

# Empirical Game Theoretic Analysis: A Survey

**Michael P. Wellman**  
*University of Michigan*  
*Ann Arbor, Michigan, USA*

WELLMAN@UMICH.EDU

**Karl Tuyls**  
*Meta AI*  
*Paris, France*

KARLTUYLS@META.COM

**Amy Greenwald**  
*Brown University*  
*Providence, Rhode Island, USA*

AMY@BROWN.EDU

## Abstract

In the empirical approach to game-theoretic analysis (EGTA), the model of the game comes not from declarative representation, but is derived by interrogation of a procedural description of the game environment. The motivation for developing this approach was to enable game-theoretic reasoning about strategic situations too complex for analytic specification and solution. Since its introduction over twenty years ago, EGTA has been applied to a wide range of multiagent domains, from auctions and markets to recreational games to cyber-security. We survey the extensive methodology developed for EGTA over the years, organized by the elemental subproblems comprising the EGTA process. We describe key EGTA concepts and techniques, and the questions at the frontier of EGTA research. Recent advances in machine learning are accelerating progress in EGTA, and promise to significantly expand our capacities for reasoning about complex game situations.

## 1. Introduction

When agents make decisions that interact with decisions of other rational entities, they play a *game*. Understanding how to play games effectively is a central concern of AI, as is anticipating the outcomes of games played by agents using AI methods. *Game theory* offers a rich conceptual and mathematical framework for describing and analyzing game situations [Leyton-Brown and Shoham, 2008]. As such, game-theoretic ideas and methods are ubiquitous in AI research, employed in theoretical treatments as well as computational approaches to reasoning about games. AI today is thus a major consumer of game theory, and also a significant contributor—to concepts, representations, algorithms, and more.

At the heart of game-theoretic analysis is a formal representation of the game situation: a *game model*. Classically (e.g., in game theory or AI textbooks), such models involve tables or trees, spelling out the agents' actions, information, and values, and the outcomes of the various decisions they might take. Modern developments, often under the label of *computational* or *algorithmic game theory* [Nisan et al., 2007], have significantly augmented these constructs in clever ways, for example by supporting compact expression of complex strategic environments. Given a game model, the logic of game theory can be applied to identify or characterize game-theoretic *solutions*:

configurations of agent behavior satisfying well-defined conditions representing criteria for rational behavior.

Requiring a game model in analytic form, however, poses an impediment to the application of game theory in many settings of interest for AI. In practice, formally specifying a multiagent scenario has proven feasible only for games that are artificially defined (e.g., recreational games in worlds of boards, cards, and dice) or are highly stylized representations of realistic situations. Applied game theorists can be quite adept at stylization: isolating the salient features of a strategic situation and capturing them in an analytic form that is tractable for game-theoretic representation and reasoning. Vast literatures document deep strategic insights about markets, conflicts, organizations, ecologies, and many other social domains—obtained through application of game-theoretic concepts to stylized models of such settings. Inevitably, however, game-theoretic conclusions obtained in this way hinge on simplifying assumptions, adopted by necessity, and thus their application to real-world instances entails judgment about the relevance of the complications abstracted away. Moreover, the necessities of stylization may be systematic in the kinds of actually relevant features that can be accommodated, and those that cannot.

An alternative is to express strategic environments *procedurally*, in the form of a *simulation model*. Simulation can readily accommodate important forms of complexity: for example, agent heterogeneity, and nonlinear dynamics generated through complicated state and information structures. The methodology of *empirical game-theoretic analysis* (EGTA) aims to combine the flexibility of simulation with the strategic logic of game theory. Simulating interacting agent decisions allows consideration of complex environments that would be difficult to express in an analytic game form, or that would be intractable for reasoning.

This motivation is similar to that of *agent-based modeling* (ABM), which emerged from the social sciences as a simulation-based alternative to mainstream economic theories [Miller and Page, 2007, Tesfatsion, 2006]. Pioneering advocates of ABM tended to eschew strict rationality and equilibrium assumptions, instead embracing tools from evolution and adaptive systems. However, ABM alone is in a sense too flexible, as the possible outcomes achievable through *some* agent behaviors can be extremely open-ended. The outcomes we more specifically care about are those that follow from *what rational agents would do*, according to some concept of rationality (including approximate or boundedly rational concepts) judged appropriate for our setting [Wellman, 2016]. This is exactly what game theory provides and what EGTA inherits: a mathematical framework with representations for strategic situations and concepts for characterizing and identifying implications of rational choice.

The core idea of EGTA is to employ agent-based simulation to generate data from which we can induce a game model, which we call the *empirical game*. A high-level diagram of the EGTA process is shown in Figure 1. The process of inducing an empirical game from agent-based simulation is shown in the upper right. That the fundamental game specification is in the form of a simulator affords a high degree of expressivity in a convenient manner. That the method produces a game model, tailored to the question at hand, in turn affords exploitation of a comprehensive algorithmic toolbox of computational game-theoretic techniques. The result is an expansion of the scope of game-theoretic reasoning beyond domains where analytic game models can be plausibly and usefully constructed.

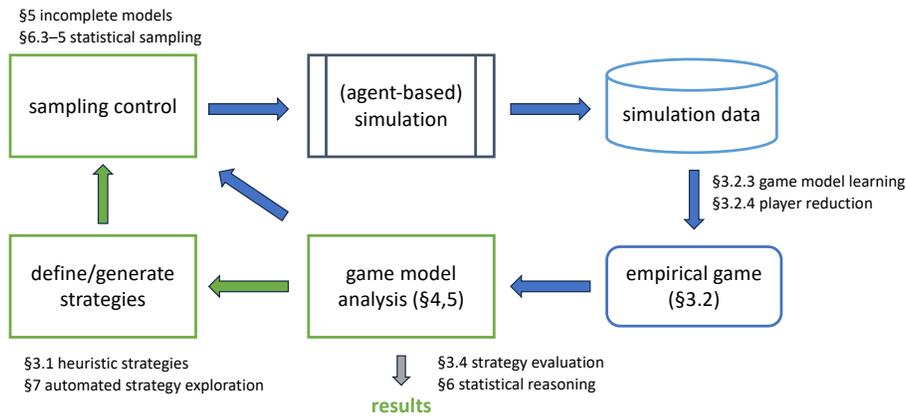


Figure 1: EGTA: A high-level view. The underlying game is represented by an agent-based simulator, depicted in the top row, center. Proceeding clockwise, data from the simulation is used to induce an empirical game model. The green-box modules show how we develop the empirical game and ultimately obtain results, through a repeated process of game model analysis, extension, and refinement. Section references indicate where key issues and techniques are surveyed herein.

The empirical game is extended and refined through an iterative process, where analysis of the game model drives identification of new strategies to consider and determination of which strategy profiles to simulate. The techniques by which these decisions are made based on interim analyses comprise the core of EGTA methodology, as described in this survey.

It is important to keep in mind that the empirical game is a model, and as such it departs in important ways from the fundamental game situation of interest, sometimes referred to as the *underlying game*. First, the simulator itself may only approximate this underlying game, as after all, a simulation model is still a model. Second, the empirical game typically covers only a strict subset of the possible agent strategies (i.e., those admissible by the simulator). It may also incorporate other abstractions or even structural modifications relative to the game defined by the simulator. For instance, an empirical game may be expressed in normal form, even though the strategies themselves (i.e., as implemented in the simulation) are typically highly sequential and conditional on implicit observations. Finally, the empirical game’s payoffs are induced from noisy and/or sparse simulation data, and so are subject to approximation error compared to the true payoffs of the underlying game. As we survey here, much of the EGTA literature is devoted to characterizing these departures and developing techniques to promote the fidelity of EGTA results given the inherent imperfections of empirical modeling.

The EGTA approach was named “empirical” because it embraces empirical methods: simulation, sampling, search, machine learning (ML), etc. [Wellman, 2006]. This contrasts with the prevalent mode of game-theoretic analysis at the time, which was based on deductive inference. Nowadays it is far more common to invoke techniques like machine learning in service of game-theoretic reasoning, beyond what we are here labeling EGTA. This trend is accelerating, for example as foundation models engender an entirely new form of procedural interrogation by which to construct game models.

This survey presents an organized view of EGTA methodology as developed in the first decades of this 21st century. The next section starts with some historical context and a presentation of the technical background in game theory, including evolutionary game theory, underpinning EGTA. Section 3 describes the core concepts defining the EGTA approach, and enabling implementation of EGTA. Concepts and methods associated with the evolutionary game perspective are presented in Section 4. Reasoning about incomplete game models is of particular relevance for EGTA, and the subject of Section 5. Statistical techniques for EGTA and works addressing statistical questions about empirical games are surveyed in Section 6. Section 7 introduces the idea of automated strategy generation, and discusses the PSRO framework which today is one of the most popular approaches to EGTA. EGTA applications and approaches to mechanism design are the topics of Sections 8 and 9, respectively. We conclude with a reflection on EGTA, discussing some promising future directions and fundamental challenges.

## 2. Background

### 2.1 History

To our knowledge, the concept of an empirical game defined by a constrained strategy set was first articulated by Armantier et al. [2000]. The earliest published instance we can identify of an explicit game model estimated by simulation of heuristic strategies appears in the PhD dissertation of William Walsh [2001, Chapter 6]. In that work, Walsh analyzed a game of supply chain formation [Walsh et al., 2000], where agents bid in a combinatorial auction to determine whether a chain forms and who participates. He used simulation to estimate payoffs for combinations of bidding strategies selected from a parametric family, and from those estimates identified approximate Nash equilibria among the evaluated strategy instances.

Around the same time, a group of IBM researchers developed some intriguing agent-based models of the emerging digital economy [Kephart et al., 1998, Tesauro and Das, 2001], including some expressly appealing to game-theoretic concepts [Greenwald and Kephart, 1999, Greenwald et al., 1999]. Walsh eventually joined this group, and together they studied pricing and bidding games with up to twenty agents and three heuristic strategies. Their paper [Walsh et al., 2002] was the first to explicitly advocate equilibrium-based analysis (both strategic and evolutionary) of game models empirically derived by simulation. They introduced the concept of *heuristic payoff table* as a representation of expected payoffs over a heuristic profile space.

Research following these early works branched off in two directions: the first focusing on strategic reasoning for simulation-based games, and the second focusing on evolutionary dynamical analysis of agent behavior inspired by evolutionary game theory. The overall methodology was given the name “EGTA”, and the strategic reasoning direction systematically developed in a program of sustained research at the University of Michigan [Wellman, 2006], shortly following the Walsh et al. [2002] paper. This began with a study of heuristic strategies for simultaneous ascending auctions [Wellman et al., 2003], which derived constrained equilibria for empirical games over selected parametric instances. A series of PhD dissertations over the next two decades advanced the methodology in a variety of directions. A significant thread of EGTA work from this group was driven by the Trading Agent Competition (TAC), a series of market games posed as challenges to the research

community [Collins et al., 2010, Wellman et al., 2007]. In these competitions, AI researchers developed innovative trading strategies for a variety of complex market environments. Since the strategies were typically developed independently by diverse groups focusing on different approaches, understanding the strategic interactions among them often required careful post-hoc analysis. For example, an empirical game study of strategic procurement in the TAC supply chain game provided insight into why the 2003 tournament was prone to bouts of ruinous price cutting [Wellman et al., 2005a]. EGTA was subsequently employed with some regularity in analyses of TAC tournaments [Jordan et al., 2007, 2010b], as well as other research competitions [Baarslag et al., 2013].

