Effectiveness of Game-Theoretic Strategies in Extensive-Form General-Sum Games

Jiří Čermák[1], Branislav Bošanský[2], and Nicola Gatti[3]

[1] Dept. of Computer Science, Faculty of Electrical Engineering,
Czech Technical University in Prague, Czech Republic,
jiri.cermak@agents.fel.cvut.cz
[2] Aarhus University, Aarhus, Denmark,
bosansky@cs.au.dk
[3] Politecnico di Milano, Milano, Italy,
nicola.gatti@polimi.it

Abstract. Game theory is a *descriptive* theory defining the conditions for the strategies of rational agents to form an equilibrium (a.k.a. the solution concepts). From the *prescriptive* viewpoint, game theory generally fails (e.g., when multiple Nash equilibria exist) and can only serve as a heuristic for agents. Extensive-form general-sum games have a plethora of solution concepts, each posing specific assumptions about the players. Unfortunately, there is no comparison of the effectiveness of the solution-concept strategies that would serve as a guideline for selecting the most effective algorithm for a given domain. We provide this comparison and evaluate the effectiveness of solution-concept strategies and strategies computed by Counterfactual regret minimization (CFR) and Monte Carlo Tree Search in practice. Our results show that (1) CFR strategies perform typically the best, (2) the effectiveness of the Nash equilibrium and its refinements is closely related to the correlation between the utilities of players, and (3) that the strong assumptions about the opponent in Strong Stackelberg equilibrium typically cause ineffective strategies when not met.

1 Introduction

One of the most ambitious aims in Multiagent Systems community is the design of artificial agents able to behave optimally in strategic interactions. Well-known examples include security domains [27], auctions [29], or classical AI problems such as general game-playing [12].

Non-cooperative game theory provides the most elegant mathematical models to deal with the strategic interactions. However, in spite of the recent enormous success of its applications, game theory is a *descriptive theory* describing equilibrium conditions (a.k.a. the *solution concepts*) for strategies of rational agents (e.g., the best known solution concept Nash equilibrium). On the other hand, game theory may fail when used to *prescribe* strategies as needed when designing artificial agents (except for specific classes of games, such as strictly competitive games). Failure appears, for example, when multiple Nash equilibria co-exist in a game. Game theory does not specify which Nash equilibrium to choose and playing strategies from different Nash equilibria may lead to an arbitrarily bad utility.[4] Other well-known examples of failure are due to the assumptions needed to adopt a given solution concepts that are rarely met (e.g., player 2 breaking ties in favor of player 1 in Strong Stackelberg equilibrium).

The following question thus naturally appears: *Is it meaningful to follow prescribed game-theoretic strategies in an actual game play regardless of the lack of the theoretical guarantees?* We experimentally investigate this question, focusing on general-sum two-player finite extensive-form games (EFGs) with imperfect information and uncertainty, since they offer a general representation of sequential interaction of two players with differing preferences, commonly found in real world scenarios (e.g., security scenarios where thieves value items differently than guards). EFGs have been a subject of a long-term research from the perspective of game theory and there is a large collection of solution concepts defined for this class of games. Nash equilibrium is considered to be a weak solution concept in EFGs since it can use non-credible threats and prescribe irrational actions (we provide an example in the following sections). Therefore, several refinements of Nash equilibria have been introduced over the years posing additional constraints over the strategies to rule out undesirable equilibria. In particular, we investigate the refinements *undominated equilibrium* [8] and *Quasi-Perfect equilibrium* [7], since we are interested in the correlation between increasing theoretical guarantees and practical effectiveness of strategies. Furthermore we evaluate Nash equilibria that maximize expected utility of a given player, or sum of the expected utilities of both players. While refinements remove some undesirable equilibria, they do not resolve the main issue of game-theoretic solution concepts and do not guarantee the expected outcome. To obtain guarantees, one can use maximin (prudential) strategies that maximize the expected outcome in the worst case by assuming that the opponent is playing to hurt the player the most, with a complete disregard of his own gains. However, this assumption is often too pessimistic. Alternatively, strategies that are the outcome of a learning algorithm are often used in practice. Two types of learning algorithms are well-established in EFGs: Counterfactual Regret Minimization (CFR; [30]) and Monte Carlo Tree Search (MCTS; [13]). Although none of these algorithms have any guarantees to converge to a game-theoretic solution in general-sum games, they are both widely used (e.g., [23, 10]), since this lack of theory does not necessarily weaken their usefulness as game theory does not provide any guarantees either.

In this paper we provide a thorough experimental comparison of the strategies that are either described by various solution concepts, or that are a result of a learning algorithm. We measure their *effectiveness* as an expected utility value of player 1 against some strategy of player 2 on a set of games with differing characteristics. A similar analysis has been previously conducted only on zero-sum extensive-form games [31], where Nash equilibrium strategies are guaranteed

[4] Let us note that Nash equilibrium would be not the appropriate solution concept if the communication before playing is available.

to achieve at least the value of the game. Interestingly, our results show that the most effective strategies are the outcomes of the CFR algorithm that typically outperform all other strategies. The CFR is widely used in poker domain, but as our results show, the CFR strategies are effective and robust beyond the domain of poker. The effectiveness of the strategies based on Nash equilibria typically varies with the correlation factor between the evaluation of the players (i.e., whether the players are adversarial, or cooperative). Finally, using additional assumptions about the opponent common in solution concepts can result in low effectiveness of the strategies when these assumptions are not met.

