Game Theory Solution Concepts for an Age of High-Stakes Multiagent Decision-Making

Sam Ganzfried

http://www.ganzfriedresearch.com/

sam.ganzfried@gmail.com

1-card poker

- Both players ante $1 and are dealt a card from a 13-card deck. Player 1 can either bet $1 or check. If player 1 bets, player 2 can either call or fold. If player 1 checks, player 2 can bet $1 or check. If player 1 checks and player 2 bets, then player 1 can call or fold.

1-card poker

- http://www.cs.cmu.edu/~ggordon/poker/
- With optimal play P2 wins $0.064 per hand.

Holding:	2	3	4	5	6	7	8	9	T	J	Q	K	A
1st round:	0.454	0.443	0.254	0.000	0.000	0.000	0.000	0.422	0.549	0.598	0.615	0.628	0.641
2nd round:	0.000	0.000	0.169	0.269	0.429	0.610	0.760	1.000	1.000	1.000	1.000	1.000	1.000

Holding:	2	3	4	5	6	7	8	9	T	J	Q	K	A
On pass:	1.000	1.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
On bet:	0.000	0.000	0.000	0.251	0.408	0.583	0.759	1.000	1.000	1.000	1.000	1.000	1.000

Nash equilibrium

- Nash equilibrium is the central solution concept in game theory. A strategy profile (vector of strategies for all players) is a Nash equilibrium if no player can improve their payoff by deviating to another strategy.
- Superhuman play in no-limit Texas hold ‘em was achieved by trying to approximate Nash equilibrium.
- However, in order to make important decisions when the stakes are high (e.g., self-driving cars or national security), Nash equilibrium on its own is not sufficient.
- In this talk I will present recent research that addresses three distinct shortcomings of Nash equilibrium.

Nash equilibrium assumes perfect rationality

	A	B
C	9, 2	9, 1.99
D	9.01, 2	-1,000,000, 1.99

Rationality vs. safety

- Most classic game-theoretic solution concepts, such as Nash equilibrium, assume that all players are behaving rationally.
- On the other hand, a maximin strategy plays a strategy that has the largest worst-case guaranteed expected payoff; this limits the potential downside against a worst-case and potentially irrational opponent, but can also cause us to achieve significantly lower payoff against rational opponents.
- In two-player zero-sum games these are equivalent.
- We can potentially obtain arbitrarily low payoff by following a Nash equilibrium strategy, but if we follow a maximin strategy will likely be playing far too conservatively.
- Neither Nash equilibrium nor maximin is definitively compelling on its own in multiplayer and non-zero-sum games.

Safe equilibrium

- We propose a new solution concept that balances between these two extremes. In a two-player general-sum game, we define an **ε-safe equilibrium** (ε-SE) as a strategy profile where each player i is playing a strategy that minimizes performance of the opponent with probability ε_i, and is playing a best response to the opponent's strategy with probability 1- ε_i , where $\varepsilon = (\varepsilon_1, \varepsilon_2)$.
  - As a special case, if we are interested in constructing a strategy for player 1, we can set $\varepsilon_1 = 0$, assuming irrationality just for player 2.
- We can generalize this to an n-player game by assuming that all players $i \neq 1$ are playing a strategy that minimizes player 1's payoff with probability ε_i, and are playing a best response to all other players' strategies with probability 1- ε_i, while player 1 plays a best response to all other players' strategies.

Safe equilibrium for two-player games

- Let G be a two-player strategic-form game. Let $\varepsilon = (\varepsilon_1, \varepsilon_2)$, where $\varepsilon_i \in [0,1]$ for $i = 1, 2$. A strategy profile σ^* is an **ε-safe equilibrium** if there exist mixed strategies $\tau^*_i, \rho^*_i \in \Sigma_i$ where $\sigma^*_i = \varepsilon_i\tau^*_i + (1 - \varepsilon_i)\rho^*_i$ for $i = 1, 2$ such that

  $$\rho^*_i \in \text{argmax}_{\sigma i \in \Sigma i}\, u_i(\sigma_i, \sigma^*_{-i})$$

  $$\tau^*_i \in \text{argmin}_{\sigma i \in \Sigma i}\, u_{-i}(\sigma^*_{-i}, \sigma_i)$$

- In practice player i would likely want to set $\varepsilon_i = 0$ and $\varepsilon_j > 0$ $j \neq i$ when determining their own strategy, though the definition allows for an arbitrary value of ε_i.
- Note that Σ_i is the set of mixed strategies for player i in G, u_i is the expected utility for player i, and u_{-i} is the expected utility for the opponent.

Existence of safe equilibrium

Theorem 1. *Let* $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$ *be a two-player strategic-form game, and let* $\epsilon = (\epsilon_1, \epsilon_2)$, *where* $\epsilon_1, \epsilon_2 \in [0, 1]$. *Then* G *contains an* ϵ*-safe equilibrium.*

