Equilibrium Refinements Improve Subgame Solving in Imperfect-Information Games

Ondřej Kubíček[1,2] , **Viliam Lisý**[1] and **Tuomas Sandholm**[2,3,4,5]
[1]Czech Technical University in Prague
[2]Carnegie Mellon University
[3]Strategy Robot, Inc.
[4]Strategic Machine, Inc.
[5]Optimized Markets, Inc.
kubicon3@fel.cvut.cz, viliam.lisy@agents.fel.cvut.cz, sandholm@cs.cmu.edu

Abstract

Subgame solving is a technique for scaling algorithms to large games by locally refining a precomputed blueprint strategy during gameplay. While straightforward in perfect-information games where search starts from the current state, subgame solving in imperfect-information games must account for hidden states and uncertainty about the opponent's past strategy. *Gadget games* were developed to ensure that the improved subgame strategy is robust against any possible opponent's strategy in a zero-sum game. Gadget games typically contain infinitely many Nash equilibria. We demonstrate that while these equilibria are equivalent in the gadget game, they yield vastly different performance in the full game, even when facing a rational opponent. We propose *gadget game sequential equilibria* as the preferred solution concept. We introduce modifications to the sequence-form linear program and counterfactual regret minimization that converge to these refined solutions with only mild additional computational cost. Additionally, we provide several new insights into the surprising superiority of the *resolving gadget game* over the *max-margin gadget game*. Our experiments compare different Nash equilibria of gadget games in several standard benchmark games, showing that our refined equilibria consistently outperform unrefined Nash equilibria, and can reduce the exploitability of the overall strategy by more than 50%.

1 Introduction

Settings where parties such as people, companies, or nations interact are ubiquitous, and game theory provides the solution concepts of how the parties should act rationally. It is therefore of significant real-world importance to be able to compute (approximate) game-theoretic solutions to large games. Recreational games have often served as computational benchmarks for this purpose.

In perfect-information games, *real-time subgame solving* techniques have been the core of game-solving algorithms since the beginning of AI. Algorithms like minimax, alpha-beta search, or Monte Carlo tree search, explore the subgame tree from the single current state of the game forward [Shoham and Leyton-Brown, 2008; Silver *et al.*, 2018].

However, most real-world settings are imperfect-information games. There are multiple real states consistent with a single decision point, and changing a part of an agent's strategy typically changes what that agent's optimal strategy is later *and earlier* in the game. Thus, a subgame cannot be solved based solely on information from that subgame.

Therefore, unlike in perfect-information games where subgame solving was a core technique from the outset, imperfect-information games where approached as a monolithic whole for the longest time. Real-time subgame solving started relatively recently [Gilpin and Sandholm, 2006]. Since then, a host of better and better subgame solving techniques have been developed [Gilpin and Sandholm, 2007; Burch *et al.*, 2014; Ganzfried and Sandholm, 2015; Moravčík *et al.*, 2016; Moravčík *et al.*, 2017; Brown and Sandholm, 2017; Brown and Sandholm, 2018; Brown and Sandholm, 2019; Zhang and Sandholm, 2021; Kovařík *et al.*, 2023; Schmid *et al.*, 2023; Zhang and Sandholm, 2026]. In fact, subgame solving has made the central difference in achieving superhuman play in games such as two-player no-limit Texas hold'em poker [Brown and Sandholm, 2018], multiplayer no-limit Texas hold'em [Brown and Sandholm, 2019], and dark chess (aka. Fog-of-War chess) [Zhang and Sandholm, 2026].

A common approach is to compute a coarse *blueprint strategy* for each agent for the entire game, and then to conduct subgame solving on the fly to refine the blueprint strategy. *Gadget games* were developed to compute robust strategies in the subgames. These are modifications of the subgame that allow the opponent to adversarially choose a strategy it could have played in the past. The resulting optimal strategy of the player in such a gadget game is robust against any past strategy of the opponent [Burch *et al.*, 2014; Moravčík *et al.*, 2016; Brown and Sandholm, 2017]. However, existing gadget constructions, such as *resolving* and *max-margin*, only ensure that the player's strategy, in the worst case, is no worse than the blueprint. As a result, if the blueprint is poor, all of these methods may yield poor results.

We make the following contributions. 1. We show that different equilibria, which all yield the same value in the gadget game, have vastly different performance (exploitability) in the full game, when combined with the blueprint. We introduce a weaker notion of sequential equilibrium, *gadget game sequential equilibrium*, which is easier to compute, and propose that gadget games should be solved according to this concept. 2. We disprove the widely held belief that max-margin gadget games are strictly superior to resolving gadget games. This provides a new understanding of some prior works that have used the resolving gadget game for subgame solving rather than max-margin [Moravčík *et al.*, 2017; Zhang and Sandholm, 2026]. 3. We introduce efficient modifications to the *sequence-form linear program (SQF)* and *counterfactual regret minimization (CFR)* that converge to these refined equilibria with minimal additional computational overhead. 4. We empirically evaluate different types of Nash equilibria in several standard benchmark games. The results show that converging to gadget game sequential equilibria can reduce exploitability by more than 50% compared to vanilla SQF and CFR. Moreover, the results show that in resolving gadget games, these equilibria exhibit similar performance gains as using unsafe subgame solving, but provably without the risk of performing worse than blueprint.

2 Background and Notation

Two-player zero-sum factored-observation stochastic game is $\mathcal{G} = (\mathcal{N}, \mathcal{W}, w^{\text{INIT}}, p, \mathcal{A}, \mathcal{T}, \mathcal{R}_1, \mathcal{O})$, where $\mathcal{N} = \{1, 2, c\}$ is a set of players and c is a chance player, which plays fixed strategy and it is used to model stochasticity of the environment. $\mathcal{W}$ is the set of world states and $w^{\text{INIT}} \in \mathcal{W}$ is an initial world state. $p : \mathcal{W} \rightarrow 2^{\mathcal{N}}$ is a function that assigns acting players to a given world state w. $\mathcal{A} = \prod_{i \in \mathcal{N}} \mathcal{A}_i$ is the set of joint actions, we also use $\mathcal{A}_i(w) \subseteq \mathcal{A}_i$ to denote legal actions of i in state w. $\mathcal{T} : \mathcal{W} \times \mathcal{A} \rightarrow \mathcal{W}$ is a transition function and $\mathcal{R}_1 : \mathcal{W} \times \mathcal{A} \rightarrow \mathbb{R}$ is a reward function of player 1. In a zero-sum setting following holds $\mathcal{R}_2(w, a) = -\mathcal{R}_1(w, a)$. $\mathcal{O} : \mathcal{W} \times \mathcal{A} \times \mathcal{W} \rightarrow \mathbb{O}$ is the observation function and $\mathbb{O} : \mathbb{O}_0 \times \mathbb{O}_1 \times \mathbb{O}_2$ is the set of joint public observations and private observations. Similarly, observation function can be factored as $\mathcal{O} = (\mathcal{O}_0, \mathcal{O}_1, \mathcal{O}_2)$, where $\mathcal{O}_0$ is a public observation given to each player and $\mathcal{O}_1, \mathcal{O}_2$ are private observations available to their respective players [Kovařík *et al.*, 2022].