2 Technical Background

We focus on general-sum extensive-form games (EFGs) that can be visualized as game trees (see Fig. 1). Each node in the game tree corresponds to a non-terminal state of the game, in which one player is making a move (an edge in the game tree) leading to another state of the game. We restrict to two-player EFGs, where $\mathcal{P} = \{1, 2\}$ is the set of players. We use i to denote a player and $-i$ to denote the opponent of i. Utility function u_i that players aim to maximize assigns a value for every terminal state (leaf in the tree).

The imperfect information is defined using the information sets (the dashed ellipses in Fig. 1). States grouped in a single information set are indistinguishable to player acting in these states. We assume perfect recall, where the players perfectly remember the history of their actions and all observed information.

A pure strategy s_i for player i is a mapping of an action to play to each information set assigned to player i. $\mathcal{S}_i$ is a set of all pure strategies for player i. A mixed strategy δ_i is a distribution over $\mathcal{S}_i$, set of all mixed strategies of i is denoted as Δ_i. When players follow a pair of strategies, we again use u_i to denote the expected utility of player i. We say that strategy s_i weakly dominates s'_i iff $\forall s_{-i} \in \mathcal{S}_{-i} : u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i})$ and $\exists s_{-i} \in \mathcal{S}_{-i} : u_i(s_i, s_{-i}) > u_i(s'_i, s_{-i})$.

Strategies can be represented as behavioral strategies B_i in EFGs, where a probability distribution is assigned over the actions in each information set. Behavioral strategies have the same expressive power as mixed strategies in games with perfect recall [18], but can be more compactly represented using the sequence-form representation [14]. A sequence is a list of actions of player i ordered by their occurrence on the path from the root of the game tree to some node. A strategy is then formulated as a realization plan r_i that for a sequence represents the probability of playing actions in this sequence assuming the other players play such that this actions can be executed.

3 Overview of Solution Concepts

We now provide an overview of the analyzed solution concepts. As a running example we use the game from Fig. 1. The game starts by dealing 1 secret card on the table. Player 1 moves next and decides between (i) action E ending the game right now leading to the immediate outcome 0 for both players, (ii)

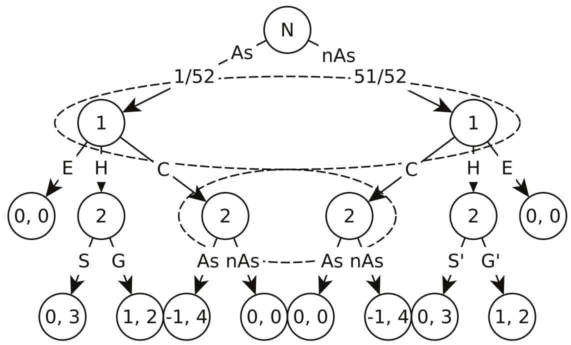

Fig. 1. Example of a EFG termed Informer's Ace of Spades.

showing the player 2 the card dealt H, and (iii) letting player 2 guess without any information C. If player 1 lets player 2 guess, the game reaches a state where player 2 tries to guess whether the hidden card is Ace of spades (As) or not (nAs). A correct guess yields outcome of 4 for player 2, while player 1 gets -1. If player 2 guesses incorrectly both players get 0. If player 1 tells player 2 which card has been dealt, the game proceeds to a state where player 2 decides whether to be selfish (actions S, S' leading to outcomes $(0, 3)$), or to be generous and reward player 1 (actions G, G' leading to outcomes $(1, 2)$).

Nash Equilibrium. The concept of Nash equilibrium (NE) is based on the notion of *best response*. A best response (BR) to a strategy of the opponent δ_{-i} is a strategy δ_i^* for which $u_i(\delta_i^*, \delta_{-i}) \geq u_i(\delta_i, \delta_{-i}),\ \forall \delta_i \in \Delta_i$. We denote the set of the BRs to the strategy δ_{-i} as $BR_i(\delta_{-i})$. We say that strategy profile $\{\delta_i, \delta_{-i}\}$ is a NE iff $\delta_i \in BR_i(\delta_{-i}),\ \forall i \in P$. A game can have more than one NE.

The game from Fig. 1 has 4 pure strategy NE $\{\{(H), (E)\} \times \{(S, S', As), (S, S', nAs)\}\}$. Our example shows why NE is a weak solution concept in EFGs. Some of the NE strategy profiles are not reasonable – for example, playing As for player 2, since the appearance of any other card is more probable. However, playing As is consistent with NE conditions, since player 2 does not expect the information set after action C to be reached as C is not a part of any NE strategy of player 1. Playing E over H is also unreasonable, since player 1 can never lose by preferring H over E (H weakly dominates E). However since any rational player 2 will play S and S' there is, according to NE, nothing to gain by preferring H over E.