Proof. Define $G' = (N', (S'_i)_{i \in N}, (u'_i)_{i \in N})$ to be the following game. $N' = \{1, 2, 3, 4\}$, $S'_1 = S'_2 = S_1$, $S'_3 = S'_4 = S_2$. For $s'_i \in S'_i$, define u'_i as follows for $i \in N$:

$$u'_1(s'_1, s'_2, s'_3, s'_4) = -\epsilon_2 u_2(s'_1, s'_3) - (1 - \epsilon_2) u_2(s'_1, s'_4)$$

$$u'_2(s'_1, s'_2, s'_3, s'_4) = \epsilon_2 u_1(s'_2, s'_3) + (1 - \epsilon_2) u_1(s'_2, s'_4)$$

$$u'_3(s'_1, s'_2, s'_3, s'_4) = -\epsilon_1 u_1(s'_1, s'_3) - (1 - \epsilon_1) u_1(s'_2, s'_3)$$

$$u'_4(s'_1, s'_2, s'_3, s'_4) = \epsilon_1 u_2(s'_1, s'_4) + (1 - \epsilon_1) u_2(s'_2, s'_4)$$

Player 1's strategy corresponds to τ_1^*, player 2's strategy corresponds to ρ_1^*, player 3's strategy corresponds to τ_2^*, and player 4's strategy corresponds to ρ_2^*. By Nash's existence theorem, the game G' has a Nash equilibrium, which corresponds to an ϵ-safe equilibrium of G. □

Computational complexity of safe equilibrium

Theorem 2. *Let $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$ be a two-player strategic-form game, and let $\epsilon = (\epsilon_1, \epsilon_2)$, where $\epsilon_1, \epsilon_2 \in [0, 1)$ are fixed constants. The problem of computing an ϵ-safe equilibrium is PPAD-hard.*

Proof. Let σ^G be a Nash equilibrium of G. Suppose that k is the smallest possible payoff for any player in G, and let $k' = k - 1$. Define the game $G' = (N', (S'_i)_{i \in N}, (u'_i)_{i \in N})$ as follows. $N' = \{1, 2\}$, $S'_1 = S_1 \cup t$, $S'_2 = S_2 \cup t$. For $s'_i \in S'_i$, define u'_i as follows for $i \in N$:

$$u'_i(s'_1, s'_2) = u(s'_1, s'_2) \text{ for } s_1 \in S_1, s_2 \in S_2.$$

$$u'_i(t, s'_2) = k' \text{ for } s'_2 \in S_2.$$

$$u'_i(s'_1, t) = k' \text{ for } s'_1 \in S_1.$$

$$u'_i(t, t) = k'.$$

Define $\rho^*_i = \sigma^G_i$, $\tau^*_i = t$, $\sigma^{G'}_i = \epsilon_i \tau^*_i + (1 - \epsilon_i)\rho^*_i$, for $i = 1, 2$. Clearly $\rho^*_i \in \arg\max_{\sigma'_i \in \Sigma'_i} u_i(\sigma'_i, \sigma^{G'}_{-i})$, since σ^G is a Nash equilibrium of G and t is strictly dominated for both players in G' and can be removed without any effect on best responses. It is also clear that $\tau^*_i \in \arg\min_{\sigma'_i \in \Sigma'_i} u_{-i}(\sigma^{G'}_{-i}, \sigma'_i)$, since t minimizes each player's payoff against any possible strategy for the opposing player. This shows that $\sigma^{G'}$ is an ϵ-safe equilibrium of G'. Since the problem of computing a Nash equilibrium is PPAD-hard and we have reduced it to the problem of computing an ϵ-safe equilibrium, this shows that the problem of computing an ϵ-safe equilibrium is PPAD-hard. □

Safe equilibrium for n-player games

- Let G be an n-player strategic-form game. Let $\varepsilon = (\varepsilon_2,\ldots,\varepsilon_n)$, where $\varepsilon_i \in [0,1]$. A strategy profile σ^* is an **ε-safe equilibrium** if there exists a mixed strategy σ^*_1 for player 1 and mixed strategies $\tau^*_i, \rho^*_i \in \Sigma_i$ where $\sigma^*_i = \varepsilon_i\tau^*_i + (1 - \varepsilon_i)\rho^*_i$ for i=2,…,n such that

$$\rho^*_i \in \operatorname{argmax}_{\sigma i \in \Sigma i}\, u_i(\sigma_i, \sigma^*_{-i})$$

$$\tau^*_i \in \operatorname{argmin}_{\sigma i \in \Sigma i}\, u_1(\sigma^*_1, \sigma')$$

$$\sigma^*_1 \in \operatorname{argmax}_{\sigma 1 \in \Sigma 1}\, u_1(\sigma_1, \sigma^*_{-1})$$

  where σ' is the strategy profile for players 2-n where player i plays σ_i and the other players $j \neq i$ play σ^*_j.
- Player 1 is a special player. Player 1 best responds to other players, while each opposing player mixes between playing a best response and minimizing player 1's expected payoff.

Existence and complexity of n-player safe equilibrium

The proof of Theorem 1 extends naturally to $n > 2$ players as well by creating a $2(n-1)+1 = 2n-1$ player game with 2 new players corresponding to each player in the initial game for $i > 1$, plus player 1.

Theorem 3. *Let $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$ be an n-player strategic-form game, and let $\epsilon = (\epsilon_2, \ldots, \epsilon_n)$, where $\epsilon_i \in [0, 1]$. Then G contains an ϵ-safe equilibrium.*

The proof of Theorem 2 also straightforwardly extends to n players.