History $h = w^0 a^0 \dots a^{l-1} w^l \in (\mathcal{W}\mathcal{A})^* \mathcal{W}$ is a finite sequence of world states and actions, which starts in the initial state $w^0 = w^{\text{INIT}}$ and each timestep $t \in \{0, \dots, l-1\}$ is a valid transition $w^{t+1} = \mathcal{T}(w^t, a^t)$. $\mathcal{H}$ is a set of all possible histories in a game $\mathcal{G}$. We use $h \sqsubseteq h'$ to denote h' that contains h as a prefix and we will denote the initial history as $h^{\text{INIT}} := w^{\text{INIT}}$. Every history ends with some world state w^l, we will often use history in the game functions instead of that world state, for example $\mathcal{A}(h) := \mathcal{A}(w^l)$. The players do not observe the whole world state at each timestep, but they only observe public and their corresponding private observations. As a result, the player may not distinguish several different histories. s_i is an information set of the player, which is a set of all histories consistent with the observations of player i. $\mathcal{S}_i$ is then set of all information sets. We will overload the notation and use $s_i(h)$ to denote the information set corresponding to history h and $\mathcal{H}(s_i)$ as all histories consistent with information set s_i. Similarly, $s_0 \in \mathcal{S}_0$ is a public state, which is an information set of the external player, that does not have any private observations, so it contains all the histories consistent with public observations. As a result, each public state contains one or more information sets of each player and we will use $s_0(s_i)$ to denote the public state which contains information set s_i and $\mathcal{S}_i(s_0)$ as a set of all information sets consistent with s_0.

The behavioral strategy of the player i is $\pi_i : \mathcal{S}_i \rightarrow \Delta\mathcal{A}_i$ a mapping from information sets to probability distribution over actions. We will often use $\pi_i(s_i, a_i)$ as a probability that a_i is played in s_i, when following π_i. The strategy profile $\pi = (\pi_1, \pi_2)$ is a joint strategy profile of both players. For any two histories h, h', such that $h \sqsubseteq h'$, the probability of reaching h' from h under strategy profile π is $P^\pi(h'|h) = \prod_{h'' a w \sqsubseteq h'} \prod_{i \in \mathcal{N}} \pi_i(s_i(h''), a_i)$. Any reach probability can be factored into the components of the individual players $P^\pi(h'|h) = \prod_{i \in \mathcal{N}} P_i^\pi(h'|h)$. Sometimes we use $P^\pi(h) := P^\pi(h|h^{\text{INIT}})$.

The expected utility of history h if all players follow strategy profile π is $u_i^\pi(h) = \sum_{h \sqsubseteq h' a} P^\pi(h'|h) \mathcal{R}_i(h', a) \prod_{i \in \mathcal{N}} \pi_i(s_i(h'), a_i)$. Moreover, $v_i^\pi(s_j) = \frac{\sum_{h \in \mathcal{H}(s_j)} P_{-j}^\pi(h) u_i^\pi(h)}{\sum_{h \in \mathcal{H}(s_j)} P_{-j}^\pi(h)}$ are a counterfactual values. If $\sum_{h \in \mathcal{H}(s_j)} P_{-j}^\pi(h) = 0$, then $v_i^\pi(s_j) = 0$. The best response against a strategy π_i is a strategy $\pi_{-i}^{BR} \in BR_{-i}(\pi_i)$ that maximizes the opponents utility $\pi_{-i}^{BR} = \arg\max_{\pi_{-i}} u_{-i}^{(\pi_i, \pi_{-i})}(h^{\text{INIT}})$. The subscript $-i$ is used to identify the other player then i. When each player follows a best response strategy to each other, the resulting strategy profile is the Nash equilibrium π^* [Nash Jr, 1950]. In two-player zero-sum games, the Nash equilibrium is the desired solution concept and we use exploitability as a metric to evaluate quality of a strategy $\mathcal{E}(\pi_i) = u_{-i}^{\pi_i, \pi_{-i}^{BR}}(h^{\text{INIT}}) - u_{-i}^{\pi^*}(h^{\text{INIT}})$, which is how much can the opponent gain if it plays a best response compared to the value received by both players following Nash equilibrium. The exploitability is always non-negative and is zero if and only if the π_i is a part of Nash equilibrium.

2.1 Subgame Solving

There are two main approaches to subgame solving in imperfect-information games. The first, common-knowledge subgame solving, considers all histories that share the public state [Burch *et al.*, 2014; Moravčík *et al.*, 2017; Brown and Sandholm, 2018; Schmid *et al.*, 2023; Kubíček *et al.*, 2024]. The second, knowledge-limited subgame solving, considers only a subset of public state histories [Zhang and Sandholm, 2021; Liu *et al.*, 2023; Zhang and Sandholm, 2026]. Although our work focuses on common-knowledge subgame solving, the principles apply to both paradigms.

A subgame $\mathcal{G}^{s_0}$ of a game $\mathcal{G}$ is defined the same as the original game, but it has different initial world state w^{INIT, s_0}

in which the chance player chooses between all the histories $h \in \mathcal{H}(s_0)$, where s_0 is the public state which contains current decision. Assume all players followed a precomputed blueprint strategy $\overline{\pi}$, the chance distribution in the initial state of the subgame is proportional to $P^{\overline{\pi}}(h)$. This results in a valid game that can be solved with any algorithm for imperfect-information games.