The second line of work was inspired by the discovery that canonical reinforcement learning (RL) algorithms, including learning automata [Narendra and Thathachar, 1989] and Q-learning [Watkins and Dayan, 1992], are formally related to the process of replicator dynamics from evolutionary game theory (discussed in depth in Section 4). The connections were illustrated by various authors, including Börgers and Sarin [1997], Sato and Crutchfield [2003], and Tuyls et al. [2002, 2003].<sup>1</sup> Agent-based research in the vein of evolutionary dynamics took off with the organization of the GTDT and EGTMAS workshops at AAMAS conferences starting in 1999 and 2003, respectively. For instance, Phelps et al. [2005] employ the evolutionary dynamics approach in an EGTA study comparing two alternative double-auction mechanisms. In follow-on work, this team showed how to derive a new strategy through genetic search over a parametric strategy space, optimizing performance against the equilibrium derived from an empirical game model [Phelps et al., 2006]. They further built on these ideas to consider basin size under an evolutionary dynamic as a fitness measure, and proposed an algorithm for extending the strategy space by repeated iteration [Phelps et al., 2010a].

At the same time, researchers from Maastricht University, building on the links between RL methods and evolutionary game theory [Tuyls et al., 2002, 2003], studied the evolutionary dynamics of heuristic strategies for Texas Hold'em Poker [Ponsen et al., 2009] and continuous double auctions [Tuyls and Parsons, 2007, Kaisers et al., 2008]. Later, these works were extended by Tuyls and colleagues at DeepMind, as they appealed to EGTA to analyze and evaluate the RL breakthroughs achieved in Go, Capture the Flag, StarCraft, and other games [Balduzzi et al., 2018, Tuyls et al., 2018a,b, 2020]. By providing a flexible template for combining game-theoretic reasoning with deep RL, the PSRO framework (see Section 7.2) developed by DeepMind researchers [Lanctot et al., 2017] spurred a significant expansion of interest in EGTA.

## 2.2 Technical Preliminaries

### 2.2.1 BASIC TERMINOLOGY AND NOTATION

Formally, a **normal-form game**  $\Gamma = \langle N, (S_i), (u_i) \rangle$  consists of a finite set of players,  $N$ , indexed by  $i$ ; a nonempty set of (pure) strategies  $S_i$  for player  $i$ ; and a utility function  $u_i : \prod_{j \in N} S_j \rightarrow \mathbb{R}$  for each player. Let  $n = |N|$  be the number of players.

Player  $i$  may play a **pure strategy**,  $s_i \in S_i$ , or a **mixed strategy**,  $\sigma_i \in \Delta(S_i)$ , which is a probability distribution over the pure strategies in  $S_i$ . Let  $\sigma_i(s_j)$  denote the probability that strategy  $s_j$  is played

---

1. Not coincidentally, some of the early EGTA works noted above also prominently employed replicator dynamics in their analyses [Walsh et al., 2002].

in  $\sigma_i$ . The **support**,  $\text{supp}(\sigma_i)$ , of a mixed strategy  $\sigma_i$  is the set of pure strategies played with positive probability:  $\text{supp}(\sigma_i) = \{s_j \in S_i \mid \sigma_i(s_j) > 0\}$ . A **strategy profile**  $\sigma = (\sigma_1, \dots, \sigma_n)$ , assigns a (generally mixed) strategy to each player. If the assignments are to pure strategies,  $s = (s_1, \dots, s_n)$  is a **pure profile**. We use notation  $\sigma_{-i} = (\sigma_1, \dots, \sigma_{i-1}, \sigma_{i+1}, \dots, \sigma_n)$  to denote a strategy profile for all players excluding  $i$ .

Thus,  $u_i(s_i, s_{-i})$  denotes player  $i$ 's **utility** for playing strategy  $s_i$  when the other players play  $s_{-i}$ , also termed  $i$ 's **payoff** for that situation. We extend  $u_i$  to mixed profiles by taking expectations over the realization of mixtures in pure profiles,

$$u_i(\sigma_i, \sigma_{-i}) \equiv \mathbb{E}_{s \sim \sigma_i, s_{-i} \sim \sigma_{-i}}[u_i(s_i, s_{-i})].$$

The aim of game-theoretic analysis is typically to characterize and identify **solutions** of a given game, according to a specified **solution concept**. Commonly employed solution concepts are based on notions of **equilibrium** among strategies. The classic concept of strategic equilibrium, due to Nash, selects profiles such that no player can benefit from a unilateral deviation. Define  $BR_i(\sigma_{-i})$  as player  $i$ 's **best response** when player  $-i$  plays strategy  $\sigma_{-i}$ :

$$BR_i(\sigma_{-i}) \equiv \arg \max_{\sigma_i \in \Delta(S_i)} u_i(\sigma_i, \sigma_{-i}).$$

Formally, profile  $\sigma^*$  is a **Nash equilibrium** (NE) if and only if (iff) for all  $i$ ,  $\sigma_i^* \in BR_i(\sigma_{-i}^*)$ . Equivalently,  $\sigma^*$  satisfies

$$\forall s'_i \in S_i. u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(s'_i, \sigma_{-i}^*).$$

For a non-NE profile  $\sigma$ , at least one player can benefit by **deviating** to an alternative strategy. Player  $i$ 's potential gain to deviation from  $\sigma$  in game  $\Gamma$  is termed their **regret**,  $\epsilon_i^\Gamma(\sigma)$ , and is given by

$$\epsilon_i^\Gamma(\sigma) = \max_{s'_i \in S_i} u_i(s'_i, \sigma_{-i}) - u_i(\sigma_i, \sigma_{-i}).$$

The **(game) regret** for a profile is the maximum over player regrets:

$$\epsilon^\Gamma(\sigma) \equiv \max_{i \in N} \epsilon_i^\Gamma(\sigma). \quad (1)$$

We drop the superscript  $\Gamma$  when the game is clear from context.

We can use the notion of regret to define approximate solution concepts. In particular,  $\sigma^*$  is an  $\epsilon$ -**Nash equilibrium** iff no player can gain more than  $\epsilon$  in payoff by deviating:

$$\forall i \forall s'_i \in S_i. u_i(\sigma_i^*, \sigma_{-i}^*) + \epsilon \geq u_i(s'_i, \sigma_{-i}^*).$$

Equivalently, a profile  $\sigma$  is an  $\epsilon(\sigma)$ -NE. Exact NE have zero regret.

Additional solution concepts are discussed in Section 2.2.4.

### 2.2.2 SYMMETRIC GAMES

A game is *anonymous* if all players have the same strategy set ( $\forall i, j. S_i = S_j = S$ ) and for all  $i$ ,  $u_i(s_i, s_{-i})$  is invariant to permutations of the other players ( $-i$ ). In other words, the utility function depends only on how many of the others play each strategy—not which ones. An anonymous game is further called *symmetric* if the utility functions are the same for every player. For symmetric games, we may drop the subscript on  $u$  and write  $\mathbb{E}[u(\sigma, \sigma')]$  for the expected utility of (any player) playing strategy  $\sigma$  when the remainder are playing according to other-player profile  $\sigma'$ .

A game is *role symmetric* if the players can be partitioned into roles  $R_1, \dots, R_K$ , such that symmetry applies within roles. That is, players within each role have the same strategy sets and utility functions, and the utility functions depend only on how many (other) players in each role play each strategy. Role symmetry is without loss of generality, with  $K = n$ ; with  $K = 1$ , the game is fully symmetric. Thus, role symmetry interpolates between these two extremes.

### 2.2.3 BIMATRIX GAMES

In the special case of two-player finite-strategy normal-form games, payoffs can be specified by a pair of matrices. Formally, a *bimatrix game* has  $n = 2$ , and is given by  $\Gamma = \langle N, (S_1, S_2), (u_1, u_2) \rangle$ , with the utility functions  $(u_1, u_2)$  represented by a pair of matrices  $(A, B)$  giving the payoffs for the respective players.

Figure 2 illustrates a two-strategy example, in which one player (dubbed the *row player*) chooses one of the two rows (each corresponding to a pure strategy), and the *column player* chooses a column (each corresponding to a pure strategy), with the combination determining their joint payoff.

$$\begin{pmatrix} a_{11}, b_{11} & a_{12}, b_{12} \\ a_{21}, b_{21} & a_{22}, b_{22} \end{pmatrix}$$

Figure 2: General payoff  $(A, B)$  for a two-action bimatrix game.

In case  $S_1 = S_2$  and  $A = B^T$ , the players are interchangeable and we have a symmetric game.

An example of an asymmetric bimatrix game is “Bach or Stravinsky” (BoS), illustrated in Figure 3. In this game, both players have the same strategy set ( $S_1 = S_2 = \{B, S\}$ ), choosing whether to go to a Bach ( $B$ ) or Stravinsky ( $S$ ) concert. They prefer to attend the same concert, however their payoffs differ, expressing divergent preferences between the two options.

$$\begin{array}{cc} & \begin{array}{cc} B & S \end{array} \\ \begin{array}{c} B \\ S \end{array} & \begin{pmatrix} 3, 2 & 0, 0 \\ 0, 0 & 2, 3 \end{pmatrix} \end{array}$$

Figure 3: Payoff matrix for the BoS game. Strategies  $B$  and  $S$  correspond to attending concerts of Bach or Stravinsky music, respectively.

#### 2.2.4 ADDITIONAL SOLUTION CONCEPTS

By far the most commonly employed solution concept in EGTA (and game-theoretic analysis more broadly) is Nash equilibrium, defined in Section 2.2.1. Game theorists have introduced numerous refinements of and alternatives to NE to address a variety of considerations, such as special game structure or bounded rationality [Gintis, 2009]. For example, *correlated equilibrium* generalizes Nash to allow strategy profiles with dependencies among player choices. *Quantal response equilibrium* is another generalization, designed to model approximately rational behavior. Such concepts are likewise applicable to empirical game models, as the choice of solution concept is generally orthogonal to whether the model is based on simulation data, or any other knowledge source.

Some solution concepts are defined with respect to structured game forms. For example, *subgame perfection* is a property of solutions applicable to *extensive-form games*. Considering such properties in EGTA would require that the empirical game model express that structure. Our survey focuses on normal-form representations, as that covers the majority of EGTA literature to date.<sup>2</sup>

The evolutionary perspective is also a rich source of solution concepts. These have played a salient role in EGTA methodology, including proposals for new solution concepts motivated in part by EGTA ( $\alpha$ -Rank, discussed in Section 4.3). We introduce technical background on evolutionary stability within a broader discussion of evolutionary game theory in the next section.