Theorem 4. *Let $G = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$ be an n-player strategic-form game, and let $\epsilon = (\epsilon_2, \ldots, \epsilon_n)$, where $\epsilon_i \in [0, 1)$ are fixed constants. The problem of computing an ϵ-safe equilibrium is PPAD-hard.*

Example: the game of chicken

	Swerve	Straight
Swerve	0, 0	-1, +1
Straight	+1, -1	-10, -10

Fig. 2: Chicken with numerical payoffs

- Unique mixed-strategy Nash equilibrium σ^{NE} is for each player to swerve with probability 0.9 (there are also two pure equilibria where one player swerves). Unique maximin strategy σ^{M} is to swerve with probability 1.
- If we set $\varepsilon_1 = 0$, σ^{NE} is an ε-safe equilibrium strategy for player 1 for $0 \leq \varepsilon_2 \leq 0.1$, and σ^{M} is an ε-safe equilibrium strategy for player 1 for $0.1 \leq \varepsilon_2 \leq 1$.
- If we set $\varepsilon_1 = 0.05$ and $\varepsilon_2 = 0.15$, an ε-safe equilibrium strategy profile is for player 1 to swerve prob 0.95, and player 2 to swerve prob 0.

Example: security game

	Target 1	Target 2	Target 3
Target 1	4,-3	-1,1	-7,2
Target 2	-5,5	2,-1	-1,4
Target 3	-9,1	-1,8	9,-4

- Nash equilibrium for player 1 (row player) σ^{NE} is to defend the targets with probabilities (0.3136, 0.4661, 0.2203), and a maximin strategy σ^{M} is (0.6144, 0.0131, 0.3725).
- If we set $\varepsilon_1 = 0$, σ^{NE} is an ε-safe equilibrium strategy for player 1 for $0 \leq \varepsilon_2 \leq 0.314$, σ^{M} is an ε-safe equilibrium strategy for $0.569 \leq \varepsilon_2 \leq 1$. But for $0.314 \leq \varepsilon_2 \leq 0.569$ ε-safe equilibrium strategy for player 1 is (0.4437, 0.366, 0.1897), which is neither a Nash equilibrium strategy nor a maximin strategy.

Nash equilibrium refinements

- While Nash equilibrium (NE) has emerged as the central game-theoretic solution concept, many important games contain several Nash equilibria and we must determine how to select between them in order to create real strategic agents. Several Nash equilibrium refinement concepts have been proposed and studied for sequential imperfect-information games, the most prominent being trembling-hand perfect equilibrium, quasi-perfect equilibrium, and recently one-sided quasi-perfect equilibrium. These concepts are robust to certain arbitrarily small mistakes, and are guaranteed to always exist.
- NE refinements in sequential imperfect-information games:
  - Perfect Bayesian equilibrium, subgame perfect equilibrium, sequential equilibrium, trembling-hand perfect equilibrium (normal-form and extensive-form), quasi-perfect equilibrium, one-sided quasi-perfect equilibrium, proper equilibrium (normal-form and extensive-form).

Trembling-hand perfect equilibrium

- Given a strategic-form game G, define a *perturbed game* as a game which is is identical to G except only totally mixed strategies (i.e., strategies that play all pure strategies with non-zero probability) can be played. A strategy profile σ^* in G is a trembling-hand perfect equilibrium (THPE) if there is a sequence of perturbed games that converges to G in which there is a sequence of Nash equilibria of the perturbed games that converges to σ^*.
  - Guaranteed to exist in every finite strategic-form game.
- An extensive-form trembling-hand perfect equilibrium (EFTHPE) is defined analogously by requiring that every action at every information set for each player is taken with non-zero probability. EFTHPE are then limits of equilibria of such perturbed games as the tremble probabilities go to zero.

Trembling-hand perfect equilibrium

	L	R
U	1, 1	2, 0
D	0, 2	2, 2

- Two pure strategy equilibria (U,L) and (D,R).
- Assume row player is playing $(1-\varepsilon, \varepsilon)$ for $0 < \varepsilon < 1$.
- Player 2's expected payoff of L is $1+\varepsilon$, and of R is 2ε.
- For small ε, player 2 should place maximal weight on L.
- By symmetry, player 1 should place maximal weight on U if player 2 is playing mixed strategy $(1-\varepsilon, \varepsilon)$.
- So (U,L) is THPE, but similar analysis fails on (D,R).

Two-player one-step extensive-form imperfect-information games (OSEFGs)

- There are two players, P1 and P2.
- Player 1 is dealt private information τ_1 from a finite set T_1 uniformly at random.
  - Our analysis still holds for arbitrary probability distributions.
- Player 1 can then choose action a_1 from finite set A_1.
- Player 2 observes the action a_1 but not τ_1.
- Player 2 then chooses action a_2 from finite set A_2.
- Both players are then given payoff $u_i(\tau_1, a_1, a_2)$.

Observation of irrational play

- Suppose we are player 2 responding to observed action a_1 of player 1. Suppose we are following our component from Nash equilibrium profile σ^* in which $\sigma^*(\tau_1, a_1) = 0$ for all τ_1. Then our observation is clearly inconsistent with player 1 following σ^*. Since our strategy is part of a Nash equilibrium, it ensures that player 1 cannot profitably deviate from σ^*_1 with any τ_1 and take a_1; however, there may be many such strategies, and we would like to choose the best one given that we have actually observed player 1 irrationally playing a_1.
- Playing an EFTHPE may ensure that we play a stronger strategy against this opponent, who has selected an action that they should not rationally play, since EFTHPE explicitly ensures robustness against the possibility of "trembling" and playing such an action with small probability.