Solving such a subgame $\mathcal{G}^{s_0}$ results in a strategy profile $\pi^{\mathcal{G}^{s_0}}$. We denote combined strategy as $\overline{\pi_i} \leftarrow \pi_i^{\mathcal{G}^{s_0}}$, where player i follows the new strategy $\pi_i^{\mathcal{G}^{s_0}}$ within the subgame and the blueprint elsewhere $\overline{\pi}$. This approach is known as unsafe subgame solving. While it often performs suprisingly well in practice [Gilpin and Sandholm, 2006; Gilpin and Sandholm, 2007; Ganzfried and Sandholm, 2015; Brown and Sandholm, 2017], it lacks theoretical guarantees [Burch *et al.*, 2014; Ganzfried and Sandholm, 2015; Kovařík *et al.*, 2023]. The unsafety stems from assuming the opponent followed a fixed strategy in the past. However, if the opponent has played according to a different strategy, the game will look exactly the same, but the computed solution would perform vastly differently.

To rectify this, gadget games were developed to compute robust strategies [Burch *et al.*, 2014; Moravčík *et al.*, 2016; Brown and Sandholm, 2017]. The gadget game $\mathcal{G}_i^{G,s_0}$ is similar to $\mathcal{G}^{s_0}$, but instead of assuming the opponent followed the blueprint $\overline{\pi}$, the gadget game introduces auxiliary information sets $\mathcal{S}^G$ that allow the opponent to choose their reach probabilities adversarially.

Definition 1. *Let $\mathcal{G}^{s_0}$ be a subgame and $\mathcal{G}_i^{G,s_0}$ be a gadget game of player i corresponding to $\mathcal{G}^{s_0}$.* Auxiliary information sets *$\mathcal{S}^G$ are all information sets that are in $\mathcal{G}_i^{G,s_0}$ but not in $\mathcal{G}^{s_0}$.*

Different types of gadget games introduce different auxiliary information sets $\mathcal{S}^G$. The *resolving gadget game* contains an auxiliary information set for each opponent's information set $s_{-i} \in \mathcal{S}_{-i}(s_0)$ in the root of the subgame, while the *max-margin gadget game* contains only a single auxiliary information set. In Figure 1 we show the construction of resolving and max-margin gadget games for a specific game. The *reach max-margin gadget game* has the same construction as max-margin, but uses different utility shifts [Brown and Sandholm, 2017].

The strategy resulting after solving the gadget game $\overline{\pi} \leftarrow \pi_i^{\mathcal{G}_i^{G,s_0}}$ is guaranteed to be safe, which means it does not increase the exploitability compared to that of the agent's blueprint strategy: $\mathcal{E}(\overline{\pi} \leftarrow \pi_i^{\mathcal{G}_i^{G,s_0}}) \leq \mathcal{E}(\overline{\pi})$.

3 Example

To demonstrate the issues with Nash equilibria in gadget games, we present a modified version of Matching Pennies in Figure 1a. The game begins with Player 2 choosing either heads (H) or tails (T). Player 1, unaware of this choice, must choose among three actions: Forfeit (F), Heads (H), and Tails (T). By forfeiting, player 1 loses with a reward of -1. Otherwise, Player 1 receives a reward of 1 if they match Player 2's action and 0 if they do not. This game has a unique Nash equilibrium in which both players play H and T with probability 0.5.

Assume a blueprint strategy for player 1, where $\overline{\pi_1} = [0.5, 0.5, 0]$ for actions F, H, T, respectively. Under this blueprint, the counterfactual values of the best response in the left and right branches of the tree are $v_1^{\overline{\pi_1}}(H) = 0$, $v_1^{\overline{\pi_1}}(T) = -0.5$.

In a resolving gadget game, Player 2 has the option to either enter the subgame (C) or terminate and receive the counterfactual best response value to the blueprint (T). The initial state of the subgame is decided by a chance node with probabilities proportional to the reaches $P_1^{\overline{\pi_1}} \cdot P_c$. These "probabilities" do not necessarily sum to 1. Figure 1b illustrates the extensive-form representation of this gadget game. Any strategy of Player 1 that forces Player 2 to play into a terminating action is a Nash equilibrium. Crucially, this includes even the blueprint, which plays suboptimal action F, which is strictly dominated.

The max-margin gadget game allows Player 2 first to select an information set before the chance node determines the specific state based on the reaches $P_1^{\overline{\pi_1}} \cdot P_c$. This construction also subtracts the counterfactual best response values from each terminal utility. Figure 1c shows the extensive-form representation of this game. In this specific case, there is a unique Nash equilibrium $\pi_1 = [0, 0.75, 0.25]$, which is not part of a Nash equilibrium of the whole game.

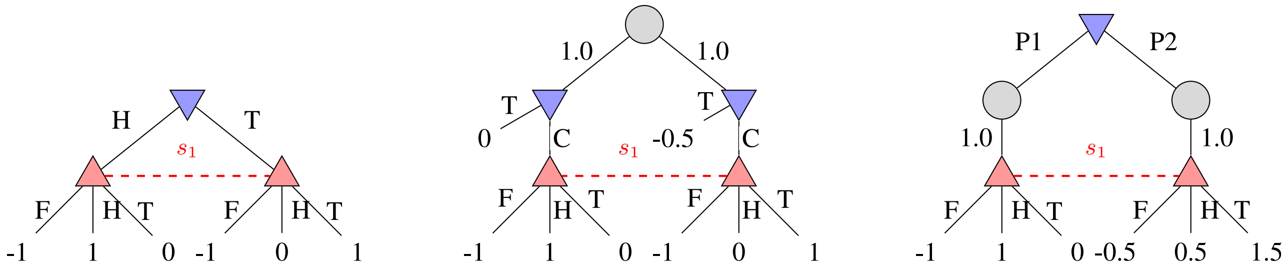
This example already hints at several of our results, which we discuss in more detail with more complicated examples in Appendix A. First, gadget games can have infinitely many equilibria that yield different values in the original game (Observation 1). Second, with some blueprint strategy, one might not be able to construct the optimal *continuation strategy* (Observation 2). Third, some equilibria in resolving gadget games can outperform all Nash equilibria of a max-margin gadget game (Observation 3).