### 2.3 Evolutionary Game Theory

Canonical game-theoretic solution concepts, starting with the classic equilibrium condition proposed by Nash [1950], tend to be defined by a *static* relationship among player strategies. But from the earliest days, game theorists have sought to produce more *dynamic* accounts of how such equilibrium configurations might arise through player interactions. A notable example is the method of *fictitious play* (FP), introduced by Brown [1951], which defines an iterative process where each player's strategy at a given time step is a best response to its belief about the play of others, based on past iterations. In the standard version, the belief is that each strategy is played with probability equal to its frequency in past play. Various analyses over the years have identified conditions for which FP and variants are guaranteed to converge (and to what), and studied FP's performance as a game-solving heuristic.

Biological evolution has been a particularly rich source of ideas about dynamically adapting behavior. The field of *evolutionary game theory* (EGT) applies such ideas to strategic interaction, building dynamic accounts of adaptive game play [Börgers and Sarin, 1997, Tuyls et al., 2003, Tuyls and Parsons, 2007], based on biological operators such as natural selection and mutation [Maynard Smith and Price, 1973, Zeeman, 1980, 1981, Weibull, 1997, Hofbauer and Sigmund, 1998]. The simplicity and concreteness of these operators provides a constructive basis for determination of joint behavior in complex strategic environments. As evolutionary computation meshes well with agent-based simulation, a simulation-based approach to game theory (i.e., EGTA) is naturally suited to incorporate EGT principles and techniques. We lay out some of these techniques in Section 4, demonstrating the role of evolutionary concepts in EGTA methodology.

---

2. We note where relevant some emerging works that extend EGTA beyond normal form.

## 2.4 Running Example

To illustrate the concepts and methods of EGTA, we present an example game that can be studied with this approach. The domain is *bidding in sequential auctions*. Like many auction games that have been tackled with EGTA, the setting is descriptively simple but strategically complex—beyond the realm of analytic tractability except in well-crafted special cases. In its general form, the game comprises a series of one-shot sealed-bid auctions, each for a single good. In each auction, the players submit bids, representing positive amounts of a standard currency. The good is awarded to the highest bidder (i.e., *winner* of the auction), who pays an amount that is a function of their own and the other players' bids, as dictated by specified auction rules. The auction then reveals information about the result (e.g., the identity of the winner and price paid), which the players may use in determining their bids in subsequent auctions. At the end of the series, players receive a payoff, namely the difference between their value for the combination of goods won and their total payment for these goods.

Greenwald et al. [2012] provide a formal specification of this game along with some insightful theoretical observations and a constructive computational approach. For our purposes, we may make do with the following notation:

- $G$  is the set of goods, indexed according to the auction sequence,  $g = |G|$ .
- $\mathbb{1}_{i \text{ wins } j}$  is an indicator function, 1 if player  $i$  wins auction  $j$  and 0 otherwise.
- $p_j$  is the price (i.e., payment) outcome from auction  $j$ .
- $Y_{ij} = \{k \mid \mathbb{1}_{i \text{ wins } k} = 1\}$  is the set of goods won by player  $i$  in the first  $j$  auctions.
- **valuation**  $v_i : 2^G \rightarrow \mathfrak{R}$  is a function that describes player  $i$ 's value for obtaining any set of goods.
- $V_i$  is the set of possible valuations  $v_i$ .
- $h_{ij}$  is the history of observations received by player  $i$  from auctions  $1, \dots, j$ .
- $H_{ij}$  is the set of possible histories  $h_{ij}$ .

The payoff to player  $i$  is determined by auction outcomes: goods won  $Y_{ig}$  and their prices. To describe the utility function in terms of strategies, we must account for the uncertainty due to incomplete and imperfect information. Auctions are Bayesian games, in that players know their own valuations but have only probabilistic information about the others'. Sequential auctions also feature imperfect information, as players do not fully observe the others' actions. As the sequence of auctions proceeds, the players receive partial evidence about the other bids, summarized in auction results. A player's bid in any given situation, therefore, depends on the player's valuation as well as their observations up to that point. Formally, the strategy set for player  $i$  is given by  $S_i : V_i \times H_{ij} \rightarrow \mathfrak{R}$ .

A complete game specification would include probability distributions over valuations (if these distributions are the same for every player, the game is symmetric). Given such distributions, we

can write the utility function as an expectation over the auction outcomes:

$$u_i(s_i, \sigma_{-i}) = \mathbb{E}_{p_j, \mathbb{1}_{i \text{ wins } j} | s_i, \sigma_{-i}} \left[ v_i(Y_{ig}) - \sum_j p_j \mathbb{1}_{i \text{ wins } j} \right].$$

### 3. EGTA: Key Concepts

We are now ready to introduce EGTA’s defining concepts, most importantly, that of an empirical game model induced by simulation over a restricted strategy space. We illustrate the definition of a space of heuristic strategies (Section 3.1), and consider a range of issues for the construction of empirical games (Section 3.2). The remaining subsections raise issues for game solving, evaluating strategies, and game visualization that have been addressed within the EGTA approach.

#### 3.1 Heuristic Strategies and Restricted Games

All game analysis is with respect to some set of included strategies  $X_i$  for each player  $i$ . This set is typically a strict subset of all possible strategies, since these are generally infeasible to cover. We use the symbol  $\Gamma_{\downarrow X}$  to denote a **restricted game** with respect to the **base game**  $\Gamma$ , where each player  $i$  in  $\Gamma_{\downarrow X}$  is restricted to playing strategies in  $X_i \subseteq S_i$ , with  $X = \prod_{i \in N} X_i$ .

Restricted strategy sets may be formed in a variety of ways. One basic approach is to design **heuristic strategies**, based on domain knowledge and intuitive simplifications, or perhaps based on known benchmarks. We often specify a parameterized space of such strategies by exposing some controllable features of the heuristics. Another approach that yields a parametric strategy space is to specify a more generic representation, such as a neural network policy implementation. This parametric space is itself a restriction, as not all strategies in the base game may be expressible as parameter settings. The actual set of included strategies may be further restricted with respect to this space, by either manual or automated selection.

For example, consider a heuristic strategy for bidding in sequential auctions based on **myopic marginal value**. The strategy must specify how to bid for good  $j + 1$ , given one’s valuation and the results from the first  $j$  auctions. The myopic marginal value for this good is the increase in value it provides, given current winnings and assuming no further winnings:  $\mu_{i, h_{ij}} \equiv v_i(Y_{ij} \cup \{j + 1\}) - v_i(Y_{ij})$ , for  $j \in \{0, \dots, g - 1\}$ . One possible strategy is to bid the myopic marginal value in every auction. A simple parameterized extension is to bid a constant fraction of this value, that is, bid  $\rho \mu_{i, h_{ij}}$  in each auction  $j + 1$ , for some  $0 < \rho \leq 1$ . We refer to  $\rho$  as a strategy parameter—in this case, a shading factor—and note that a parameterized heuristic strategy plus a set of allowed parameter settings defines a set of strategies. Let  $s^\rho$  denote the strategy of shading myopic marginal value by fraction  $\rho$  as described above.  $X^P = \{s^\rho \mid \rho \in P\}$  would then comprise a set of strategies corresponding to a set of possible parameter settings  $P$ . Such strategy sets could, in turn, define a restricted game—for instance, if  $\Gamma$  is a game of sequential auctions, then a corresponding game restricted to myopic marginal value bidding could be written  $\Gamma_{\downarrow \{X^P_i\}}$ , where  $P_i$  is the set of shading factors allowed for player  $i$ .

Though myopic marginal value is limited as a strategy, analyzing a restricted game among these strategies could provide useful strategic insights about sequential auctions. Such an analysis is directly relevant in cases where the restricted strategies are representative of real-world behavior. More generally, it provides a baseline for gauging the benefit of incorporating more sophisticated reasoning into the bidding, most naturally transcending myopia by accounting for the opportunities of future auctions. By incrementally extending the restricted strategy set, we can refine the strategic analysis, ultimately approaching the base game.

### 3.2 Empirical Game Models

The hallmark of EGTA is that the source of strategic information comes in the form of a simulation model of the environment, rather than an analytic game form. In lieu of directly specified utility functions, the analyst must induce a utility model from payoff samples generated by simulation. The simulator generally produces a noisy sample of the payoff vector on each run, reflecting stochastic factors in the agent strategies or game environment. Sample information is thus accompanied by some error, and so we add a hat to notationally distinguish an *empirical game model*  $\hat{\Gamma}$  induced from simulation data from the game  $\Gamma$  itself, and similarly for the empirical utility functions  $\hat{u}_i$  defining the empirical game.

As illustrated in Figure 1, the empirical game evolves throughout the EGTA process, continually refined as new simulation data accrues over an expanding strategy space based on results from intermediate analysis. Iterative development of a sequence of models is common in game-theoretic reasoning, in standard analytic approaches as well as EGTA [Wellman and Mayo, 2024]. Here we consider the empirical game at a particular point in the process, induced from a fixed body of simulation data collected to that point.

The most straightforward way to construct an empirical game model from data is through *direct estimation*. In this approach, the empirical payoff for agent  $i$  in profile  $s$ ,  $\hat{u}_i(s)$ , is estimated as the sample average over a set of observations taken in simulation runs of profile  $s$ . More sophisticated sampling methods may weight observations non-uniformly, or employ other statistical techniques to sharpen estimates for a given body of data. Statistical reasoning for EGTA is discussed further in Section 6.

Let us illustrate game estimation with the running example game of sequential auctions with myopic marginal value bidding. Consider a small version with  $n = g = 3$ , and three strategies, defined by shading factors:  $\rho \in \{0.3, 0.5, 0.7\}$ . Valuations are drawn as in the homogeneous-good model employed by Wellman et al. [2017], and the auction in each round is first-price. We ran 10,000 simulations of all distinct profiles in this game, yielding the empirical game displayed in Table 1.

Inspection of Table 1 reveals that  $(0.3, 0.3, 0.3)$  is a PSNE in this simple empirical game. Indeed, it is uniquely so. The setting  $\rho = 0.3$  is not quite dominant, as the strategy  $\rho = 0.5$  is a best response when the other-player profile is  $(0.5, 0.7)$  or  $(0.7, 0.7)$ . However,  $\rho = 0.7$  is dominated in this game, and after eliminating  $\rho = 0.7$ ,  $\rho = 0.5$  becomes dominated as well. By iterated dominance, only  $\rho = 0.3$  survives. Another way to visualize this game and identify the unique PSNE is depicted in Figure 4.

$\rho_3 = 0.3$	0.3	0.5	0.7	$\rho_3 = 0.5$	0.3	0.5	0.7
0.3	51.5	44.7	39.8	0.3	44.7	36.9	32.0
0.5	44.2	40.9	37.6	0.5	40.9	36.8	33.3
0.7	28.9	27.0	25.6	0.7	27.0	25.6	23.9
	$\rho_3 = 0.7$			0.3	0.5	0.7	
	0.3			39.8	32.0	28.0	
	0.5			37.6	33.3	29.9	
	0.7			25.6	23.9	22.0	

Table 1: Normal forms for three-player sequential auctions with three heuristic strategies. Each  $3 \times 3$  table presents estimated payoffs (sample average rounded to tenths) for the row player, given the combination of row and column player, with a third player fixed at  $\rho_3$ .