Core assumption of game theory

- EFTHPE assumes that all players take all actions at all information sets with non-zero probability. In this situation, we know that player 1 is taking a_1 at some information set with non-zero probability; however, we really have no further information beyond that. It is possible that there are playing a strategy that plays some other action a'_1 with zero probability for all τ_1.
- The core assumption of game theory is that, in the absence of any information to the contrary, we assume that all players are behaving rationally. Clearly this assumption is violated here. However, it is extreme to now assume that all players are playing all actions with nonzero probability. If we assume that the opponent is playing *as rationally as possible given our observations*, then we would only consider trembles that are consistent with our observations of their play.

Observable Perfect Equilibrium (OPE)

- Trembles consistent with our observations must satisfy $\sigma_1(\tau_1, a_1) > 0$ for at least one τ_1, or alternatively, $\sum_{\tau 1}\sigma_1(\tau_1, a_1) > 0$.
- The concept of OPE captures this assumption that all players are playing as rationally as possible subject to the constraint that their play is consistent with our observations.

Definition 1. *Let G be a two-player one-step extensive-form game of imperfect information, and suppose that player 2 has observed public action a_1 from player 1. Then σ^* is an* observable perfect equilibrium *if there is a sequence of perturbed games, in which player 1 is required to play a_1 with non-zero probability for at least one $\tau_1 \in T_1$, that converges to G, in which there is a sequence of Nash equilibria of the perturbed games that converges to σ^*.*

Proposition 1. *Every observable perfect equilibrium is a Nash equilibrium.*

- We can extend the definition to general n-player imperfect-info games by adding analogous constraints for all observed actions.
- No longer required to reason about trembles off path of play.

Quasi-perfect Equilibrium

- For general EFGs there is a further consideration about what trembles should be considered for future moves beyond the path of play. EFTHPE assumes that all players may tremble in future actions, while quasi-perfect equilibrium (QPE) assumes that only the opposing players tremble for future actions (even if we have trembled previously ourselves). One-sided quasi-perfect equilibrium (OSQPE) assumes that only the opposing players tremble at all and we cannot. OSQPE is most computationally efficient and is also the most similar to OPE.
- In OSEFGs QPE and EFTHPE are identical, since both players only take a single action along the path of play. Both potentially differ from OSQPE, which from the perspective of player 2 requires only that player 1 puts non-zero probability on all possible actions. All of these potentially differ from OPE.

Computation and Existence of OPE

- We can adapt results from OSQPE [Farina, Sandholm NeurIPS '21] to construct an efficient algorithm for OPE computation and a proof of its existence by solving a sequence of linear programs. Given a fixed $\varepsilon > 0$, we can define a trembling linear program whose solution is an ε-OPE. We can solve the problem for consecutively smaller values of ε until a termination criterion is met, providing a polynomial-time algorithm (as well as proving existence).
- Like for OSQPE, the OPE LP formulation depends on ε only through the objective (while EFTHPE depends on ε through LHS of constraints and QPE depends on ε in both RHS of constraints and objective). Our objective also has far fewer ε terms in the objective than OSQPE, suggesting that OPE can be computed faster.

No-limit poker

- Major AI challenge problem for which we have recently achieved superhuman performance for both 2 and 6-player.
  - Abstraction algorithms have no performance guarantee.
  - Endgame solving has no performance guarantee.
  - Counterfactual regret minimization does not guarantee convergence to Nash equilibrium for more than two players, and furthermore there can be multiple NE and following one has no performance guarantee.
- It turns out that even ignoring all of these theoretical limitations, there is an additional challenge present. Even if we are in the two-player zero-sum setting and are able to compute an exact Nash equilibrium, the game may contain many Nash equilibria, and we would like to choose the "best" one. As we will see, even the simplest two-player no-limit poker game contains infinitely many Nash equilibria.

No-limit clairvoyance game (NLCG)

- Player 1 is dealt a winning hand (W) and losing hand (L) each with probability ½.
  - While P2 is not explicitly dealt a "hand", we can view P2 as always being dealt a medium-strength hand that wins against a losing hand and loses to a winning hand.
- Both players have initial chip stacks of n and ante $0.50 (creating an initial pot of $1).
- P1 is allowed to bet any integral amount x in [0,n].
  - A bet of 0 is called a *check*.
- Then P2 is allowed to call or fold (but not raise).
- The game clearly falls into the class of OSEFGs.

Nash equilibrium of NLCG

- The game is small enough that its solution can be computed analytically (even for continuous bet sizes) [Ankenman, Chen "The Mathematics of Poker"].
  - P1 bets n with probability 1 with winning hand.
  - P1 bets n with probability n/(1+n) with losing hand (and checks otherwise).
  - For all x in (0,n], P2 calls a bet of size x with probability 1/(1+x).
- Despite the game's simplicity, the solution has been used by strong no-limit Texas hold'em agents for interpreting bet sizes for the opponent that fall outside an abstracted game model.

Full set of equilibria in NLCG

- It turns out that player 2 does not need to call a bet of size $x \neq n$ with exact probability 1/(1+x): they need only not call with such an extreme probability that player 1 has an incentive to change their bet size to x (with either a winning or losing hand).