Sequential Equilibrium. The sequential equilibrium (SE) due to Kreps and Wilson [16] is the natural refinement of Subgame Perfect equilibrium for games with imperfect information. SE uses a notion of beliefs μ, which are probability distributions over the states in information sets symbolizing the likeliness of being in a state when the information set is reached. Assessment is a pair $(\mu, \mathcal{B})$

representing beliefs for all players and a behavioral strategy profile. A SE is an assessment which is consistent (μ is generated by $\mathcal{B}$) and sequential best response against itself ($\mathcal{B}$ is created by BRs when considering μ). The set of SE forms a non-empty subset of NE [16].

The set of SE of the game from Fig. 1 is $\{\{(H),(E)\} \times \{(S, S', nAs)\}\}$. The SE eliminated the NE that contained the action As, since the SE reasons about the action As and nAs with the additional knowledge of beliefs over the game states in the information set. Since the probability of being in the state where ace of spades was dealt is smaller than the opposite, SE rules out the action As.

Undominated Equilibrium. An undominated equilibrium (UND) is a NE using only undominated strategies in the sense of weak dominance. For two-player EFGs it holds that every UND of EFG G is a perfect equilibrium (as defined by Selten [24]) of its corresponding normal-form game G' and therefore forms a Normal-form perfect equilibrium of G. The set of UND forms a non-empty subset of NE [24].

The UND of the game from Fig. 1 are $\{\{(H)\} \times \{(S, S', As), (S, S', nAs)\}\}$. UND removed two NE where player 1 plays E, since H weakly dominates E.

Quasi-Perfect Equilibrium. Informally speaking, Quasi-Perfect equilibrium (QPE) is a solution concept due to van Damme [7] that requires that each player at every information set takes a choice which is optimal against mistakes of all other players. The set of all QPE forms a non-empty subset of intersection of SE and UND [7].

The only QPE of the game in Fig. 1 is $\{(H), (S, S', nAs)\}$, as it is the NE lying in the intersection of UND and SE.

Outcome Maximizing Equilibria. Besides refinements of NE, it is common that agents follow specific NE strategy based on the expected utility outcome. We use two examples of such specific NE: (1) we use player 1 outcome maximizing equilibrium (P1MNE) and (2) welfare equilibrium (WNE), which is a NE maximizing the sum of expected utilities of both players.

The set of P1MNE of the game from Fig. 1 is equal to the set of NE since all the NE generate the same expected value for player 1. The set of WNE of the game from Fig. 1 is $\{\{(H)\} \times \{(S, S', As), (S, S', nAs\}\}$.

Stackelberg Equilibrium. The Stackelberg equilibrium assumes player 1 to take the role of a leader while player 2 plays a follower. Leader commits to a strategy δ_l observed by the follower. The follower then proceeds to play a pure BR to δ_l. The pair of strategies $(\delta_l, g(\delta_l))$, where $g(\delta_l)$ is a follower best response function $g : \Delta_l \rightarrow S_f$, is said to be a Stackelberg equilibrium iff $\delta_l \in \arg\max_{\delta'_l \in \Delta_l} u_l(\delta'_l, g(\delta'_l))$ and $u_f(\delta'_l, g(\delta'_l)) \geq u_f(\delta'_l, g'(\delta'_l)), \forall g', \delta'_l$. When g breaks ties optimally for the leader, the pair $(\delta_l, g(\delta_l))$ forms a Strong Stackelberg equilibrium (SSE) [19, 4].

The set of SSE of the game from Fig. 1 is $\{\{(H)\} \times \{(S, S', As), (S, S', nAs)\}\} \cup \{\{(E)\} \times \{(S, S', As), (S, S', nAs), (S, G', As), (S, G', nAs)\}\}$.

4 Computing Strategies

We now focus on the algorithms for computing the solution concepts. Computing different solution concepts has different computational complexity; it is PPAD-complete to compute one NE [9], QPE [22] and UND [26], while it is NP-hard to compute some specific NE (i.e., WNE, P1MNE) [5] and SSE [20].

4.1 Computing Game-Theoretic Strategies

We describe the algorithms used to compute NE, UND, P1MNE, WNE, QPE, SSE and maximin strategies in this order. Due to the space constraints we only describe the key idea behind the algorithm. The baseline algorithm is the LCP formulation due to Koller et al. [15], which guarantees all its feasible solutions to be NE. LCP formulation, however, does not allow us to choose some specific NE (e.g., required for UND, WNE, P1MNE). Therefore, we also use a MILP reformulation of the LCP obtained by linearisation of the complementarity constraints [1].

(1) NE is computed by LCP formulation due to Koller et al. [15].

(2) UND is computed by using the MILP formulation [1] maximizing $r_1^\top U_1 r_2^m$, where r_2^m is a uniform strategy of the player 2. This objective ensures finding NE that is a BR to a fully mixed strategy. By definition of dominance, such a NE cannot contain dominated strategies and thus it is a UND [8].

(3) WNE is computed by using the MILP formulation [1] maximizing the sum of expected values of both players.