Constructing an entire empirical utility function by direct estimation takes simulation time proportional to the number of strategy profiles. For  $n$  players and  $|S|$  available strategies per player, there are  $|S|^n$  possible profiles. For instance, in the 3-player 3-strategy game of Table 1 there are  $3^3 = 27$  table entries, one for each profile. With symmetry, the number of distinct profiles is somewhat smaller: indeed, the payoffs in Table 1 can be derived from simulations of only ten profiles. The savings, however, are limited, as a symmetric game comprises  $\binom{n+|S|-1}{n}$  distinct profiles, which is still exponential in the smaller of  $n$  and  $|S|$ . So we cannot expect simulation to be performed exhaustively for games with many players and strategies (or even large numbers of one and modest of the other). The alternatives are to reason about an incompletely specified game model (i.e., where only a subset of profiles are evaluated; see Section 5), or to extend the game model to unevaluated profiles through some kind of generalization (i.e., machine learning; see Section 3.2.3) process.

### 3.2.1 HEURISTIC PAYOFF TABLES

The natural representation of a finite  $n$ -player normal-form game is as an  $n$ -dimensional matrix, with cell  $(j_1, \dots, j_n)$  containing the *payoff vector*

$$(u_1(s_{j_1}, s_{-1}), \dots, u_n(s_{j_n}, s_{-n})), \text{ where } s_{-i} = (s_{j_1}, \dots, s_{j_{i-1}}, s_{j_{i+1}}, \dots, s_{j_n}).$$

An equivalent way to express these vectors is as a set of  $n$   $n$ -dimensional matrices, with cell  $(j_1, \dots, j_n)$  of matrix  $i$  containing the payoff scalar  $u_i(s_{j_i}, s_{-i})$ . The special case of  $n = 2$  is the bimatrix game representation presented in Section 2.2.3.

Such a matrix representation of game  $\Gamma$  has size  $n \prod_{i \in N} |S_i|$ . This size is required to fully represent  $\Gamma$ , but an empirical game model is typically incomplete in the sense of including payoffs for only a subset of strategy profiles (Section 3.2.2), and moreover may incorporate special structure such as symmetry that would afford more compact representation. We therefore commonly employ a sparse-matrix representation that requires space proportional to the distinct strategy profiles for which payoffs are evaluated. This representation has been termed a *heuristic payoff table* (HPT) [Walsh et al., 2002], and special-purpose infrastructure for data management of large HPTs has been developed [Cassell and Wellman, 2013].

Consider a symmetric normal-form game with  $n$  players and  $|S|$  strategies. For a symmetric game, what matters is not *which* players choose a given strategy, but just *how many*. Therefore we can represent a strategy profile by a vector of *counts*: for each strategy, the number of players who choose to play it. As noted in Section 3.2, there are  $\binom{n+|S|-1}{n}$  distinct count vectors for a symmetric game with  $n$  players and  $|S|$  strategies, compared with the  $|S|^n$  profiles ignoring symmetry.

Formally, let the HPT  $\mathcal{H} = (\mathcal{N}, \mathcal{U})$ , where  $\mathcal{N}$  is a  $\binom{n+|S|-1}{n} \times |S|$  matrix of counts, and  $\mathcal{U}$  is a matrix of utilities of the same dimension. Each row represents a profile, such that entry  $\mathcal{N}_{k,j} \in \{0, \dots, n\}$  indicates the number of players choosing strategy  $s_j$  in the  $k^{\text{th}}$  profile. The rows of  $\mathcal{N}$  are all distinct, and satisfy  $\sum_j \mathcal{N}_{k,j} = n$  for all  $k$ . Entry  $\mathcal{U}_{k,j}$  is undefined if  $\mathcal{N}_{k,j} = 0$ , and otherwise represents the payoff to playing strategy  $s_j$  in the  $k^{\text{th}}$  profile. To connect  $\mathcal{U}$  to the standard utility function, let  $s^k = (s_{j_1}, \dots, s_{j_n})$  be a profile in the standard strategy vector representation consistent with the  $k^{\text{th}}$  HPT profile, that is,  $\forall j. |\{i \mid j_i = j\}| = \mathcal{N}_{k,j}$ . Then  $\forall i. \mathcal{U}_{k,s_{j_i}} = u_i(s^k)$ .

Table 2a presents the HPT representation for the  $n = |S| = 3$  sequential auction game specified in standard matrix form in Table 1. A partial HPT for a five-player version is presented in Table 2b.

$\mathcal{N}$	$\mathcal{U}$	$\mathcal{N}$	$\mathcal{U}$	
3 0 0	51.5 – –	5 0 0	37.6	– –
2 1 0	44.7 44.2 –	4 1 0	32.0	39.2 –
2 0 1	39.8 – 28.9	4 0 1	28.5	– 26.7
1 2 0	36.9 40.9 –	3 2 0	26.1	36.3 –
1 1 1	32.0 37.6 27.0	3 1 1	22.5	33.8 25.4
1 0 2	28.0 – 25.6	3 0 2	19.2	– 24.0
0 3 0	– 36.8 –	...	...	...
0 2 1	– 33.3 25.6	2 1 2	15.2	27.5 22.4
0 1 2	– 29.9 23.9	...	...	...
0 0 3	– – 22.0	0 0 5	–	– 16.1

a 3 players (cf. Table 1)

b 5 players (partial: 8 of 21 profiles shown)

Table 2: Example heuristic payoff tables for 3- and 5-player versions of the sequential auctions game, with three strategies. Each row represents a profile, with counts and payoffs for strategies  $\rho = 0.3, 0.5, 0.7$ , respectively.

The HPT construct can be straightforwardly extended to role-symmetric games. Each role has a fixed strategy set and number of players, with symmetry holding within the role. We define the HPT  $\mathcal{H} = (\mathcal{N}^1 \times \dots \times \mathcal{N}^K, \mathcal{U})$ , where  $\mathcal{N}^r$  is a counts matrix for role  $r$ , and  $\mathcal{U}_{(k_1, \dots, k_K), j, r}$  gives the payoff for playing strategy  $s_j$  in role  $r$  given the rest of the strategy profile.

### 3.2.2 INCOMPLETE EVALUATION OF PROFILE SPACE

When evaluating all profiles by simulation is computationally infeasible, a natural approach is to infer what we can from whatever part of the restricted game we can feasibly evaluate. In many cases we can certify solutions or approximate solutions far short of exhaustively evaluating the profile space. Verifying a Nash equilibrium or its degree of approximation is a matter of calculating a profile's regret. This requires only the payoffs for that profile and all *deviation profiles*: profiles formed by unilateral deviations. For a pure profile, there are  $\sum_i |S_i| - n$  deviation profiles. For

a mixed profile  $\sigma$ , the number of deviation profiles is generally exponential in  $\max_i |\text{supp}(\sigma_i)|$ , which may still be modest for small-support profiles.

Finding an approximate solution is generally more expensive than verifying one, but can also often be accomplished short of exhaustive evaluation of the profile space. We survey EGTA techniques for reasoning about incomplete game models in Section 5.

### 3.2.3 GAME MODEL LEARNING

The machine learning approach to empirical game modeling is essentially a form of regression where the input is a set of (profile, payoff-vector) pairs, and the output is the vector of empirical utility functions  $\hat{u}_i$ . These techniques can be used to infer a complete empirical game model from an incomplete one.

For illustration, suppose we wish to extend the example three-player sequential auction above (Table 1) to include a fourth strategy,  $\rho = 0.4$ . There are many possible ways to extend the payoff matrix to include profiles with the new strategy. In this case, since the strategies are defined parametrically, we can apply a nearest-neighbor approach with linear interpolation. For example, consider a profile of the form  $(0.4, \rho_2, \rho_3)$ , with  $\rho_2, \rho_3 \in \{0.3, 0.5, 0.7\}$ , the set of strategies for which we already have estimates. To estimate  $u(0.4, \rho_2, \rho_3)$ , we could simply average  $\hat{u}(0.3, \rho_2, \rho_3)$  and  $\hat{u}(0.5, \rho_2, \rho_3)$ , for example  $\hat{u}_1(0.4, 0.3, 0.3) = (51.5 + 44.2)/2 = 47.85$ . The actual value, based on 10,000 samples, is 50.48. Similarly, the interpolated estimate  $\hat{u}_2(0.4, 0.3, 0.3) = (51.5 + 44.7)/2 = 48.1$  (actual sampled estimate is 47.2). More sophisticated approaches could fit the payoff function to the data using richer hypothesis spaces, for example based on neural networks.

The first work to apply regression to learn payoff functions from simulation data was that of Vorobeychik et al. [2007]. Ficici et al. [2008] employed clustering to partition a large number of players into two roles, then regression to find a best-fit representative payoff function for each role. Jordan and Wellman [2009] applied cross-validation methods to select the best empirical game model, for instance to decide whether some strategies are similar enough to merge samples. Wiedenbeck et al. [2018] introduced methods to learn large symmetric games, using a representation that encodes the number of players choosing each strategy, rather than vectors based on a player ordering. Sokota et al. [2019] propose an approach that learns deviation payoffs (defined in Section 4.4) with respect to role-symmetric mixed strategies, an alternative to directly learning payoff functions, which provides some advantages for equilibrium computation. Li and Wellman [2021] employ this representation for game learning in an EGTA algorithm for symmetric Bayesian games, evaluated in a simultaneous-auction setting. Shao et al. [2025] propose a matrix completion approach, adopting a conservative bias to steer toward equilibria that are relatively well supported by the available data.

Some recent research in game representation has developed structural model forms that support succinct representation of games exhibiting particular regularities. An example is graphical games [Kearns, 2007], which compactly capture situations where agents are affected by others' actions only in a local neighborhood. There has also been some work on learning such graphical game models from payoff data [Duong et al., 2009, Fearnley et al., 2015]. Li and Wellman [2020] propose an approach that interleaves structure learning and payoff regression to induce tractable game models with many players.

Gatchel and Wiedenbeck [2023] broaden the problem to learn a model covering *families* of games, defined by specified context features. For example, a family of sequential auction games could encompass a range of scenarios parametrized by the number of goods, number of players, or features of the valuation distributions. This approach supports reasoning about the relationship between environmental features and game solutions, with greater robustness and sample efficiency compared to learning separate models for an enumerated set of game instances.

There is extensive additional literature at the intersection of machine learning and game theory, which is relevant but not specifically oriented toward learning game models from simulation data. This includes a great deal of work on algorithms that learn to play games (i.e., converge to equilibria) through repeated interaction [Fudenberg and Levine, 1998]. Some more recent work has exploited advances in deep neural networks to learn equilibrium behavior [Bichler et al., 2021, Gatti and Restelli, 2011, Marchesi et al., 2020] or optimal mechanisms [Dütting et al., 2024] directly from strategy simulations (i.e., without constructing a game model). In the realm of game modeling, there is interesting research on learning to predict behavior of human players given a normal-form game specification [Wright and Leyton-Brown, 2017]. Yet another category is work on learning games from behavioral data [Honorio and Ortiz, 2015, Waugh et al., 2011], generally based on fitting structure and parameters of a game model under assumption that the behavior is generated rationally. Gao and Pfeffer [2010] suggest an approach that combines payoff data with behavioral data in this way. In all of these categories we could cite many other works; satisfactory coverage would require a full-length survey treatment.