Proposition 5. *A strategy profile σ^* in the no-limit clairvoyance game is a Nash equilibrium if and only if under σ^* player 1 bets n with probability 1 with a winning hand and bets n with probability $\frac{n}{1+n}$ with a losing hand, and for all $x \in (0, n]$ player 2 calls vs. a bet of size x with probability in the interval $\left[\frac{1}{1+x}, \min\left\{\frac{n}{x(1+n)}, 1\right\}\right]$.*

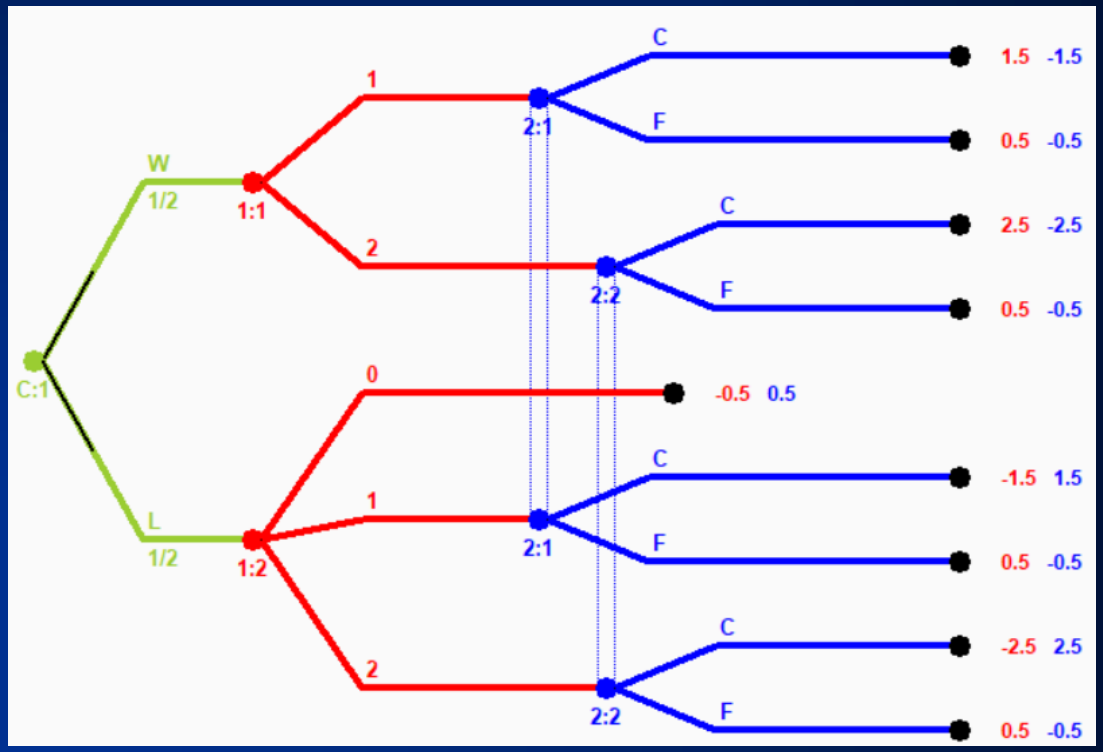
- Player 2 has infinitely many NE strategies that differ in their frequencies of calling vs "suboptimal" bet sizes of player 1.

Equilibrium selection in NLCG

- Argument for calling at interval lower threshold 1/(1+x):
  - If the opponent bets x as opposed to the optimal size of n, a reasonable deduction is that they aren't even aware that n would have been the optimal size, and believes that x is optimal. Therefore, it would make sense to play a strategy that is an equilibrium in the game where the opponent is restricted to only betting x (or to betting 0), which corresponds to calling a bet of x with prob. 1/(1+x). The other equilibria pay more heed to the concern that the opponent could exploit us by deviating to bet x instead of n; but we need not be as concerned about this possibility, since a rational opponent who knew to bet n would not bet x.
- One could use a similar argument to defend upper threshold.
- First argument seems much more natural, but both could be appropriate depending on assumptions on opponent's reasoning process. Without such information, we assume both are rational and look for a theoretically principled NE refinement selection.

NLCG with n = 2

- Unique NE strategy for player 1 is to bet 2 with prob. 1 with winning hand, bet 2 with prob. 2/3 with losing hand, and check prob. 1/3 with losing hand. Player 2 calls bet of 2 with prob. 1/3, and calls a bet of 1 with prob. in [1/2, 2/3].



NE refinement solutions for NLCG

- The unique EFTHPE is for player 2 to call a bet of 1 with probability 2/3 (upper interval threshold). Since this is a OSEFG, it is also the unique QPE. And since player 2's strategy is fully mixed, this is also the unique OSQPE.
- However, the unique OPE for player 2 is to call with prob. 5/9.
  - Note that the OPE does not simply correspond to the average of the two interval boundaries, which would be 7/12.
- In the OPE solution, player 1 loses an equal amount of expected payoff by betting 1 instead of 2 or 0 with both bluffs and value bets (both lose 1/9), while in the other equilibria player 1 loses an unequal amount between both types of mistakes.

Opponent modeling

- A Nash equilibrium strategy is static and does adapt its strategy to information on the opponents even if it is available (e.g., historical data or observations of repeated play).
- We will show that opponent modeling can lead to significantly stronger performance against a range of opponents in a multiplayer imperfect-information game (3-player Kuhn poker).
- Our approach has the following features:
  - It is computationally efficient and scalable to large games.
  - It applies to any number of opponents (not just two-player zero-sum).
  - It does not assume that any historical data is available.
  - It assumes the general setting of partial observability of opponents' private information (which includes full and none as special cases).
  - It is domain independent and assumes only access to partial observations of opponents' play in the current series of games.