(4) P1MNE is computed by using the MILP formulation [1] maximizing the expected value of player 1.

(5) QPE is computed using a LCP with symbolic perturbation in ε. The perturbation model used in this paper to produce QPE is due to Miltersen et al. [22]. It restricts the strategies in the game by requiring every sequence σ_i for all players to be played with a probability at least $\varepsilon^{|\sigma_i|+1}$ where $|\sigma_i|$ is the length of σ_i. All the feasible solutions of the perturbed LCP are guaranteed to be QPE.

(6) SSE is computed by the MILP formulation due to Bosansky et al. [3].

(7) Maximin strategies are computed using the LP formulation for finding NE of zero-sum games due to Koller et al. [14] from the perspective of player 1. The strategy resulting from this LP is guaranteed to be maximin, as this formulation assumes player 2 to have the utilities exactly opposite to the utilities of player 1. By maximizing her expected outcome, player 2 minimizes the expected outcome of player 1 and therefore behaves as the worst case opponent.

4.2 Counterfactual Regret Minimization (CFR)

CFR due to Zinkevich et al. [30] iteratively traverses the whole game tree, updating the strategy with an aim to minimize the overall regret. The overall regret is bounded from above by the sum of additive regrets (counterfactual regrets, defined for each information set), which can be minimized independently.

4.3 Monte Carlo Tree Search (MCTS)

MCTS is an iterative algorithm that uses a large number of simulations while iteratively building the tree of the most promising nodes. We use the most typical game-playing variant: UCB algorithm due to Kocsis et al. [13] is used as the selection method in each information set (Information Set MCTS [6]). Additionally we use nesting [2]—MCTS algorithm runs for a certain number of iterations, then advances to each of the succeeding information sets and repeats the whole procedure. This method ensures that all parts of the game tree are visited often enough so that the frequencies, with which the MCTS algorithm selects the actions in this information set, better approximate the optimal behavioral strategy. Moreover, using nesting simulates online decision making, where the players are making the decision in a limited time in each information set.

5 Experiments

We now describe the main results of the paper and provide the comparison of the effectiveness of the different strategies in EFGs – i.e., we assume that player 1 is using a strategy prescribed by some solution concept, or an iterative algorithm, and we are interested in the expected outcome for player 1 against some strategy of player 2. We first describe the domains used for the comparison that include randomly generated games and simplified general-sum poker. Since finding a NE (especially some specific equilibrium) in general-sum games is a computationally hard problem, we follow with the experimental evaluation of the scalability of the algorithms. Next we describe the opponents (strategies for player 2) used for the strategy effectiveness experiments. Finally we measure the effectiveness of strategies against the NE minimizing expected utility of player 1.

5.1 Domains

There is no set of standard benchmark domains available to measure the effectiveness of strategies in general sum EFGs. We try to create such a set by using games with differing characteristics to provide as thorough evaluation of the strategies as possible. We primarily use randomly generated games, since they provide variable sources of uncertainty for the players and we can alter the properties of the games (e.g., utility correlation factor) to investigate the sensitivity to different characteristics. On the other hand, we also use simplified poker games since they are widely used for benchmark purposes, even though they are limited in terms of uncertainty that is only caused by the unobservable action of nature at the beginning of the game.

Extended Kuhn Poker. Extended Kuhn poker is a modification of the poker variant Kuhn poker [17]. In Extended Kuhn poker the deck contains only 3 cards $\{J, Q, K\}$. The game starts with a non-optional bet of 1 called ante, after which each of the players receives a single card and a betting round begins. In this round player 1 decides to either *bet*, adding 1 to the pot, or to *check*. If he bets, second player can either *call*, adding 1 to the pot, *raise* adding 2 to the pot or *fold* which ends the game in the favor of player 1. If player 1 checks, player 2 can either *check* or *bet*. If player 2 raises after a *bet*, player 1 can *call* or *fold* ending the game in the favor of player 2. This round ends by *call* or by *check* from both players. After the end of this round, the cards are revealed and the player with the bigger card wins the pot. Our games are general-sum assuming that 10% of the pot is claimed by the dealer regardless of the winner. Furthermore there are bounties assigned to specific cards, paid when player wins with such card.

Leduc Holdem Poker. Leduc holdem poker [25] is a more complex variant of a simplified poker. The deck contains $\{J, J, Q, Q, K, K\}$. The game starts as in Extended Kuhn. After the end of the first betting round one card is dealt on the table. Second betting round with the same rules begins. After the second betting round, the outcome of the game is determined. A player wins if (1) her private card matches the table card, or (2) none of the players' cards matches the table card and her private card is higher than the private card of the opponent. If no player wins, the pot is split. As before, 10% of the pot is claimed by the dealer. Furthermore there are bounties assigned to specific card combinations.

Randomly Generated Games. Finally, we use randomly generated games without nature. We alter several characteristics of the games: the *depth* of the game (number of moves for *each* player), the *branching factor* representing the number of actions in each information set; number of observation signals generated by the actions (each action of a player generates observation signal, represented as a number from a limited set, for the opponent—the states that share the same history and the sequence of observations belong to the same information set). The utility for player 1 is uniformly generated from the interval $[-100, 100]$. The utility of player 2 is computed based on the correlation parameter from $[-1, 1]$. Correlation set to 1 leads to identical utilities, -1 to a zero-sum game. Otherwise, the utility of player 2 is uniformly generated from a restricted interval around the multiplication of utility of player 1 and the correlation factor.