4 Gadget Game Sequential Equilibria

The goal of subgame solving is to find a local strategy $\pi_i^{\mathcal{G}^{s_0}}$ that ensures the updated strategy improves upon the blueprint strategy $\overline{\pi_i}$. This improvement is formally characterized in Theorem 1 (inspired by [Moravčík *et al.*, 2017]), which shows that the resulting exploitability depends on the counterfactual best response values in the root information sets of the opponent.

Theorem 1. *Given a blueprint strategy $\overline{\pi_i}$, a subgame $\mathcal{G}^{s_0}$, and a subgame strategy $\pi_i^{\mathcal{G}^{s_0}}$, let $\pi_i = \overline{\pi_i} \leftarrow \pi_i^{\mathcal{G}^{s_0}}$. Let $\overline{\pi_{-i}^{BR}} \in BR_{-i}(\overline{\pi_i})$ and $\pi_{-i}^{BR} \in BR_{-i}(\pi_i)$. If $v_i^{(\pi_i, \pi_{-i}^{BR})}(s_{-i}) \geq v_i^{(\overline{\pi_i}, \overline{\pi_{-i}^{BR}})}(s_{-i})$ for each $s_{-i} \in \mathcal{S}_{-i}(s_0)$, then $u_i^{(\pi_i, \pi_{-i}^{BR})}(h^{INIT}) - u_i^{(\overline{\pi_i}, \overline{\pi_{-i}^{BR}})}(h^{INIT}) \geq \sum_{s_{-i} \in \mathcal{S}_{-i}(s_0)} P_{-i}^{\pi_{-i}^{BR}}(s_{-i}) \big(v_i^{(\pi_i, \pi_{-i}^{BR})}(s_{-i}) - v_i^{(\overline{\pi_i}, \overline{\pi_{-i}^{BR}})}(s_{-i})\big)$*

The proof is in Appendix A. Any Nash equilibrium of the resolving gadget game ensures that exploitability cannot increase by satisfying $v_i^{(\pi_i, \pi_{-i}^{BR})}(s_{-i}) - v_i^{(\overline{\pi_i}, \overline{\pi_{-i}^{BR}})}(s_{-i}) \geq 0$. In a max-margin gadget game, any Nash equilibrium maximizes



(a) Original game

(b) Resolving gadget game

(c) Max-margin gadget game

Figure 1: Example game, that highlights the problem with sequential rationality in subgame solving. Player 1 (red) is maximizing the value, while Player 2 (blue) is minimizing. Chance player (grey) is playing according to a shown fixed "probability" distribution.

the smallest improvement in these counterfactual values.

$$\max_{\pi_i^{\mathcal{G}^{s_0}}} \min_{s'_{-i} \in \mathcal{S}_{-i}} v_i^{(\overline{\pi_i} \leftarrow \pi_i^{\mathcal{G}^{s_0}}, \pi_{-i}^{BR})}(s'_{-i}) - v_i^{(\overline{\pi_i}, \overline{\pi_{-i}^{BR}})}(s'_{-i})$$

Because this minimal improvement is always at least zero, every solution of the max-margin gadget game is also a solution of the corresponding resolving gadget game.

Proposition 1. *Every Nash equilibrium of max-margin gadget game is also a Nash equilibrium of the corresponding resolving gadget game.*

The max-margin gadget game can be further improved by allowing the player to pay back the "gifts" that the opponent has given through mistakes so far [Brown and Sandholm, 2017]. However, this typically requires knowledge outside of the current subgame. We demonstrate that strategies can be improved without additional information by accounting for sequential rationality.

Observation 1. *There exists a game $\mathcal{G}$, a subgame $\mathcal{G}^{s_0}$ of $\mathcal{G}$, and blueprint $\overline{\pi_i}$ of player i, with corresponding resolving or maxmargin gadget game $\mathcal{G}^{G,s_0}$, in which different Nash equilibria result in different exploitability in the original game.*

We provide an example for this and all future observations in Appendix A. In two-player zero-sum games, all Nash equilibrium yield the same unique value of the game [Kuhn and Tucker, 1951]. This is true even in a gadget game. However, thanks to Observation 1, we see that in the gadget game, not all equilibria are equal, as some lead to lower exploitability when used in the whole game. This is because the exploitability of the global strategy does not depend on the value in the root of the gadget game $\mathcal{G}^{G,s_0}$, but on the values in the information sets where the subgame $\mathcal{G}^{s_0}$ begins.

Some Nash equilibria may play irrationally in parts of the gadget game that have zero reach. However, those same parts may be reached with high probability in the original game even if the opponent is playing an equilibrium of the whole game. Therefore, ensuring rational play even in unreachable nodes would be an improvement.

There has been a long line of work studying different Nash equilibria refinements in extensive-form games that ensure rationality in unreachable parts of the game [Kreps and Wilson, 1982; Selten, 1988; Van Damme, 1984; Van Damme, 2012]. The common problem with many of those refinements is that their computation necessitates playing each action with a small non-zero probability, and that the mistakes by the players should not be correlated. These conditions often make the computation numerically unstable. We observe that for gadget games, we do not need the full complexity of these refinements. Since the discrepancy between the gadget and the original game is in the auxiliary information sets, we only need to ensure rationality at the transition from the gadget game to the subgame.

Both the max-margin and resolving gadget game only add a single auxiliary information set to each trajectory, so it is enough to force at most one decision by the opponent to be played with non-zero probability in each trajectory. This means that the desired solution concept in gadget games is a weaker notion than some other refinements studied in the past [Van Damme, 2012], which simplifies their computation.

Definition 2. *Let $\mathcal{S}_{-i}^G$ be the opponent's auxiliary information sets in the gadget game $\mathcal{G}_i^{G,s_0}$, and let $\pi^P(s_{-i}^G, a_{-i}) > 0$ be a prior strategy $s_{-i}^G \in \mathcal{S}_{-i}^G$.* Gadget game sequential equilibrium (GGSE) *is a Nash equilibrium of a Gadget game $\mathcal{G}_i^{G,s_0}$, where every action $a_{-i} \in \mathcal{A}_{-i}(s_{-i}^G)$ is played with a probability greater than $\epsilon \cdot \pi^P(s_{-i}^G, a_{-i})$ as $\epsilon \to 0^+$.*

Gadget game sequential equilibria are a subset of Nash equilibria. They play rationally also in information sets to which the opponent plays with probability 0 from the auxiliary information sets.