Meta-algorithm for two-player games

Algorithm 1 Meta-algorithm for Bayesian opponent exploitation in two-player imperfect-information games [7]

Inputs: Prior distribution p_0, response functions r_t for $0 \leq t \leq T$

$M_0 \leftarrow \overline{p_0(\sigma_{-i})}$
$R_0 \leftarrow r_0(M_0)$
Play according to R_0
for $t = 1$ to T **do**
    $x_t \leftarrow$ observations of opponent's play at time step t
    $p_t \leftarrow$ posterior distribution of opponent's strategies given prior p_{t-1} and observations x_t
    $M_t \leftarrow$ mean of p_t
    $R_t \leftarrow r_t(M_t)$
    Play according to R_t

Algorithm for opponent modeling in multiplayer imperfect-information games

Algorithm 2 Opponent modeling algorithm for multiplayer imperfect-information games

Inputs: Rounding threshold ϵ, number of samples k, switching time H, our default strategy σ^*, prior strategy means $\{\sigma_{i,j}\}$ for all opposing players i in position j, Dirichlet factor parameter η.

Simulate k strategies for each opponent i in each position j by $\{\tau_{i,j,s}\}$ = CreateSamples(ϵ, k, $\{\sigma_{i,j}\}$, η).
Initialize prior probabilities of all samples to be equal.
for $t = 1$ to H **do**
    Follow σ^* and collect observations on opponents' play.
    Update posterior strategy probabilities from the new observations to create opponent models.
for $t = H + 1$ to T **do**
    Update posterior strategy probabilities from the new observations to create opponent models.
    Compute and play best response strategy to the opponent models.

Creating sampled strategies for each opponent player label/position

Algorithm 3 CreateSamples

Inputs: Rounding threshold ϵ, number of samples k, prior strategy means $\{\sigma_{i,j}\}$ for all opposing players i in position j, Dirichlet factor parameter η.

for each opposing player i and position j **do**
  for each information set $q_{i,j}$ **do**
    for each action a at information set $q_{i,j}$ **do**
      if $\sigma_{i,j}^{q_{i,j}}(a) < \epsilon$ **then**
        $\sigma_{i,j}^{q_{i,j}}(a) = \epsilon$
      else if $\sigma_{i,j}^{q_{i,j}}(a) > 1 - \epsilon$ **then**
        $\sigma_{i,j}^{q_{i,j}}(a) = 1 - \epsilon$
    Normalize $\sigma_{i,j}^{q_{i,j}}$
    for $s = 1$ to k **do**
      $z_s = 0$
      for each action a at information set $q_{i,j}$ **do**
        x_a = sample from Gamma distribution with parameters $\eta \cdot \sigma_{i,j}^{q_{i,j}}(a)$ and 1
        $z_s = z_s + x_a$
      for each action a at information set $q_{i,j}$ **do**
        $\tau_{i,j,s}^{q_{i,j}}(a) = \frac{x_a}{z_s}$
return $\{\tau_{i,j,s}, s = 1 \text{ to } k\}$

Updating posterior probabilities and computing opponent models

Algorithm 4 Update posterior strategy probabilities

Inputs: Current posterior probabilities $p(s_1, \ldots, s_n)$ for sample s_i for opposing player i, sampled strategies $\{\tau_{i,j,s}\}$ for all opposing players, our current strategy σ, observations from current round of play o.

$z = 0$

for each combination of sample indices $s = (s_1, \ldots, s_n)$ **do**

z_s = probability we observe play consistent with o when we follow σ and opponents follow sampled strategies $\tau_{1,j_1,s_1}, \ldots, \tau_{n,j_n,s_n}$

$q(s_1, \ldots, s_n) = p(s_1, \ldots, s_n) * z_s$

$z = z + p(s_1, \ldots, s_n) * z_s$

for each combination of sample indices $s = (s_1, \ldots, s_n)$ **do**

$q(s_1, \ldots, s_n) = \frac{q(s_1, \ldots, s_n)}{z}$

return $q(s_1, \ldots, s_n)$

Algorithm 5 Compute opponent models

Inputs: Current posterior probabilities $p(s_1, \ldots, s_n)$ for sample s_i for opposing player i in position i_j, sampled strategies $\{\tau_{i,j,s}\}$ for all opposing players i in position i_j

$m_{i,j} \leftarrow$ vector of all zeros for all info. sets and actions.

for opposing players i in position i_j **do**

for each information set $q_{i,j}$ **do**

for each action a at information set $q_{i,j}$ **do**

for each combination of sample indices $s = (s_1, \ldots, s_n)$ **do**

$m_{i,j}^{q_{i,j}}(a) = m_{i,j}^{q_{i,j}}(a) + p(s_1, \ldots, s_n)\tau_{i,j,s_i}^{q_{i,j}}(a)$

return $\{m_{i,j}\}$

Three-player Kuhn poker

- Three-player Kuhn poker is a simplified form of limit poker that has been used as a testbed game in the AAAI Annual Computer Poker Competition for several years. There is a single round of betting. Each player first antes a single chip and is dealt a card from a four-card deck that contains one Jack (J), one Queen (Q), one King (K), and one Ace (A). The first player has the option to bet a fixed amount of one additional chip or to check. When facing a bet, a player can call or fold. No additional bets or raises beyond the additional bet are allowed. If all players but one have folded, then the player who has not folded wins the pot, which consists of all chips in the middle. If more than one player has not folded by the end there is a showdown, at which the players reveal their private card and the player with the highest card wins the pot (which consists of the initial antes plus all additional bets and calls).
- As one example of a play of the game, suppose the players are dealt Queen, King, Ace respectively, and player 1 checks, player 2 checks, player 3 bets, player 1 folds, and player 2 calls; then player 3 wins a pot of 5, for a profit of 3 from the start of the hand (while player 1 loses 1 and player 2 loses 2).