5.2 Scalability Results

Table 1 provides the computation times (in milliseconds) it takes to compute the equilibrium strategies. Presented numbers are means of 50 runs of different randomly generated games of fixed size. We used Lemke algorithm implementation in C language using unlimited precision arithmetic to compute QPE (similar to [11]), IBM CPLEX 12.5.1 was used to compute all the MILP-based solution concepts.

Table 1. Average computation times (from 50 different games, in [ms]) for computing different solution concepts on random games.

	NE	UND	QPE	P1MNE	WNE	SSE
Random games, bf=2 d=2	10	17	9	52	228	30
Random games, bf=3 d=2	45	250	16	32,749	605,780	60
Random games, bf=3 d=3	7,071	>7,200,000	8,589	>7,200,000	>7,200,000	17,700

By increasing the depth and the branching factor the size of the game grows exponentially and so does the computation time for all the algorithms. There is a notable difference between the solution concepts that use an optimization criteria (UND, P1MNE, WNE), and the ones that do not. While computing an arbitrary NE or QPE takes less than 10 seconds in the game with branching factor 3 and depth 3 (i.e., each player makes 3 moves in the game), MILPs with an objective computing specific equilibria do not finish in 2 hours. Interestingly, while SSE is also NP-hard and uses an objective function, the scalability is better compared to UND, P1MNE, or WNE, since it uses the binary variables only for the sequences of player 2. Scaling to even larger games is possible with NE or QPE solution concepts (e.g., it takes on average almost 39 minutes to compute NE in games with branching factor 4 and depth 4 with MILP). However, it would be impractical for the full range of experiments due to the large number of different settings (different correlation factors, different opponents), hence, the largest games for which we report the results for games are limited to branching factor 3 and depth 3. The scalability results for poker games were similar; hence, we omit the results for UND, P1MNE, and WNE for larger games.

5.3 Imperfect Opponents

We use three types of the opponents to evaluate the effectiveness of the strategies. First we assume that the opponent uses some iterative algorithm (CFR, or MCTS) and we measure the change in expected utility for different solution concepts as the number of iterations used by opponents increases. Both the MCTS and CFR learn their strategy in self-play with no knowledge about the strategies they will be facing. Second, we assume human opponents and use a mathematical model of human behavior called quantal-response equilibrium (QRE) [21]. Calculation of QRE is based on a logit function with precision parameter λ [28]. The logit function prescribes the probability for every action in every information set as follows.

$$B(I,a) = \frac{e^{\lambda u(I,a|B)}}{\sum_{a' \in A(I)} e^{\lambda u(I,a'|B)}} \tag{1}$$

By setting $\lambda = 0$ we get uniform fully mixed strategy and when increasing λ we obtain a behavior, where players are more likely to make less costly mistakes rather than completely incorrect moves, with guaranteed convergence to SE when λ approaches ∞.

Finally, we assume that the opponent is computing NE, however, we do not know which NE will be selected. Therefore, we are interested in the expected value of player 1 against NE strategy of player 2 that minimizes the expected value of player 1 (i.e., the worst case out of all NE strategies player 2 can select).

5.4 Measuring the Effectiveness of Strategies

We present the relative effectiveness of the strategies. The upper and lower bound on the value are formed by the best and the worst response of player 1 against strategy of player 2 (relative effectiveness is equal to 1 if the strategy obtains the same expected utility as the best response, 0 if it obtains the same expected utility as the worst response). We use the relative values to eliminate the inconsistency caused by randomly generated utility values.

Where possible (due to computation-time constraints), we show the relative effectiveness within an interval for best and worst NE strategies—i.e., we use a MILP formulation for computing NE and maximize (or minimize) the expected utility value of player 1 against the specific strategy of player 2. This interval contains *all NE and its refinements* and it is visualized as a grey area in the results. Finally, we use 10^5 iterations of CFR and MCTS algorithms to produce the strategies of player 1 (both algorithms learn in self play with no information about the strategies they will face) used for the effectiveness comparison.

5.5 Results

Results on Random Games. The graphs in Fig. 2 show the effectiveness of the strategies on the random games against different opponents (columns of the grid correspond to CFR, MCTS and QRE opponents). Each curve is computed as the mean over 50 different random games with branching factor 2, depth 2 (first 3 columns) and branching factor 3 and depth 3 (last 3 columns). We also varied the utility correlation (rows of the grid) from -0.75 to 0.75. In each graph, the y-axis shows the relative expected utility of strategy of player 1 against the opponent. The x-axis shows the number of iterations used to generate the imperfect opponent in case of CFR and MCTS, or the λ parameter value in the case of QRE. The standard errors of the means with 95% confidence lie in an interval from 0 to 0.03 and are omitted due to readability of the graphs.