### 3.2.4 PLAYER REDUCTION

Another technique developed to support scalability of empirical game modeling is *player reduction*. The goal of this technique is to approximate a many-player game by one with considerably fewer ( $\ll$ ) players. The intuition is that in a large game, it may be approximately correct to reason coarsely about agent aggregates, expecting results not to be unduly sensitive to the exact number of agents adopting a particular strategy. The approach inherently requires symmetry, otherwise aggregation would not be clearly meaningful.<sup>3</sup>

Let  $p < n$ ; typically  $p$  is a small fraction of  $n$ . Formally, a *reduced game*  $\Gamma^{(p)} = \langle P, (S_i), (u_i^{(p)}) \rangle$ ,  $|P| = p$ , is a  $p$ -player game derived from an  $n$ -player *full game*  $\Gamma = \langle N, (S_i), (u_i) \rangle$ . The reduced and full games have the same strategy sets. The key to defining the reduction is specifying how the utility function  $u_i^{(p)}$  for the reduced game can be derived from that of the full game. Reductions of this sort mesh well with the simulation-based approach since payoff data for estimating the reduced game can be taken from simulations of the full game.

The idea of *hierarchical reduction* [Wellman et al., 2005b] is to model the  $p$ -player game as if each player controlled  $n/p$  of the players in the full game. Suppose for simplicity  $p$  divides  $n$ , so that  $n/p$  is integral.<sup>4</sup> Let  $s_{-i}^{(p)}$  be an other-agent strategy profile in the reduced game, a vector of  $p - 1$  strategies. Then we can define  $u_i^{(p)}(s, s_{-i}^{(p)}) = u_i(s, s_{-i}^{(n)})$ , where the full-game other-agent profile

3. We describe the techniques below for the case of full symmetry. Generalization to role symmetry is straightforward, limiting all aggregation to be applied within roles.

4. The case where  $n/p$  is fractional can be handled by careful rounding or interpolation.

$s_{-i}^{(n)}$  contains  $n/p$  copies of  $s_{-i}^{(p)}$ , plus  $n/p - 1$  copies of  $s$ . (Note that given symmetry, the ordering of strategies in  $s_{-i}^{(n)}$  does not matter.)

In **twins reduction** [Ficici et al., 2008], an  $n$ -player game is reduced to a 2-player game where each player views the other as representing an aggregate of all the rest playing the other strategy. The derived utility function can be written simply as  $u_i^{(2)}(s, s_{-i}) = u_i(s, s_{-i}^{(n)})$ , where the full-game other-agent profile  $s_{-i}^{(n)}$  contains  $n - 1$  copies of  $s_{-i}$ . An interesting property of the twins reduction is that it preserves symmetric PSNE, that is,  $(s, s)$  is an NE of  $\Gamma^{(2)}$  twins-reduced from  $\Gamma$  iff everyone playing  $s$  is an NE of  $\Gamma$ . This is because the twins reduction (unlike hierarchical reduction) preserves the effect of single-player deviations.

**Deviation-preserving reduction** (DPR) [Wiedenbeck and Wellman, 2012] generalizes twins reduction for  $p > 2$ . The idea of DPR is to consider each reduced player as controlling one player in the full game, but to treat the other reduced players as full-game aggregates. For simplicity suppose  $p - 1$  divides  $n - 1$ . The DPR mapping is given by  $u_i^{(p)}(s, s_{-i}^{(p)}) = u_i(s, s_{-i}^{(n)})$ , where the full-game other-agent profile  $s_{-i}^{(n)}$  contains  $(n - 1)/(p - 1)$  copies of  $s_{-i}^{(p)}$ . For  $p = 2$ , DPR is the same as twins, and for any  $p$  it also preserves symmetric PSNE.

To apply player reduction in EGTA, one typically analyzes the reduced game as an approximation for the full game. Symmetric (mixed or pure) profiles can be interpreted as profiles over any number of players, so we can consider symmetric solutions of the reduced game as rough or candidate solutions of the full game. If the support of the reduced-game solution is sufficiently small, evaluation of accuracy with respect to the full game by sampling can be quite tractable. Although in the worst case an approximation by player reduction can be arbitrarily bad, DPR in particular has proved reasonably accurate for a range of natural games. For example, Brinkman [2018] used DPR ( $p = 4$  or  $6$ ) to analyze financial market games with up to 216 agents, and verified statistically that regret due to the reduced-game approximation was small relative to the inherent variability of the models.

### 3.3 Game Solving

An empirical game model is fundamentally a model of a game, and so all the concepts and tools of computational game theory apply in principle to games induced from agent-based simulation or other data sources. This includes game forms (e.g., normal form, extensive form), solution concepts, and algorithms for computing equilibria or other objects of game-theoretic analysis. In practice, certain representations and methods have proven particularly relevant for EGTA studies. An example is the HPT representation for symmetric games presented in Section 3.2.1. This encoding is particularly convenient for applying *replicator dynamics* (Section 4.4), hence RD is commonly employed in EGTA. EGTA-specific approaches have also been developed for special forms of learned game models, as well as representations that accommodate incompleteness in game models (Section 5.2).

### 3.4 Strategy Evaluation

By the very definition of a game environment, it is generally not possible to evaluate the quality of one player's strategy absent consideration of the strategies chosen by others. Thus, when we

refer to solutions or solution concepts for games, we characterize these on the joint strategy space. Nevertheless, we often are particularly interested in the perspective of a particular player, and would like some way to assess the efficacy of that player’s strategy in absolute or relative terms (i.e., scores or rankings), without explicit reference to other-player choices. There is no escaping the need for making some assumptions about these choices, but it could be that some fairly generic assumptions about strategic context are sufficiently informative for strategy evaluation.

For example, one can invoke a worst-case assumption on other-agent play, and compare strategies on that basis. The optimal strategy under this assumption is the *maximin strategy*, and the minimum payoff guaranteed by the maximin strategy is the player’s *safety value* for the game.

Except in zero-sum games, worst-case assumptions do not capture the interests of other players, and are thus not generally a realistic expectation for their choices. An alternative is to assume rational play, which brings us back to the realm of game-theoretic solution concepts. For example, we might consider the performance of a player’s strategy when the others play according to a Nash equilibrium. Jordan et al. [2007] defined the *NE-regret* of a strategy  $s_i$  as  $\epsilon_i(s_i, \sigma_{-i}^*)$ , for  $\sigma^*$  a NE profile. (In a game with multiple equilibria, there would be NE-regret measures with respect to each.) Jordan [2010] showed how to use NE-regret for ranking strategies as part of an EGTA process.

Balduzzi et al. [2018] studied the question of empirical strategy evaluation in generalized form, motivating the problem and pointing out strengths and limitations of various approaches. Their proposal to rank strategies by *Nash-averaging* is technically equivalent to ranking by NE-regret, with an additional prescription to select the benchmark equilibrium that maximizes entropy.

Alternative approaches for ranking strategies appeal to evolutionary solution concepts. In particular,  $\alpha$ -Rank has been deployed to evaluate the strength of various learning strategies in the games of Go, soccer, and poker [Omidshafiei et al., 2019]. Amenability to a variety of measures that account for strategic context make EGTA a valuable tool for evaluating novel multiagent training algorithms [Li and Wellman, 2024].

### 3.5 Visualization

Game analysts are often motivated to draw strategic insights about a game environment, beyond identifying specific solutions. Toward that end, researchers have devised various ways to visualize game analyses. An example is the use of directional field plots (e.g., Figures 5 and 6) to illustrate evolutionary dynamics over a profile space.

Another approach developed in EGTA research graphs the relationships over an enumerated set of pure profiles. A *deviation graph* connects profiles by edges indicating the most beneficial deviations.<sup>5</sup> A plot of such a graph organized to display levels of profile regret is called a *contour deviation graph* [Jordan, 2010, Chapter 3]. Figure 4 presents a contour deviation graph corresponding to the  $n = g = 3$  auction empirical game model given by Table 1. The ten nodes of the graph represent the pure profiles of the game. In this visualization, an outgoing edge represents the most beneficial one-player deviation. (One could also include edges for all deviations, weighted by the

---

5. These are related to response graphs, introduced in Section 4.3 for analyzing learning dynamics. Response graphs encode transitions for all alternative strategies, whereas here we focus on best deviations.

associated gain.) For example, from the profile  $(0.3, 0.7, 0.7)$  near the top right, a player could gain 14.2 by switching from strategy  $\rho = 0.7$  to  $\rho = 0.3$ . The sole node without an outgoing edge is  $(0.3, 0.3, 0.3)$ , the game’s unique PSNE.

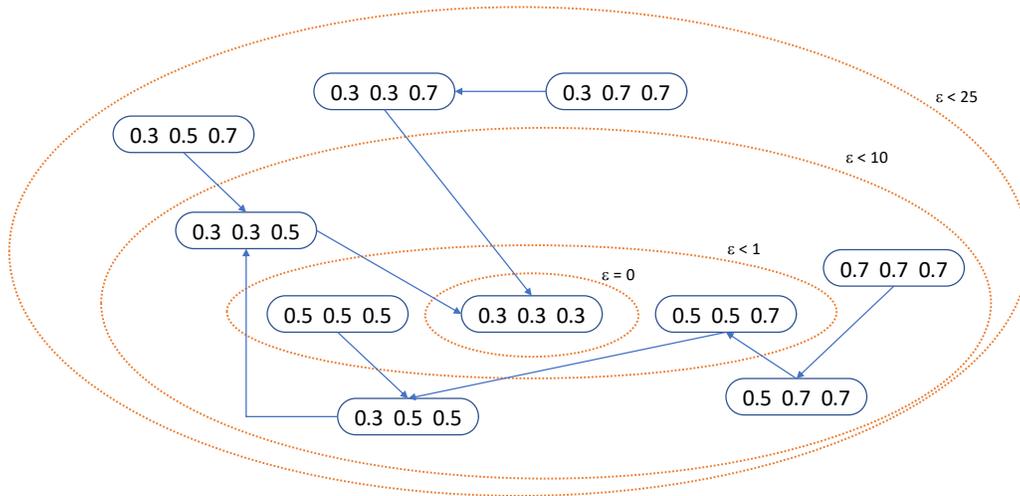


Figure 4: Contour deviation graph for the sequential auction game model of Table 1. Nodes are strategy profiles, and edges denote most beneficial one-player deviations. The dashed ellipses in orange delimit increasing levels of regret ( $\epsilon$ ) as one moves outward in the plot.

Scaling these visualizations to high-dimensional or otherwise large profile spaces is a challenge. New and creative ideas are required to support the extraction of strategic insights in game analysis. One interesting concept proposed by Czarnecki et al. [2020] is to view strategic relationships on two dimensions: *transitive* and *intransitive*. The upright axis represents transitive relationships, where strategies can be ordered relatively unambiguously in strength or competence. The radial axis represents intransitive relationships, like the canonical rock-paper-scissors interaction. Through an extensive EGTA exercise over popular two-player zero-sum games, these authors find a pervasive “spinning top” structure, in which the degree of non-transitivity tapers off as strategies improve along the upright dimension toward the game’s NE solutions. Omidshafiei et al. [2020] present a more comprehensive set of tools, based on response graphs and other constructs, for categorizing and visualizing the strategic landscapes of a wide variety of games.