Theorem 2. *Assume a subgame $\mathcal{G}^{s_0}$ and the corresponding gadget game $\mathcal{G}_i^{G,s_0}$. Let $\overline{\pi}$ be a blueprint strategy. For every Nash equilibrium π^* in the gadget game $\mathcal{G}_i^{G,s_0}$, there exists a gadget game sequential equilibrium π^G, that does not have lower counterfactual best response values in each root information set $s_{-i} \in \mathcal{S}_{-i}$:*

$$\min_{\pi_{-i}} v_i^{\overline{\pi_i} \leftarrow \pi_i^*, \pi_{-i}}(s_{-i}) \leq \min_{\pi_{-i}} v_i^{\overline{\pi_i} \leftarrow \pi_i^G, \pi_{-i}}(s_{-i})$$

Corollary 1. *If the optimal continuation strategy is a Nash equilibrium of a gadget game, then it is also gadget game sequential equilibrium.*

Thanks to Theorem 2 and Corollary 1, it is enough to search the space of gadget game sequential equilibria without the risk of excluding equilibria that would result in lower exploitability.

5 Choice of Prior

While GGSE ensures rationality in the subgame, the performance of the resulting strategy is influenced by the choice of the prior π^P. We first define the ideal target, which is the optimal strategy player i could play in the subgame, without changing strategy in the past.

Definition 3. *Let $\mathcal{G}^{s_0}$ be a subgame of game $\mathcal{G}$. Also, let $\hat{\pi}_i$ be a blueprint strategy, but only in the information sets $\mathcal{S}'_i$, which preceded some information set from $\mathcal{S}_i(s_0)$. The* optimal continuation strategy *$\hat{\pi}^*_i$ satisfies*

$$\hat{\pi}^*_i = \arg\max_{\pi_i} \min_{\pi_{-i}} u_i^{(\pi_i \leftarrow \hat{\pi}_i, \pi_{-i})}(h^{INIT}) \tag{1}$$

There could be multiple optimal continuation strategies, and ideally, we would like to find a strategy that is part of some optimal continuation strategy $\hat{\pi}^*_i$. Finding such a strategy naively requires solving the whole game with a fixed strategy of player i before reaching the subgame. A more scalable approach is to use a trained value function outside of the path to the subgame, which is called "full gadget" by Milec *et al.* [2024]. However, this approach requires a value function, which is difficult to train and grows linearly with the length of the game, making it less tractable than the resolving or max-margin gadget games.

Observation 2. *There exists a game $\mathcal{G}$, a subgame $\mathcal{G}^{s_0}$ of $\mathcal{G}$, and blueprint strategy $\overline{\pi}$ of player i, with corresponding resolving or maxmargin gadget game $\mathcal{G}^{G,s_0}$, in which no Nash equilibrium is part of the optimal continuation strategy $\hat{\pi}^*_i$*

Observation 2 demonstrates that when using the resolving or max-margin gadget game, the optimal continuation strategy may not be reconstructible.

In practice, many prior works have used unsafe resolving, which assumes the opponent followed the blueprint in the past [Gilpin and Sandholm, 2006; Gilpin and Sandholm, 2007; Ganzfried and Sandholm, 2015; Brown and Sandholm, 2017]. Although it often yields stronger performance in practice than prior safe subgame solving techniques, it runs the risk of producing strategies that are more exploitable than the blueprint. We propose using the blueprint as the prior, which combines the best of both worlds, as the strategy will always be safe, but it will leverage the information from the blueprint.

Specifically, for a resolving Player 1, we set the prior of auxiliary information set s_2^G and action a, which lead to information set s_2 in the root of the subgame to be proportional to the reach of s_2.

$$\pi_2^P(s_2^G, a_2) \propto P_2^{\overline{\pi_2}}(s_2)$$

The blueprint may play some actions with probability 0, so in our experiments, we clip the value to $\pi_2^P(s_2^G, a_2) \geq 10^{-3}$.

Using the opponent's part of the blueprint in subgame solving was explored in the poker AI Deepstack, where it was used as a warm-start strategy in the root of the gadget game [Moravčík *et al.*, 2017]. Contrary to our approach, they used a blueprint only in the first iteration, so their approach did not generally converge to GGSE.

6 Superiority of the Resolving Gadget Game

There has been a long-standing consensus that the max-margin gadget game is an improvement over the resolving gadget game. However, the authors of the AI poker system Deepstack noted that resolving gadget game performed better in early experiments [Moravčík *et al.*, 2017]. Similarly, the Obscuro AI agent for dark chess solves both the resolving and max-margin gadget games, and then heuristically chooses one of those two [Zhang and Sandholm, 2026].

Previously, these discrepancies were often attributed to the use of imperfect value functions in nested subgame solving. The logic was that if the value function is inaccurate, the safety constraints from Theorem 1 might be impossible to satisfy, leading to more exploitable strategies.

We offer another explanation that shows the max-margin gadget game can be outperformed by resolving even in endgame situations or when using a perfect value function. We show that the max-margin gadget game with a poor blueprint may eliminate some equilibria that are still obtainable by the resolving gadget game. Moreover, the max-margin gadget game may remove all equilibria that are part of the optimal continuation strategy $\hat{\pi}^*_i$.

Observation 3. *There exists a game $\mathcal{G}$, a subgame $\mathcal{G}^{s_0}$ of $\mathcal{G}$, and a blueprint strategy $\overline{\pi_i}$ of player i, in which some Nash equilibria of the corresponding resolving gadget game are less exploitable than any Nash equilibrium in the corresponding max-margin gadget game.*

The reason for this performance degradation is that in the max-margin gadget game, the opponent chooses at the root which information set it should play into based on how much improvement the player can get there, which introduces a bias to any Nash equilibrium strategy. In contrast, the resolving gadget game does not exhibit this bias toward certain parts of the game based on the blueprint. Therefore, when using an informed prior, such as the blueprint, the resolving gadget game often outperforms the max-margin.