The results show that CFR strategies typically outperform all other strategies with the average relative effectiveness over all scenarios equal to 0.93 (1 being the BR)[5]. This is caused by the fact that CFR optimizes the strategy with respect to the given domain and possible actions of the opponent in this domain, while the rest of the solution concepts assumes a fixed opponent model (e.g., QPE can eliminate an action of the opponent based on a negligible difference in the expected value, while the influence of this action on resulting strategy of CFR is noticeable, since the regret of playing this action is small). This leads to more

[5] The differences between the stated means are statistically significant as the standard error with 95% confidence is no bigger than 10^{-3}

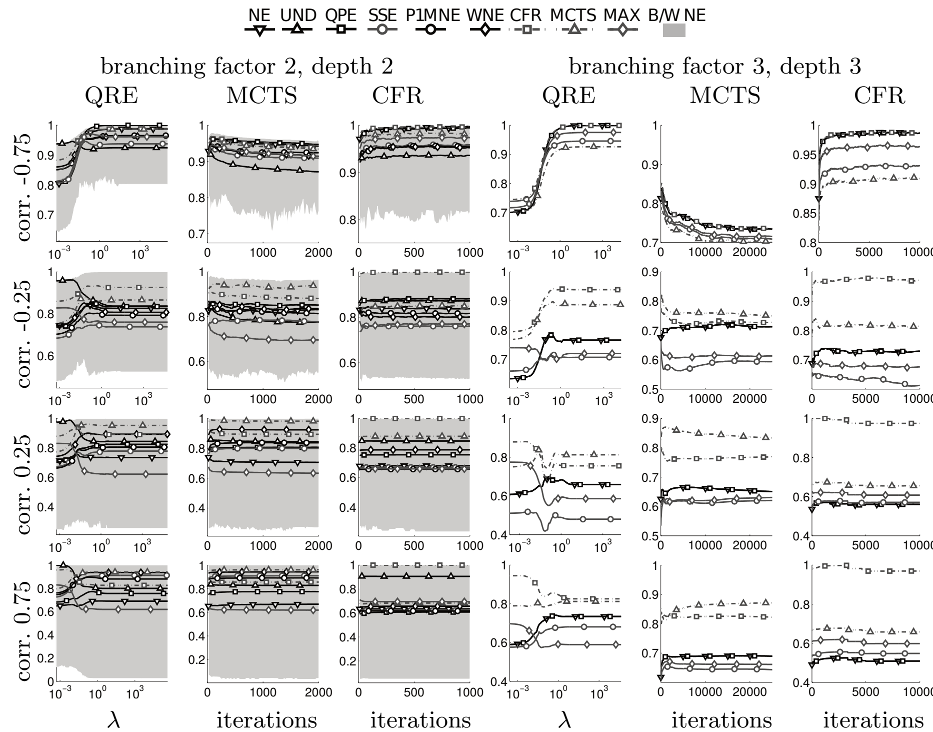

Fig. 2. Relative effectiveness of the strategies computed as a mean over 50 random games with the b.f. 2, d. 2 (columns 1-3) and b.f. 3, d. 3 (columns 4-6) and varying utility correlation against imperfect opponents

effective CFR strategies, especially in competitive scenarios (see the first row in Fig. 2, where correlation is set to -0.75).

MCTS was the second best with the mean of 0.9, however, its effectiveness is not consistently close to the best. It is weak, e.g., against the CFR opponent on random games with depth 3, branching factor 3 and correlation -0.75 (first row last graph in Fig. 2). On the other hand, it has the best effectiveness, e.g., against the QRE opponent on random games with depth 3, branching factor 3 and correlation 0.25 (third row fourth graph).

In games with negative utility correlation the QPE strategies perform well, often very close to CFR (first row in Fig. 2), with the overall mean value equal to 0.85. This is due to the fact that the QPE tries to exploit the mistakes of the opponent. If the correlation is negative, the mistakes of player 2 very likely help player 1, and thus their exploitation significantly improves the expected utility. However as the correlation factor increases the advantages player 1 can get from the mistakes of the player 2 diminish, since the mistakes of player 2 decrease also the utility for player 1, and so the effectiveness of QPE decreases (last row in Fig. 2).

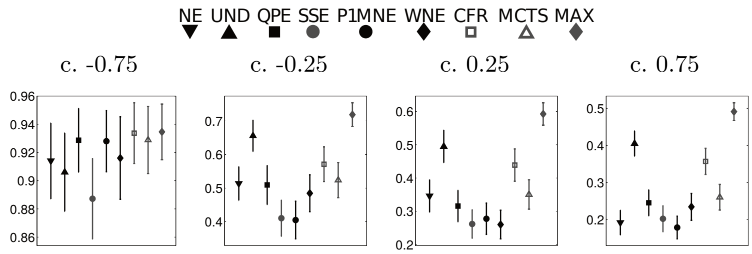

Fig. 3. Confidence intervals depicting the relative expected effectiveness of the measured strategies against the NE minimizing expected value of every measured strategy on random games with b.f. 2 and depth 2

UND strategies are often the worst for the negative correlation (first row in Fig. 2), but their effectiveness increases with higher correlation (last row in Fig. 2). This is because we compute UND using uniform strategy of the opponent in the objective of MILP; hence, the strategy of player 1 is trying to reach leafs with high outcome through the whole game tree. As the utility correlation increases, there is a higher chance that player 2 will try to reach the same leafs.