Experiments

- We experimented with our agent against ten agents created by students for a class project as well as three exact Nash equilibrium strategies. The class agents utilize a wide range of approaches that are potentially dynamic.
  - These approaches include neural networks, counterfactual regret minimization, opponent modeling, and rule-based approaches.)
- The Nash equilibrium strategies are all from the “robust” subfamily of equilibrium strategies that have been singled out as obtaining the best worst-case payoff assuming that the other agents are following one of the strategies given by the computed infinite equilibrium family. In particular, we use the strategy falling at the lower bound of this robust equilibrium subfamily, the upper bound, and the midpoint. For each grouping of 3 agents we ran 10 matches consisting of 3000 hands between each of the 6 permutations of the agents (with the same cards being dealt for the respective positions of the agents in each of the duplicated matches). The number of hands per match (3000) is the same value used in the Annual Computer Poker Competition, and the process of duplicating the matches with the same cards between the permutations is a common approach that significantly reduces the variance.

Agent parameters

- For our agent we use the following parameter values:
  - $\varepsilon = 0.05$, $k = 10$, $H = 10$, $\eta = 4$.
- For our default strategy σ^* we use the strategy at the lower bound of the space of robust Nash equilibrium strategies, and for the prior strategy means for the opponents σ_{ij} we use the strategy at the midpoint of the space of robust Nash equilibrium strategies. We chose this strategy for σ^* because it performed the best out of the lower bound, upper bound, and midpoint in preliminary experiments. We chose the midpoint as the prior means for the opponents because it has the most entropy out of the three strategies and has fewer actions with probability 0 which may be problematic when modeling unknown opponents.

Results

- MBBR denotes our multiplayer Bayesian best response agent, N1 denotes the lower bound Nash agent, N2 denotes the upper bound Nash agent, N3 denotes the midpoint Nash agent, and C1-10 are the class project agents. The values are the winrates of per hand multiplied by 1,000 (i.e., a winrate of 30 means winning 0.03 chips per hand). The standard errors for all agents are between 0.4 and 0.6 (using the same units). The MBBR agent clearly outperformed all other agents with statistical significance. We can see that the exact Nash equilibrium agents also performed well, coming in 2nd, 3rd, and 6th.

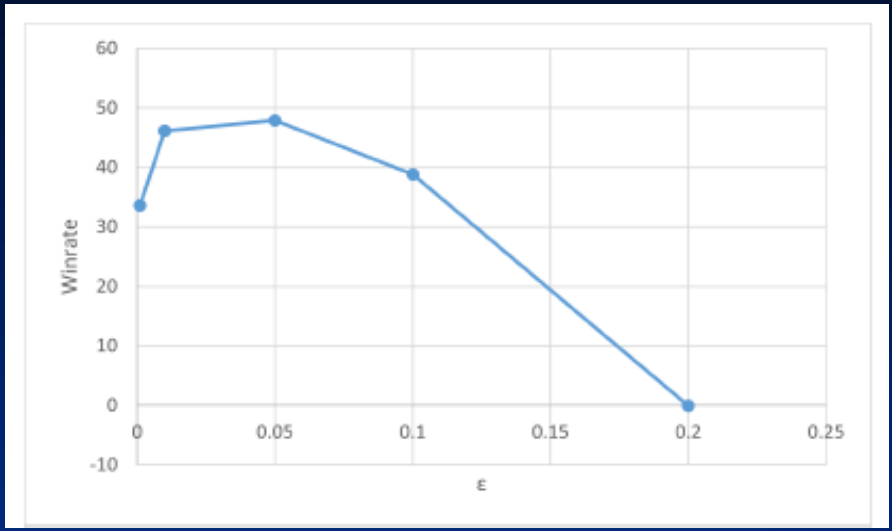
MBBR	N1	N3	C1	C2	N2	C3	C4	C5	C6	C7	C8	C9	C10
48	37	35	34	33	30	22	14	11	7	-22	-33	-44	-172