7 Converging to Gadget Game Sequential Equilibria

A key advantage of GGSE is that it can be computed using standard tabular algorithms with only minor modifications. We describe how to modify the sequence-form linear program and counterfactual regret minimization to converge to GGSE. Both of these approaches are based on the idea of assuming the opponent makes small mistakes in the auxiliary information sets $\mathcal{S}^G$

7.1 Sequence-Form Linear Program

Consider the SQF for finding Player 1's part of the Nash equilibrium [Koller *et al.*, 1996]:

$$\max_{\boldsymbol{\pi_1}, \boldsymbol{v_2}} \boldsymbol{f_2}^\top \boldsymbol{v_2} \tag{2a}$$

$$\boldsymbol{F_1}\boldsymbol{\pi_1} = \boldsymbol{f_1}; \quad \boldsymbol{A}^\top \boldsymbol{\pi_1} - \boldsymbol{F_2}^\top \boldsymbol{v_2} \geq \boldsymbol{0}; \quad \boldsymbol{\pi_1} \geq \boldsymbol{0} \tag{2b}$$

Now consider a gadget game $\mathcal{G}^{G,s_0}$ and a perturbation vector $\boldsymbol{p} \in \mathbb{R}^{|\mathcal{S}_2|}$ that is zero everywhere except for sequences that end with action in the auxiliary information set

$a_2 \in \mathcal{A}_2(s_2^G)$, where it contains the prior $\pi_2^P(s_2^G, a_2)$. The following linear program approximates GGSE for a sufficiently small ϵ:

$$\max_{\boldsymbol{\pi_1}, \boldsymbol{v_2}} (\boldsymbol{f_2}^\top + \epsilon \boldsymbol{p}^\top) \boldsymbol{v_2} \tag{3a}$$

$$\boldsymbol{F_1} \boldsymbol{\pi_1} = \boldsymbol{f_1}; \quad \boldsymbol{A}^\top \boldsymbol{\pi_1} - \boldsymbol{F_2}^\top \boldsymbol{v_2} \geq \boldsymbol{0}; \quad \boldsymbol{\pi_1} \geq \boldsymbol{0} \tag{3b}$$

This formulation resembles the LP used to find one-sided quasi-perfect equilibria [Farina and Sandholm, 2021]. It similarly does not require an arbitrarily small ϵ, but it suffices to solve for a small ϵ. If the solution remains optimal when ϵ is set to 0, then the solution is the desired equilibrium. Otherwise, the ϵ decreases, and the program is solved again. However, this LP presents two key distinctions over one-sided quasi-perfect equilibria. First, the trembles are assumed only in auxiliary information sets $\mathcal{S}_2^G$ rather than every decision point. Second, within the same decision points, two actions may be played with different minimal probabilities, depending on the prior π^P.

7.2 Counterfactual Regret Minimization

The majority of successful subgame-solving applications have used counterfactual regret minimization (CFR), which iteratively minimizes counterfactual regret at each decision node. To approximate GGSE with CFR, we adopt the approach of Farina *et al.* [2017], which modifies the Regret Matching within CFR framework by introducing the concept of basis matrix $\boldsymbol{B}_{s_i} \in \mathbb{R}^{|\mathcal{A}_i(s_i)| \times |\mathcal{A}_i(s_i)|}$. Each entry in the matrix is non-negative, and each row and column sum to 1. This matrix serves as a transformation from the non-perturbed strategy to the perturbed one, and then transforms the regrets from the perturbed space to the non-perturbed one.

$$\pi_i'(s_i) = \boldsymbol{B}_{s_i} \pi_i(s_i) \tag{4}$$

$$r_i'(s_i) = \boldsymbol{B}_{s_i}^\top r_i(s_i) \tag{5}$$

We use this basis in auxiliary information sets s_2^G:

$$\boldsymbol{B}_{s_2^G} = (1-\epsilon)\boldsymbol{I} + \epsilon \boldsymbol{P} \tag{6}$$

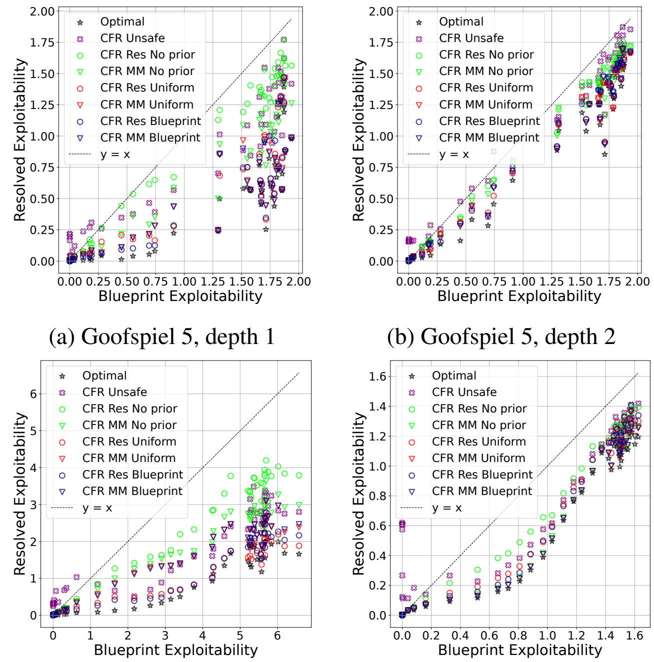
$\boldsymbol{I}$ is an identity and $\boldsymbol{P}$ is a symmetric matrix where a column corresponding to action $a_2 \in \mathcal{A}_2(s_2^G)$ contains $\pi_2^P(s_2^G, a_2)$.

This change does not significantly increase per-iteration time, but converging to an equilibrium may require more iterations based on the ϵ and $\boldsymbol{P}$, which we show in Section 8.

When using either the CFR or SQF, the resulting equilibrium is a GGSE, but it need not be consistent with the prior π^P. The GGSE is consistent with the prior only if the equilibrium itself does not play into any part of the subgame. The solutions of the max-margin gadget game wuill always be inconsistent, but those of the resolving gadget game can be consistent (exactly when all the counterfactual values are lower than the optimal ones).

8 Experiments

We evaluated the proposed algorithms for finding GGSE across three standard benchmark games: imperfect-information Goofspiel 5, Leduc hold'em, and Liar's dice 1,4. In each of those games, we generated 25 blueprints from different checkpoints of CFR. These serve as realistic approximations of blueprint strategies typically produced by abstraction-based methods or Monte Carlo CFR. We also generated 20 blueprints, in which the behavioral strategy in each information set was drawn independently from the Dirichlet distribution $\mathrm{Dir}(\boldsymbol{\alpha})$, with $\alpha_i = 1$.