The NE and its modifications WNE and P1MNE achieve inconsistent effectiveness through the experiments, because their only guarantee is against a fixed strategy of the opponent resulting from the MILP computation. They perform well only if the opponent strategy approximates the strategy computed by the MILP. The results from Fig. 2 show that WNE and P1MNE tend to score better for highly correlated utilities, because the utilities of players are likely to be similar and so there is a higher chance that the opponent will fit to the model assumed and will try to reach the expected nodes.

While maximin strategies provide guarantees in the worst case scenario, they use very pessimistic assumptions about the opponent, assuming that he will aim to hurt player 1 as much as possible with complete disregard of his own gains. Figure 2 shows that the maximin strategies have typically very bad effectiveness, with the exception of random games with negative utility correlation, since the lower the utility correlation, the closer we are to zero-sum games where the maximin strategies form a NE. On the other hand, we compare the effectiveness of the strategies against the NE strategy of player 2 that minimizes the expected value of player 1 (Fig. 3 shows the mean of relative expected values of player 1 with 95% confidence intervals showing the standard error against this strategy). The worst case NE opponent is close to the maximin assumptions; hence, the maximin strategies have the highest effectiveness of all (except for the case of correlation equal to -0.75, where all strategies scored similarly), proving the usefulness of worst case guarantees in this setting (UND was the second best followed by CFR).

Finally, the effectiveness of SSE is typically the worst out of all compared strategies, with the relative mean equal to 0.77. This is caused by the fact that the leader assumes the follower to observe his strategy, and so the leader can

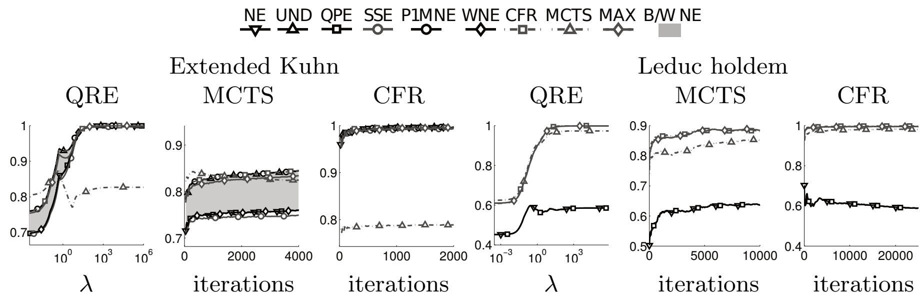

Fig. 4. Relative effectiveness of the strategies on poker against imperfect opponents

play irrationally in some parts of the tree in order to make it less desirable for the follower. The leader changes the set of BRs of the follower in this way, with an aim to maximize his profit. If the follower, however, does not choose the expected BR, the parts of the tree with irrational strategy get visited and the effectiveness decreases.

We have performed additional experiments with varying amount of imperfect information and different sizes of the randomly generated games. Due to the space constraints, the results are omitted, however, they further support the presented findings.

Results on Poker. The results on poker depicted in Fig. 4 offer similar conclusions as the experiments on randomly generated games. CFR again produces very effective and robust strategy. Interestingly, the effectiveness of MCTS strategy significantly varies depending on the opponent. When player 2 uses MCTS strategies as well, MCTS for player 1 significantly outperforms all other strategies, however, this strategy can be easily exploited by both CFR and QRE opponents, making MCTS the worst strategy. The UND and QPE strategies are both always aligned with the best possible NE strategies, while P1MNE is aligned with the worst NE strategy against QRE and MCTS. Finally, SSE is more effective in poker games compared to the randomly generated games (SSE is missing for Leduc holdem, due to the scalability issues). Our conjecture is that the publicly observable actions appearing in poker are closer to the assumption of the observable commitment of player 1.

6 Conclusion

This paper experimentally analyzes the *prescriptive* viewpoint on game theory and compares the effectiveness of game-theoretic strategies and the strategies resulting from iterative algorithms against different opponents modeling rational and human competitors. The results show that the strategies from Counterfactual regret minimization (CFR) outperform other approaches in practice in spite of no convergence guarantees in the general-sum case from the perspective of game theory. Among the strategies prescribed by the game-theoretic solution concepts, the effectiveness often depends on the utility structure of the game. In case the utility functions of players are positively correlated, strategies of Undominated equilibrium provide an effective and robust choice. In case of negative correlation, strategies of Quasi-Perfect equilibrium effectively exploit the mistakes of the opponent and thus provide a good expected outcome. Finally, our results show that strong assumptions about its opponent made by Strong Stackelberg equilibrium cause ineffective strategies when these assumptions are not met in actual experiments.