Another type of strategic question of interest in EGTA is how solutions vary as a function of the game environment. Such questions are typically addressed by conducting EGTA over a space of parametric instances, and plotting features of the solution. For example, Wang et al. [2021] used EGTA to derive empirical equilibria among trading strategies for 36 instances of a financial market game: with and without the presence of a market manipulator, for 18 configurations of trader population size and stochastic parameters of the market. Plotting features of the strategies adopted in equilibrium, as well as outcome features (in this case, spoofing effectiveness, price accuracy, and welfare) yields insights about how spoofing manipulation operates across a space of market situations. These kinds of visualizations are now quite common in EGTA studies.

## 4. Evolutionary Game Analysis

As noted in Section 2.3, evolutionary game theory is a natural fit for games defined by agent-based simulation. In this section, we elaborate on some of the basic concepts underpinning the evolutionary perspective on game theory, and discuss EGTA methods developed based on these concepts.

### 4.1 Evolutionarily Stable Strategies

Imagine a large population of agents, each playing a pure strategy, from which two are repeatedly selected at random to play a game. Imagine further that these agents reproduce according to their success in these interactions, so that successful strategies multiply, while unsuccessful ones die off. In evolutionary game theory, we often view such populations as conceptually infinite, and interpret the distribution of strategies adopted as a symmetric mixed-strategy profile. If one strategy attempts to invade another (i.e., if a small part of the population mutates), but if the reproductive success of the new one lags behind that of the original, then the new strategy will eventually disappear. In other words, the original strategy is *evolutionarily stable* against this invading strategy. More generally, an *evolutionarily stable strategy* (ESS) is a strategy that is robust against evolutionary pressure from any mutant strategy not yet present in the population, or present only as a very small fraction.

Formally, suppose the population can be described by the state vector (or mixed strategy)  $\sigma = (\sigma(s_1), \dots, \sigma(s_{|S|}))$ , with  $\forall j. 0 \leq \sigma(s_j) \leq 1$  and  $\sum_j \sigma(s_j) = 1$ , representing the fractions of the population playing pure strategies  $1, \dots, |S|$ , respectively. A strategy  $\sigma$  is an ESS if it is immune to invasion by other strategies that initially occupy only a small fraction of the population. Let  $f(\sigma, \pi)$  be the (expected) fitness of strategy  $\sigma$  against strategy  $\pi$ . Formally, then, strategy  $\sigma$  is an ESS iff, for any mutant strategy  $\pi$ , the following hold:

1.  $f(\sigma, \sigma) \geq f(\pi, \sigma)$ , and
2. if  $f(\sigma, \sigma) = f(\pi, \sigma)$ , then  $f(\sigma, \pi) > f(\pi, \pi)$ .

The first condition states that an ESS is also a Nash Equilibrium of the original game, which implies that ESS is a refinement of the Nash solution concept. The second condition states that if the invading strategy does as well against the original strategy as the original strategy does against itself, then the original strategy must do better against the invader than the invader does against itself. It turns out that every ESS is an asymptotically stable fixed point of the replicator dynamics process [Weibull, 1997], which we define next.

### 4.2 Replicator Dynamics

*Replicator dynamics* (RD), introduced by Taylor and Jonker [1978] and developed further by Schuster and Sigmund [1983], is another key concept from EGT. As a dynamical system, RD describes mathematically how a population can evolve over time, either computationally or based on biological operators such as selection, mutation, and crossover.

In their most basic form, the RD equations express the canonical biological selection mechanism: survival of the fittest. Suppose the fitness of pure strategy  $i$  is given by a fitness function  $f_i(\sigma)$ , typically defined as  $i$ 's expected payoff, with the average population fitness  $\bar{f}(\sigma) = \sum_j \sigma_j f_j(\sigma)$ . In the single population case, which is applicable to symmetric games, the RD equations are:

$$\dot{\sigma}_i = \sigma_i (f_i(\sigma) - \bar{f}(\sigma)), \tag{2}$$

If  $A$  denotes a symmetric bimatrix payoff matrix, as in Section 2.2.3, the RD equations simplify as:

$$\dot{\sigma}_i = \sigma_i ((A\sigma)_i - \sigma^T A\sigma). \tag{3}$$

We illustrate these dynamics on a classic example, the Prisoners' Dilemma, whose payoff table is shown in Figure 5. The axes of the field plot correspond to the probability of the respective players playing action  $D$  (Defect). The gradient flow indicates that all probability mass flows to coordinates  $(1, 1)$ , which represents the pure Nash equilibrium  $(D, D)$ . One can also observe from this plot that the Nash equilibrium  $(D, D)$  is evolutionarily stable: injecting any number of cooperators into the population would not lead the dynamics to stray from the  $(D, D)$  state.

$$\begin{array}{c} C \\ D \end{array} \begin{array}{cc} C & D \\ \left( \begin{array}{cc} 3, 3 & 0, 5 \\ 5, 0 & 1, 1 \end{array} \right) \end{array}$$

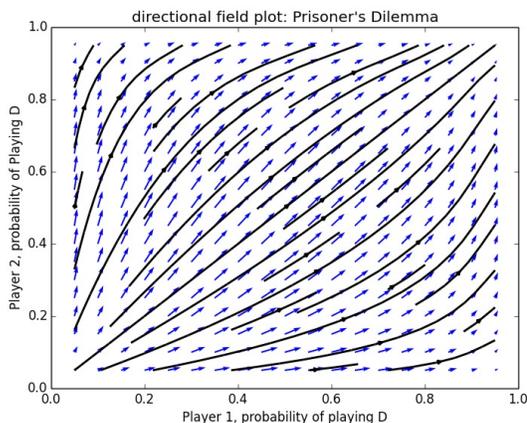


Figure 5: Payoff matrix (left) and directional field plot (right) for the Prisoners' Dilemma game. Strategies  $D$  and  $C$  represent *Defect* and *Cooperate*. Flows point towards full probability on the unique Nash equilibrium,  $(D, D)$ , which is a strong attractor.

We can also extend the replicator equations to asymmetric games, but we then need multiple populations. In a two-player asymmetric bimatrix game with payoff matrices  $(A, B)$ , and population states  $x$  and  $y$ , respectively, for populations 1 and 2 (i.e., corresponding to players 1 and 2), evolution under RD is now given by

$$\dot{x}_i = x_i ((Ay)_i - x^T Ay) \quad \dot{y}_j = y_j ((x^T B)_j - x^T B y) \quad \forall (i, j) \in S_1 \times S_2. \tag{4}$$

This system of coupled differential equations models the temporal dynamics of the interactions among agents in these two populations.

We illustrate the asymmetric replicator dynamics in the BoS game in Figure 3, which depicts the gradient dynamics of player 1 (row,  $x$ -player) and player 2 (column,  $y$ -player), respectively. This

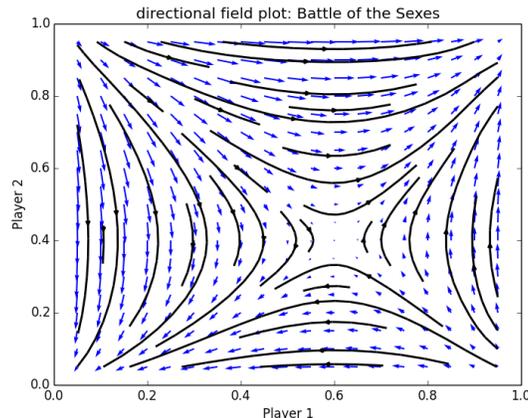


Figure 6: Directional field plot of the BoS game. Flows point toward the two PSNE,  $(B, B)$  and  $(S, S)$ , whereas the mixed Nash equilibrium at  $(\frac{2}{3}, \frac{1}{3})$  is unstable.

game has three equilibrium fixed points: two pure-strategy NE (PSNE)—at the origin and  $(1, 1)$ , and a third (evolutionarily unstable) mixed NE at coordinates  $(\frac{2}{3}, \frac{1}{3})$ . This example thus illustrates how the ESS solution concept is a refinement of Nash.

### 4.3 Evolutionarily Dynamic Solution Concepts

Whereas Nash equilibria are guaranteed to exist in all finite games [Nash, 1950], their computation is believed to be intractable in general-sum games [Daskalakis et al., 2009, Goldberg et al., 2013], and their non-uniqueness leads to an equilibrium selection problem [Harsanyi and Selten, 1988]. Such indeterminacy motivates the consideration of behavior under learning dynamics (e.g., fictitious play), viewed as stochastic processes.

One simple way to apply this approach, dating back to Young [1993], is to represent the learning dynamics of interest by a Markov chain. One such possible Markov chain is a *response graph*, with states as strategy profiles, and transitions/edges indicating the extent to which a player has an incentive to deviate unilaterally from one state/profile to another. One can then consider a stationary distribution of this Markov chain as a solution concept, but the stationary distribution is not necessarily unique. Instead, one often introduces small perturbations to this chain, which render it irreducible and its stationary distribution unique [Levin et al., 2006]. Note that such a solution concept captures agent interactions regardless of initial conditions (i.e., the agents' initial strategies).

The *stochastically stable distribution* (SSD) of a Markov chain is the limit as  $\epsilon \rightarrow 0$  of such a perturbed Markov chain,<sup>6</sup> and the *stochastically stable states* (SSS) are those in the support of the SSD. The SSS yield an alternative solution concept, possibly ranked according to the corresponding SSD probabilities. Unlike Nash equilibrium, the SSD (and hence the set of SSS) is unique, and computing it is tractable [Wicks and Greenwald, 2005]. Young [1993] thus proposed it as a solution to equilibrium selection in his work on the evolution of conventions.

6. This limit is guaranteed to exist when the perturbed Markov chain is *regular* [Young, 1993].

More recently, a related solution concept called  $\alpha$ -Rank was proposed by Omidshafiei et al. [2019]. Not only was it applied as described above to rank strategy profiles, it was also applied to rank agents, by building a Markov chain whose states are agents instead. Encoded in the Markov chain’s probabilities are evolutionary dynamics that model a selection-mutation process over a set of interacting populations. Individuals sampled from each population play an  $n$ -player game, and then the strongest individuals either reproduce, or with a very small probability mutate.

Specifically, the propagation of strong agents is driven by a selection function that compares the fitness of a resident agent  $\tau$  with a competing agent  $\sigma$ , as shown on the bottom right of Figure 7. The *ranking-intensity* of selection, which is given by a parameter  $\alpha$ , influences the probability that a mutant overtakes a population. A low  $\alpha$  corresponds to weak selection, while a large  $\alpha$  ensures that only the strongest mutants survive. This selection-mutation model is encoded as the transition matrix of the  $\alpha$ -Rank Markov chain. (See the formula at the top right of Figure 7.)