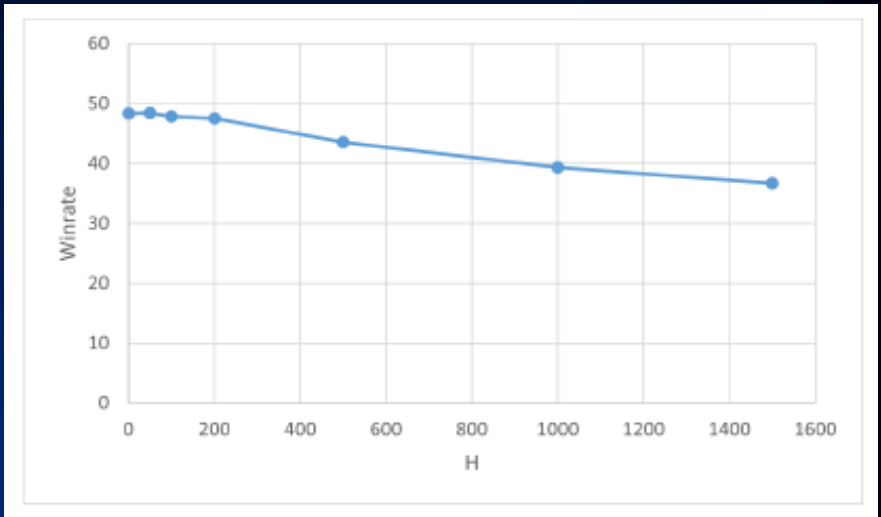
Informed vs. uninformed prior

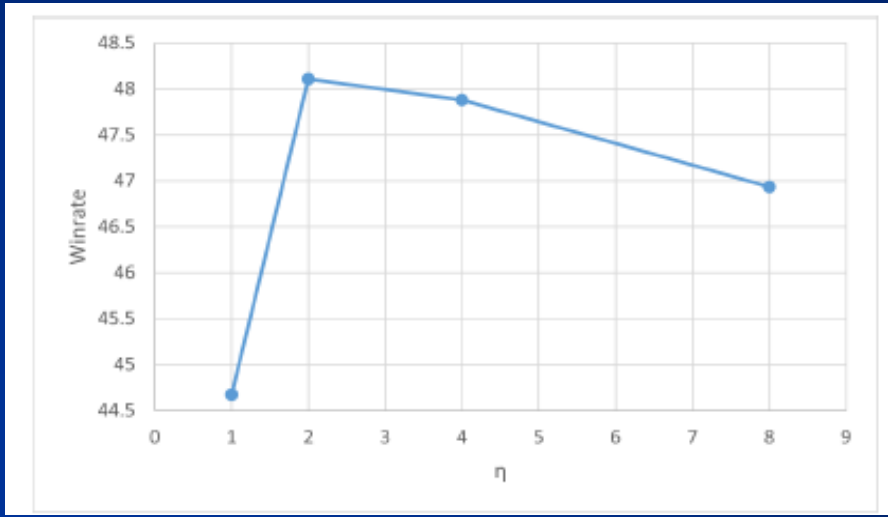
- The following table shows the results using the same parameter settings except using “uninformed priors” for the opponents that are independent Dirichlet with all parameters equal to 2, as done in prior work. We can see that this leads to extremely poor performance, as the MBBR agent now finishes in second to last place. This shows that it is critical to use an informed prior mean strategy for the opponents, such as an exact or approximate Nash equilibrium, as opposed to a naïve random strategy.

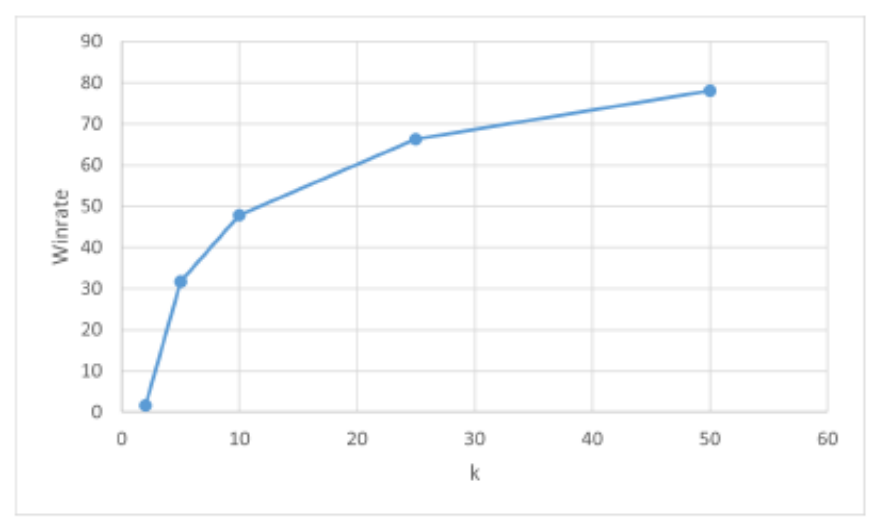
N1	C2	N3	C1	N2	C3	C4	C5	C6	C7	C8	C9	MBBR	C10
47	46	44	42	40	31	24	22	19	-19	-25	-42	-54	-175

Parameter sensitivity









Conclusion

- While Nash equilibrium is the central solution concept in game theory, it has significant limitations that hinder its effectiveness in important real-world settings.

1. It assumes all players are perfectly rational and may not be robust to even small degrees of irrationality.
2. Many games, including two-player zero-sum games, contain multiple (even infinitely-many) Nash equilibria. A real agent must select which one to play.
3. Nash equilibrium does not take into account any historical data or observations of opponents' play. Opponent modeling can lead to much higher payoffs.

- Full versions at http://www.ganzfriedresearch.com/.
- Sam Ganzfried. 2023. Safe Equilibrium. In Proceedings of the *IEEE Conference on Decision and Control (CDC)*.
- Sam Ganzfried. 2023. Observable Perfect Equilibrium. In Proceedings of the *Conference on Decision and Game Theory for Security (GameSec)*.
- Sam Ganzfried, Kevin A. Wang, and Max Chiswick. 2022. Bayesian Opponent Modeling in Multiplayer Imperfect-Information Games. ArXiv preprint.