs, *Proc. of the 5th int'l conf. on Computational Intelligence and Games*, pp. 194–200, 2009.

[19] S. Parsons, M Wooldridge, Game theory and decision theory in Multi-Agent systems, *Autonomous Agents and Multi-Agent Systems*, **vol. 5**, pp. 243–254, 2002.

[20] J. S. Rosenschein, G. Zlotkin, *Rules of encounter: designing conventions for automated negotiation among computers*, MIT Press, 1994.

[21] P. Tosic, P. Dasler, How to play well in non-zero sum games: Some lessons from generalized traveler's dilemma, *Active Media Technology*, N. Zhong, V. Callaghan, A. Ghorbani, B. Hu (eds.), *Lecture Notes in Computer Science*, vol. 6890, pp 300–311, Springer, 2011.

[22] P. Tosic, P. Dasler, Strategies for challenging iterated two-player games: Some lessons learned from iterated traveller's dilemma, *Proc. 4th Int'l Conf. on Agents and Artificial Intelligence* (ICAART'12), pp. 72–82, SciTe Press, 2012.

[23] C. Watkins, P. Dayan, Q-learning, *Machine Learning*, **vol. 8**(3-4), pp. 279–292, 1992.

[24] M. Wooldridge, *An Introduction to MultiAgent Systems*, John Wiley and Sons, 2009.

[25] F. Zeuthen, *Problems of monopoly and economic warfare / by F. Zeuthen ; with a preface by Joseph A. Schumpeter*, Routledge and K. Paul, London, 1967.