(a) Goofspiel 5, depth 1

(b) Goofspiel 5, depth 2

(c) Leduc hold'em after flop

(d) Liar's dice 1,4, depth 1

Figure 2: Exploitability of the resolved strategy from each subgame based on blueprint's exploitability when using CFR as solver.

We solve subgames at specific decision points. In Goofspiel and Liar's dice, we solve every subgame following a fixed number of moves, and in Leduc hold'em, we solve all subgames after the public card is revealed. For each game and blueprint, we computed different types of subgame equilibria using the CFR modification discussed in Section 7.2, with CFR+ updates [Tammelin *et al.*, 2015]. We compare three different prior settings: no prior, uniform prior, or blueprint prior. For comparison, we also include the optimal continuation strategy and unsafe solving, which assumes both players play a fixed strategy before the subgame.

The relationship between the exploitability of the resolved strategy and the exploitability of the blueprint is shown in Figure 2. In Appendix B, we provide statistical analysis for our results and the same results when using SQF as a solver. We also show the convergence curves in Figure 3 for one of the blueprints to demonstrate the computational overhead in convergence when using a prior. We highlight the following results from this experiment.

- Incorporating the prior does not increase the exploitability and often decreases it. The gains are more significant when using the resolving gadget game.
- In many cases, the max-margin solutions outperform the resolving ones when no prior is used, in line with prior

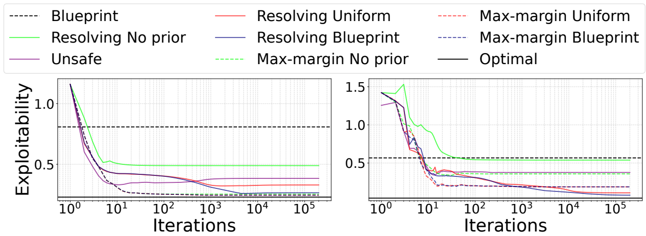


Figure 3: CFR Convergence curves of subgame solving techniques. Left is Goofspiel 5, depth 1, right is Liar's dice 1,4, depth 1

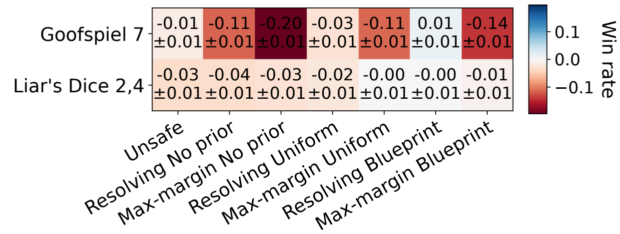


Figure 4: Head-to-head win-rate of each subgame solving technique against its 'worst-matchup' opponent out of the other techniques.

work. However, when using the prior, in Leduc hold'em, the solutions of the resolving gadget game outperform the max-margin ones, confirming that max-margin does not have to be superior to resolving.

- Although unsafe solving performs well when the blueprint is poor, its performance degrades as the blueprint improves. In contrast, when using the resolving gadget game with a blueprint prior, the solution achieves performance comparable to unsafe solving while being safe, effectively combining the strengths of both approaches.
- Converging to GGSE with CFR requires more iterations. Even with fewer iterations, the performance does not deteriorate compared to not using a prior at all, suggesting that using the prior does not come with any downside.

8.1 Nested Depth-Limited Subgame Solving

In large games, it is intractable to solve the whole game from the current decision point all the way to terminal states. The common approach to address this is to use nested depth-limited subgame solving, also known as midgame solving [Brown *et al.*, 2018; Brown and Sandholm, 2019; Kubíček *et al.*, 2024]. This approach trains a value function that is used after a fixed number of moves. In our experiments, we used a value function based on multi-valued states, which, after the depth limit, abstracts the entire rest of the game into a single decision by both players, in which they choose a strategy for the remainder of the game [Brown and Sandholm, 2019; Kovařík *et al.*, 2023].

We trained 10 different value functions in imperfect-information Goofspiel 7 and Liar's dice 2,4, both of which are large enough that exploitability cannot be computed efficiently, using the approach of Kubíček *et al.* [2024]. Then we matched different types of subgame solving at each encountered decision point. Head-to-head performance does not directly correspond to exploitability, because neither technique aims to exploit the opponent; rather, each approximates the Nash equilibrium of the entire game. Figure 4 shows the win-rate of each subgame solving technique against the worst matchup out of all the other techniques, which serves as a very limited approximation of exploitability. We provide the full head-to-head win-rates in Appendix B.

The results confirm that using the prior is beneficial even in head-to-head play. In Goofspiel, the unsafe solving performs well, while the solutions of the max-margin gadget game perform poorly, regardless of the prior. The situation in Liar's dice is the opposite. In both games, the solutions from the resolving gadget game with a blueprint prior consistently perform well across opponents.

9 Conclusions and Future Research

We studied gadget games, a subgame solving technique for imperfect-information games. While gadget games provide essential safety guarantees, they often contain infinitely many Nash equilibria. We showed that those equilibria may seem equivalent in the context of the gadget game, but when used in the original game, they lead to vastly different exploitability.

We proposed a new equilibrium refinement, specifically tailored to gadget games, *gadget game sequential equilibrium*, that improves upon vanilla Nash equilibria. We provided modifications to both the sequence-form linear program and counterfactual regret minimization to compute those refinements with negligible additional computational overhead.

We provided a novel explanation for the empirical superiority of the resolving gadget game over max-margin. We showed that the max-margin gadget game introduces an inherent bias into the solution, potentially excluding good equilibria that remain solutions of the resolving gadget game. We empirically verified that this occurs regularly in some games.

We compared different subgame-solving methods across multiple benchmarks and verified that converging to gadget game sequential equilibria consistently performs as well or better than both unsafe solving and the usual unrefined Nash equilibria to which CFR and the sequence form LP converge in gadget games. Thanks to improved theoretical understanding and empirical performance, this work establishes gadget game sequential equilibrium as the preferred solution concept for subgame solving in imperfect-information games.