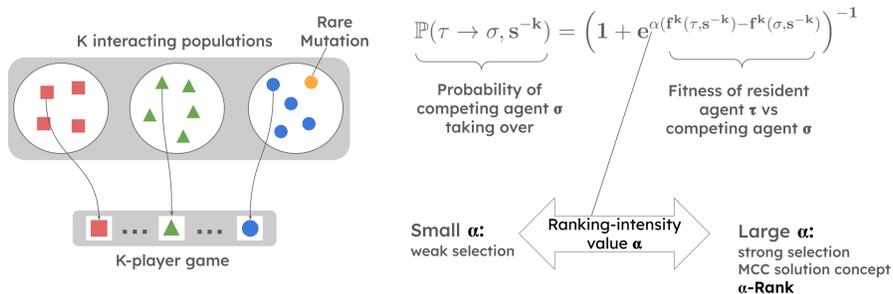


Figure 7: A discrete-time selection-mutation evolutionary process described as a Markov chain.

The  $\alpha$ -Rank algorithm proceeds as follows:

1. Estimate the empirical game.
2. Define a Markov chain where states are the agents being evaluated.
3. Compute the transition matrix according to a selection-mutation model parameterized by  $\alpha$ .
4. Compute the stationary distribution of this Markov chain.
5. Rank the agents via ordered masses of this distribution.

The final steps are repeated by sweeping over  $\alpha$  values, until the resulting rankings stabilize over a few consecutive iterations.

Figure 8 presents an example. The empirical game here is a two-player symmetric game with 56 AlphaZero Chess checkpoints, ranging from beginning to end of training. For example, AZ(99.4) has completed 99.4% of training. One note to make here is that these interactions are evaluated using the entire 56-agent dataset, though we show only the top-8 ranked agents for clarity.

The payoffs in the empirical game are the win ratios of pairwise match-ups between these agents. On the left we depict the ensuing discrete-time Markov chain, with node color indicating the mass of the stationary distribution on that agent. The majority of top-ranked agents indeed correspond to snapshots taken near the end of training. There are some more interesting outcomes, however. For example, the agent ranked 5th is AZ(86.4), which places higher than several agents with longer

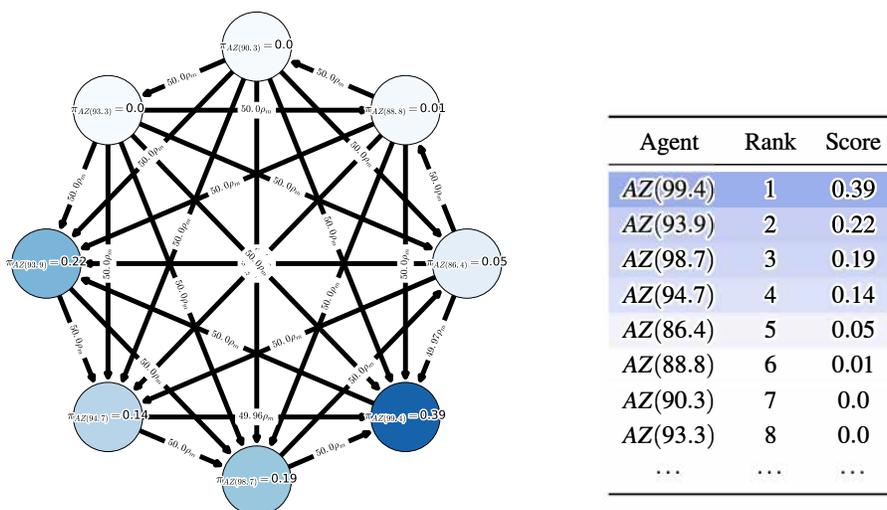


Figure 8: (LHS) Graph representation of the discrete-time dynamics applied to the AlphaZero dataset, also referred to as the response graph. Nodes represent the various agents or checkpoints during training of the Chess agent. (RHS) The outcome of  $\alpha$ -Rank applied to AlphaZero Chess, considering 56 agents and ranking the top 8 (listed).

training time. The key point is that by using this type of evaluation, we can narrow the focus to agents of interest during training, and conduct a more refined analysis of their interactions.

#### 4.4 Replicator Dynamics on HPTs

The HPT representation can be leveraged for use with any game-solving method, but is particularly well-suited for implementing algorithms that exploit symmetry, such as the method of replicator dynamics described in Section 4.2. That section presented RD as a dynamical system in terms of differential equations. Here we describe it as an equilibrium search algorithm, which traverses a simplex as does the RD dynamical system, though in discrete steps. Points in the simplex represent mixed strategies, which can also be interpreted as symmetric mixed profiles.<sup>7</sup> As a search algorithm, RD is an iterative improvement method that evaluates all strategies in the current solution, and updates their probabilities based on relative fitness. If all strategies are equally fit, RD is at a fixed point. RD may cycle or otherwise fail to converge, but if it does converge to a stable fixed point (e.g., an ESS), the result also constitutes a Nash equilibrium [Gintis, 2009].

Fitness in this context is expected payoff, under the assumption that all other players choose according to the current mixed strategy. Let  $\sigma$  be a mixed strategy, with  $\sigma(s_j)$  the probability that strategy  $s_j$  is played in  $\sigma$ . Given  $\sigma$  and HPT  $\mathcal{H} = (\mathcal{N}, \mathcal{U})$ , the **deviation payoff**  $U_{\sigma,j}$  for strategy  $s_j$  represents the expected payoff to that strategy conditional on other players following the mixture  $\sigma$ . To express this expectation, define  $\mathcal{N}_{k,\ell}^{-j}$  as a vector of other-agent counts given that the designated

7. As discussed in Section 4, it is common in evolutionary game theory to interpret the points as fractions of an infinite population of agents employing various pure strategies. We present the use of RD for studying evolutionary dynamics under this interpretation in Section 4.5.

player plays  $s_j$  in the  $k^{\text{th}}$  profile. That is,  $\mathcal{N}_{k,j}^{-j} = \mathcal{N}_{k,j} - 1$ , and  $\mathcal{N}_{k,\ell}^{-j} = \mathcal{N}_{k,\ell}$  for  $\ell \neq j$ . Then

$$\mathcal{U}_{\sigma,j} = \sum_{k|\mathcal{N}_{k,j}>0} \binom{n-1}{\mathcal{N}_{k,1}^{-j}, \dots, \mathcal{N}_{k,n}^{-j}} \prod_{\ell=1}^{|S|} \sigma(s_\ell)^{\mathcal{N}_{k,\ell}^{-j}} \mathcal{U}_{k,j}, \quad (5)$$

where  $\binom{n-1}{\mathcal{N}_{k,1}^{-j}, \dots, \mathcal{N}_{k,n}^{-j}}$  is a multinomial coefficient, which describes the number of possible partitions of  $n-1$  objects into groups of sizes  $\mathcal{N}_{k,1}^{-j}, \dots, \mathcal{N}_{k,n}^{-j}$ .

Wiedenbeck and Brinkman [2023] present a series of data structure improvements, building on the HPT representation, that facilitate representation and computations over deviation payoffs. The cumulative effect of these is a  $10^4$ -fold speedup in RD for games with many players, compared to the baseline matrix game representation.

Given a specification of deviation payoffs, the RD algorithm starts from an initial mixed strategy  $\sigma_0$  and updates the mixture at each step  $t$  using a discrete-time version of the replicator equation (2):

$$\sigma_{t+1}(s_j) \leftarrow \sigma_t(s_j) \left[ 1 + \alpha \left( \mathcal{U}_{\sigma_t,j} - \sum_{\ell} \sigma_t(s_\ell) \mathcal{U}_{\sigma_t,\ell} \right) \right], \quad (6)$$

where  $\alpha > 0$  is a learning-rate parameter.

A typical implementation of RD for game-solving will start from a uniform or random initial mixed strategy, and iterate (6) until convergence (e.g., change in probabilities fall below a numeric threshold), or until a maximum number of steps have been reached. At termination, the result can be evaluated for distance from Nash equilibrium, as measured by regret. Multiple runs from different starting points is also common practice, to deal with non-convergence and to identify multiple equilibria if they exist. Though experience with RD has confirmed it is usually effective on empirical games encountered in practice, it is not guaranteed to find solutions and therefore having backup solution algorithms is recommended. For example, Wiedenbeck and Brinkman [2023] recommend running both RD and gradient descent (also supported by HPT-based representation of deviation payoffs) from a diverse set of starting points in the strategy simplex.

## 4.5 Approximating Infinite Populations for Evolutionary Dynamics

The previous section described how the HPT representation of empirical games can facilitate RD computations, assuming the dynamic state variable  $\sigma$  is interpreted as a symmetric mixed strategy. Evolutionary game theory more commonly appeals to an alternate interpretation, where  $\sigma$  is viewed as proportions of an infinite population. RD using the HPT representation can apply under this interpretation as well [Bloembergen et al., 2015], where the population is approximated by a large, but finite, population of  $p$  *replicators*.

An evolutionary game played by a population of  $p \geq n$  replicators is derived from, but not the same as, an underlying game among a fixed set of  $n$  players. In the evolutionary game among replicators,  $n$  instances are drawn from the population, inheriting their strategies, with or without replacement. The selected replicators then play their strategies and receive payoffs per the underlying  $n$ -player game. The payoff to each replicator in this game is their deviation payoff conditional on their

selection in this sampling process, given the population distribution. The larger the population  $p$ , the better the game among replicators approximates an infinite-population evolutionary game.

Using the HPT representation, we can express an evolutionary game with  $p$  replicators as if it were a  $p$ -player game. The idea is simply to define the payoffs by the result of the sampling process. For example, consider the  $2 \times 2$  Prisoner’s Dilemma game (Figure 5). Payoffs for the game with two replicators and no replacement (Table 3a) are exactly the same as for the underlying normal-form game, since  $p = n$ . With replacement (Table 3b), we see a difference in the profile where the population comprises one  $C$  and one  $D$  ( $\mathcal{U}_C$  in red). Conditional on a  $C$  replicator being chosen for one player, the probability of the other being  $D$  is one under no-replacement (payoff 0), but 0.5 if the sampling is with replacement (payoff  $1/2(3) + 1/2(0)$ ).

Table 3: Prisoner’s Dilemma HPTs with two replicators.