This paper opens several directions for future work. Our results show that CFR strategies are effective beyond the computational poker, but there is an apparent lack of theory that would formally explain the empirical success of CFR strategies in general-sum sequential games. Next, our experiments were limited by the scalability of the current algorithms for computing different solution concepts and the results might differ with increasing sizes of the games. Therefore, new algorithms that allow scaling to larger games need to be designed in order to confirm our findings for more realistic games.

Acknowledgements. This research was supported by the Czech Science Foundation (grant no. P202/12/2054 and 15-23235S), by the Danish National Research Foundation and The National Science Foundation of China (under the grant 61361136003) for the Sino-Danish Center for the Theory of Interactive Computation. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005), is greatly appreciated.

References

1. Audet, C., Belhaiza, S., Hansen, P.: A new sequence form approach for the enumeration and refinement of all extreme Nash equilibria for extensive form games. International Game Theory Review 11 (2009)
2. Baier, H., Winands, M.H.M.: Nested Monte-Carlo Tree Search for Online Planning in Large MDPs. In: ECAI (2012)
3. Bošanský, B., Čermák, J.: Sequence-Form Algorithm for Computing Stackelberg Equilibria in Extensive-Form Games. In: AAAI (2015)
4. Conitzer, V., Sandholm, T.: Computing the optimal strategy to commit to (2006)
5. Conitzer, V., Sandholm, T.: New complexity results about Nash equilibria. Games and Economic Behavior 63(2), 621 – 641 (2008)
6. Cowling, P.I., Powley, E.J., Whitehouse, D.: Information Set Monte Carlo Tree Search. Computational Intelligence and AI in Games, IEEE Transactions on 4, 120–143 (2012)
7. van Damme, E.: A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. Game Theory 13, 1–13 (1984)
8. van Damme, E.: Stability and Perfection of Nash Equilibria. Springer-Verlag (1991)

9. Daskalakis, C., Fabrikant, A., Papadimitriou, C.H.: The Game World Is Flat: The Complexity of Nash Equilibria in Succinct Games. In: ICALP. pp. 513–524 (2006)
10. Finnsson, H., Björnsson, Y.: Simulation-based approach to general game-playing. In: AAAI. pp. 259–264 (2008)
11. Gatti, N., Iuliano, C.: Computing an extensive-form perfect equilibrium in two-player games. In: AAAI (2011)
12. Genesereth, M., Love, N., Pell, B.: General game-playing: Overview of the AAAI competition. AI Magazine 26, 73–84 (2005)
13. Kocsis, L., Szepesvári, C., Willemson, J.: Improved Monte-Carlo Search. Tech. Rep 1 (2006)
14. Koller, D., Megiddo, N., von Stengel, B.: Fast algorithms for finding randomized strategies in game trees. In: Proceedings of the 26th annual ACM symposium on Theory of computing (1994)
15. Koller, D., Megiddo, N., von Stengel, B.: Efficient computation of equilibria for extensive two-person games. Games and Economic Behavior (1996)
16. Kreps, D.M., Wilson, R.: Sequential equilibria. Econometrica (1982)
17. Kuhn, H.W.: A simplified two-person poker. Contributions to the Theory of Games 1 (1950)
18. Kuhn, H.W.: Extensive games and the problem of information. Annals of Mathematics Studies (1953)
19. Leitmann, G.: On generalized stackelberg strategies. Journal of Optimization Theory and Applications 26 (1978)
20. Letchford, J., Conitzer, V.: Computing optimal strategies to commit to in extensive-form games. In: Proceedings of the 11th ACM conference on Electronic commerce. pp. 83–92. ACM, New York, NY, USA (2010)
21. McKelvey, R.D., Palfrey, T.R.: Quantal response equilibria for normal form games. Games and economic behavior 10(1), 6–38 (1995)
22. Miltersen, P.B., Sørensen, T.B.: Computing a quasi-perfect equilibrium of a two-player game. Economic Theory (2008)
23. Risk, N.A., Szafron, D.: Using counterfactual regret minimization to create competitive multiplayer poker agents. In: AAMAS. pp. 159–166. International Foundation for Autonomous Agents and Multiagent Systems (2010)
24. Selten, R.: Reexamination of the perfectness concept for equilibrium points in extensive games. International Journal of Game Theory 4, 25–55 (1975)
25. Southey, F., Bowling, M.P., Larson, B., Piccione, C., Burch, N., Billings, D., Rayner, C.: Bayes' bluff: Opponent modelling in poker. In: UAI (2005)
26. von Stengel, B., Elzen, A.V.D., Talman, D.: Computing normal form perfect equilibria for extensive two-person games. Econometrica 70 (2002)
27. Tambe, M.: Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned. Cambridge University Press (2011)
28. Turocy, T.L.: Computing sequential equilibria using agent quantal response equilibria. Economic Theory 42, 255–269 (2010)
29. Wellman, M.P.: Trading Agents. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Pub. (2011)
30. Zinkevich, M., Bowling, M., Burch, N.: A new algorithm for generating equilibria in massive zero-sum games. In: AAAI (2007)
31. Čermák, J., Bošanský, B., Lisý, V.: Practical Performance of Refinements of Nash Equilibria in Extensive-Form Zero-Sum Games. In: ECAI (2014)