Even though we established that there are games where neither the resolving nor the max-margin gadget game have the optimal continuation strategy as a solution, there is potential future work to either develop a new gadget game or to modify the existing ones so that they can reconstruct an optimal continuation strategy without the linear growth in size of the game as "full gadget". Another future research direction is to verify whether gadget game sequential equilibria present similar performance gains even when using the reach max-margin gadget game and/or when conducting knowledge-limited subgame solving. Also, there is potential to apply gadget game sequential equilibria to subgame solving in large recreational games like Texas hold'em and dark chess as well as to real-world game applications.

Acknowledgments

This work is supported by National Science Foundation grant RI-2312342 , the Vannevar Bush Faculty Fellowship ONR N00014-23-1-2876, the Czech Science Foundation GA25-18353S, the Grant Agency of the Czech Technical University in Prague (SGS23/184/OHK3/3T/13). The access to the computational infrastructure of the OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 "Research Center for Informatics" is also gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

References

[Brown and Sandholm, 2017] Noam Brown and Tuomas Sandholm. Safe and nested subgame solving for imperfect-information games. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 689–699, 2017.

[Brown and Sandholm, 2018] Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

[Brown and Sandholm, 2019] Noam Brown and Tuomas Sandholm. Superhuman AI for multiplayer poker. *Science*, 365(6456):885–890, 2019.

[Brown *et al.*, 2018] Noam Brown, Tuomas Sandholm, and Brandon Amos. Depth-limited solving for imperfect-information games. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 7674–7685, 2018.

[Burch *et al.*, 2014] Neil Burch, Michael Johanson, and Michael Bowling. Solving imperfect information games using decomposition. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'14, page 602–608. AAAI Press, 2014.

[Farina and Sandholm, 2021] Gabriele Farina and Tuomas Sandholm. Equilibrium refinement for the age of machines: the one-sided quasi-perfect equilibrium. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, pages 8845–8856, 2021.

[Farina *et al.*, 2017] Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Regret minimization in behaviorally-constrained zero-sum games. In *International Conference on Machine Learning*, pages 1107–1116. PMLR, 2017.

[Ganzfried and Sandholm, 2015] Sam Ganzfried and Tuomas Sandholm. Endgame solving in large imperfect-information games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 37–45, 2015.

[Gilpin and Sandholm, 2006] Andrew Gilpin and Tuomas Sandholm. A competitive Texas Hold'em poker player via automated abstraction and real-time equilibrium computation. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, AAAI'06, page 1007–1013. AAAI Press, 2006.

[Gilpin and Sandholm, 2007] Andrew Gilpin and Tuomas Sandholm. Better automated abstraction techniques for imperfect information games, with application to Texas Hold'em poker. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–8, 2007.

[Koller *et al.*, 1996] Daphne Koller, Nimrod Megiddo, and Bernhard Von Stengel. Efficient computation of equilibria for extensive two-person games. *Games and economic behavior*, 14(2):247–259, 1996.

[Kovařík *et al.*, 2022] Vojtěch Kovařík, Martin Schmid, Neil Burch, Michael Bowling, and Viliam Lisý. Rethinking formal models of partially observable multiagent decision making. *Artificial Intelligence*, 303:103645, 2022.

[Kovařík *et al.*, 2023] Vojtěch Kovařík, Dominik Seitz, Viliam Lisý, Jan Rudolf, Shuo Sun, and Karel Ha. Value functions for depth-limited solving in zero-sum imperfect-information games. *Artificial Intelligence*, 314:103805, 2023.

[Kreps and Wilson, 1982] David M Kreps and Robert Wilson. Sequential equilibria. *Econometrica: Journal of the Econometric Society*, pages 863–894, 1982.

[Kubíček *et al.*, 2024] Ondřej Kubíček, Neil Burch, and Viliam Lisý. Look-ahead search on top of policy networks in imperfect information games. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 4344–4352, 2024.

[Kuhn and Tucker, 1951] Harold William Kuhn and Albert William Tucker. Contributions to the theory of games (AM-24), Volume I, 1951.

[Liu *et al.*, 2023] Weiming Liu, Haobo Fu, Qiang Fu, and Yang Wei. Opponent-limited online search for imperfect information games. In *International Conference on Machine Learning*, pages 21567–21585. PMLR, 2023.

[Milec *et al.*, 2024] David Milec, Ondřej Kubíček, and Viliam Lisý. Continual depth-limited responses for computing counter-strategies in sequential games. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, pages 2393–2395, 2024.

[Moravčík *et al.*, 2016] Matej Moravčík, Martin Schmid, Karel Ha, Milan Hladik, and Stephen J Gaukrodger. Refining subgames in large imperfect information games. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 572–578, 2016.

[Moravčík *et al.*, 2017] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513, 2017.

[Nash Jr, 1950] John F Nash Jr. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[Schmid *et al.*, 2023] Martin Schmid, Matej Moravčík, Neil Burch, Rudolf Kadlec, Josh Davidson, Kevin Waugh, Nolan Bard, Finbarr Timbers, Marc Lanctot, G Zacharias Holland, et al. Student of games: A unified learning algorithm for both perfect and imperfect information games. *Science Advances*, 9(46):eadg3256, 2023.

[Selten, 1988] Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. In *Models of strategic rationality*, pages 1–31. Springer, 1988.

[Shoham and Leyton-Brown, 2008] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

[Silver *et al.*, 2018] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419):1140–1144, 2018.

[Tammelin *et al.*, 2015] Oskari Tammelin, Neil Burch, Michael Johanson, and Michael Bowling. Solving heads-up limit Texas Hold'em. In *Proceedings of the 24th International Conference on Artificial Intelligence*, pages 645–652, 2015.

[Van Damme, 1984] Eric Van Damme. A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. *International Journal of Game Theory*, 13(1):1–13, 1984.

[Van Damme, 2012] Eric Van Damme. *Refinements of the Nash equilibrium concept*. Springer Science & Business Media, 2012.

[Zhang and Sandholm, 2021] Brian Zhang and Tuomas Sandholm. Subgame solving without common knowledge. *Advances in Neural Information Processing Systems*, 34:23993–24004, 2021.

[Zhang and Sandholm, 2026] Brian Hu Zhang and Tuomas Sandholm. General search techniques without common knowledge for imperfect-information games, and application to superhuman Fog of War chess. In *International Conference on Learning Representations*, 2026.