<p><b>a</b> No replacement</p> $\left( \begin{array}{cc cc} \mathcal{N}_C & \mathcal{N}_D & \mathcal{U}_C & \mathcal{U}_D \\ \hline 2 & 0 & 3 & - \\ 1 & 1 & 0 & 5 \\ 0 & 2 & - & 1 \end{array} \right)$	<p><b>b</b> With replacement.</p> $\left( \begin{array}{cc cc} \mathcal{N}_C & \mathcal{N}_D & \mathcal{U}_C & \mathcal{U}_D \\ \hline 2 & 0 & 3 & - \\ 1 & 1 & 1.5 & 3 \\ 0 & 2 & - & 1 \end{array} \right)$
--	---

For the infinite population case, one may consider the probability that each profile  $k$  is realized under the given  $\sigma$ :

$$\Pr(k | \sigma) = \binom{n}{\mathcal{N}_{k,1}, \dots, \mathcal{N}_{k,n}} \prod_{j=1}^{|S|} \sigma(s_j)^{\mathcal{N}_{k,j}}.$$

The deviation payoff  $\mathcal{U}_{\sigma,j}$  can then be computed (in alternative to (5)) as the normalized weighted combination of the profile payoffs:

$$\mathcal{U}_{\sigma,j} = \frac{\sum_{k|\mathcal{N}_{k,j}>0} \Pr(k | \sigma) \mathcal{U}_{k,j}}{\sum_{k|\mathcal{N}_{k,j}>0} \Pr(k | \sigma)}.$$

Tables 4 and 5 show the HPTs for six and ten replicators, respectively. As  $p$  increases, the replicator HPT provides a better approximation to the infinite game. Suppose the population is half  $C$  and half  $D$ . Figure 9 shows how the expected payoffs for  $C$  (top) and  $D$  (bottom) approach the true infinite-population value (green line) as the number of replicators increases. Convergence is relatively quick whether sampling is with (orange curve) or without (blue curve) replacement, but the former can provide a significantly better approximation for small numbers of replicators.

### 5. Incomplete Game Models

As noted in Section 3.2.2 above, it is not always feasible to simulate all strategy profiles, even over a restricted strategy space. Two approaches to deal with this are to generalize from the sampled profiles to fit a learned game model (Section 3.2.3), or to reason over a model covering only part of the profile space. Here we discuss EGTA methods developed for the latter option.

Table 4: Prisoner’s Dilemma HPTs with six replicators.

a No replacement				b With replacement.			
$\mathcal{N}_C$	$\mathcal{N}_D$	$\mathcal{U}_C$	$\mathcal{U}_D$	$\mathcal{N}_C$	$\mathcal{N}_D$	$\mathcal{U}_C$	$\mathcal{U}_D$
6	0	3	–	6	0	3	–
5	1	2.4	5	5	1	2.5	4.33
4	2	1.8	4.2	4	2	2	3.66
3	3	1.2	3.4	3	3	1.5	3
2	4	0.6	2.6	2	4	1	2.33
1	5	0	1.8	1	5	0.5	1.66
0	6	–	1	0	6	–	1

Table 5: Prisoner’s Dilemma HPTs with ten replicators.

a No replacement				b With replacement.			
$\mathcal{N}_C$	$\mathcal{N}_D$	$\mathcal{U}_C$	$\mathcal{U}_D$	$\mathcal{N}_C$	$\mathcal{N}_D$	$\mathcal{U}_C$	$\mathcal{U}_D$
10	0	3	–	10	0	3	–
9	1	2.66	5	9	1	2.7	4.6
8	2	2.33	4.55	8	2	2.4	4.2
7	3	2	4.11	7	3	2.1	3.8
6	4	1.66	3.66	6	4	1.8	3.4
5	5	1.33	3.22	5	5	1.5	3
4	6	1	2.77	4	6	1.2	2.6
3	7	0.66	2.33	3	7	0.9	2.2
2	8	0.33	1.88	2	8	0.6	1.8
1	9	0	1.4	1	9	0.3	1.4
0	10	–	1	0	10	–	1

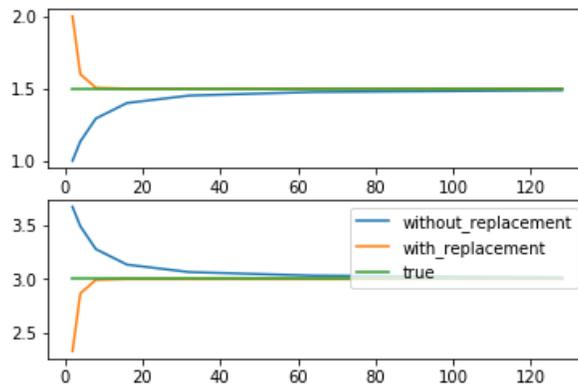


Figure 9: Expected payoffs approach the infinite-population value as we increase the number of replicators in the  $2 \times 2$  Prisoner’s Dilemma game.

To accommodate incomplete specification, we allow that the estimated payoff function in an empirical game model leave utility unassigned for some profiles. Formally, some profiles may be mapped to a null value,  $\hat{u}_i : \prod_{j \in N} S_j \rightarrow \mathbb{R} \cup \{\perp\}$ . A profile  $s$  such that  $\hat{u}_i(s) \in \mathbb{R}$  for all  $i$  is termed *evaluated*. Otherwise (i.e.,  $\hat{u}_i(s) = \perp$  for some  $i$ ), we say  $s$  is *unevaluated*.<sup>8</sup> If all  $s \in \prod_{j \in N} S_j$  are evaluated we say the empirical game  $\hat{\Gamma} = \langle N, (S_i), (\hat{u}_i) \rangle$  is *completely evaluated*.

A core problem within incomplete game models is to search for a profile (pure or mixed) that can be established to have sufficiently small regret. Several works have examined this problem from a search perspective. The first was due to Sureka and Wurman [2005], who proposed an algorithm to identify PSNE based on best-response dynamics combined with tabu search [Glover, 1989]. In their formulation, the basic operation is evaluation of a strategy profile by simulation until it is labeled evaluated, and the search successively evaluates profiles until a PSNE is reached. Jordan et al. [2008] termed this problem formulation the *revealed payoff model*, and extended the approach to include approximate equilibria, since PSNE may not exist. These authors also proposed an algorithm called *minimum-regret-first search* (MRFS). This algorithm maintains for each profile  $s$  evaluated a lower bound  $\bar{\epsilon}(s)$  on the profile's regret, defined as the maximum gain to deviating from  $s$  considering the deviation profiles evaluated. If all deviations have been evaluated, then  $\bar{\epsilon}(s)$  is the actual regret, so that  $s$  constitutes an  $\bar{\epsilon}(s)$ -Nash equilibrium. MRFS selects an evaluated profile with minimal  $\bar{\epsilon}$ , and then chooses an unevaluated deviation to simulate. This simple approach typically identifies and confirms low-regret profiles after exploring only a small fraction of the profile space.

## 5.1 Query Complexity

Fearnley et al. [2013] examined the problem from the algorithmic complexity perspective. They define the *query complexity* of a game class with respect to some solution concept as the number of profile evaluations that are required (under the revealed-payoff model) to identify a solution, in the worst case. For example, the solution concept may be  $\epsilon$ -NE, in which case the query complexity would generally depend on  $\epsilon$  as well as parameters of the game class. The authors investigated several settings, producing interesting query complexity results for bimatrix games as well as structured game classes with compact payoff representations [Fearnley et al., 2015].

For instance, one striking result for bimatrix games is that it is possible to guarantee finding an approximate NE with a linear number of queries. Consider a two-player game where each player has strategy set  $S$ , and assume payoffs lie in the range  $[0, 1]$ . Recall  $BR_i(s)$  denotes player  $i$ 's best response when player  $-i$  plays strategy  $s$ . The following algorithm identifies an  $\epsilon$ -NE for  $\epsilon = 0.5$  with  $2|S| - 1$  queries.

1. Choose an arbitrary action  $s^0 \in S$ .
2. Compute  $BR_2(s^0)$ . This can be accomplished in  $|S|$  queries; simply evaluate profiles  $(s^0, s)$  for all  $s \in S$  and select the  $s$  with greatest payoff as  $BR_2(s^0)$ .

---

8. This binary distinction is overly coarse, as profiles may also be evaluated to varying levels of accuracy or confidence. The methods described in the current section make this simplification, operating as though the evaluations are exact and correct. We present more statistical treatments of profile evaluation in Section 6.

3. Compute  $BR_1(BR_2(s^0))$ . This can be accomplished with an additional  $|S| - 1$  queries, of profiles  $(s, BR_2(s^0))$  for all  $s \in S \setminus s^0$ . Note that profile  $(s^0, BR_2(s^0))$  was already evaluated by a query in the preceding step.
4. Return the profile where player 1 plays  $s^0$  or  $BR_1(BR_2(s^0))$ , each with probability 0.5, and player 2 plays  $BR_2(s^0)$ .

To see that the profile returned by this algorithm is a 0.5-NE, we observe that each is playing a best-response to the other with probability 0.5. With the remaining probability they can be worse by at most one (the payoff range), thus 0.5 in expectation.

Fearnley et al. [2015] further show how the regret bound for bimatrix games can be tightened somewhat with more queries. Finding an exact equilibrium, however, requires exhaustive evaluation. On the other hand, games with known structure may enable alternative approaches, like generalization from queries across profiles (i.e., model learning, discussed in Section 3.2.3). If structure exists but is not known to the analyst, it still may be implicitly exploited by heuristic search methods, such as MRFS discussed above. These methods may be expected to perform better than worst-case, and thus continue to provide a practical means of performing EGTA, in the usual case where exhaustive evaluation of the profile space is infeasible.

## 5.2 Equilibrium Search over Subgames

The heuristic methods (tabu search and MRFS) mentioned above focus on finding single profiles that represent approximate pure equilibria. To verify that a mixed profile is in equilibrium (exact or approximate) from an incomplete game model, we require that the model contains full payoff information over the support of the profile, plus all one-player deviations from any pure profile in the support. Given such information, we can employ any game-solving algorithm for finding equilibria within the support space, and then check whether any of the deviations outside the support are beneficial. In this section, we define an analysis algorithm for incomplete game models based on searching over strategy subspaces for which full payoff information is specified. As necessary, the search process proposes profiles to extend the evaluated space so that an equilibrium can be identified and confirmed.

Let  $\Gamma$  be symmetric with strategy set  $S$ ,<sup>9</sup> and recall the notation  $\Gamma_{\downarrow X}$ ,  $X \subseteq S$ , for the restricted game over strategies  $X$ . We say that  $\Gamma_{\downarrow X}$  is a **complete subgame** if the empirical game  $\hat{\Gamma}_{\downarrow X}$  is completely evaluated. We typically focus attention on **maximal** complete subgames, where adding any strategy to  $X$  would render the subgame incomplete.

The game analysis algorithm maintains a set  $\Psi$  of candidate solution profiles of empirical game  $\hat{\Gamma}$ . For  $\sigma$  to qualify as a candidate, we require that  $\Gamma_{\downarrow \text{supp}(\sigma)}$  be a complete subgame, and that there be some completion of  $\hat{u}$  (i.e., assignment of payoff values to  $\perp$  entries) that renders  $\sigma$  a solution. If  $\sigma$  is a solution in *all* completions of  $\hat{u}$ , we say that  $\sigma$  is **confirmed**. Otherwise  $\sigma$  is **unconfirmed**.

For example, suppose the solution concept is symmetric mixed  $\epsilon$ -NE. To assess profile  $\sigma$  as a candidate solution, we would first verify that it is an  $\epsilon$ -NE of  $\hat{\Gamma}_{\downarrow \text{supp}(\sigma)}$ . We then consider deviation profiles of  $\sigma$ . In order to determine whether strategy  $s \in S \setminus \text{supp}(\sigma)$  is a beneficial deviation, we

9. The methods in this section can be generalized in a straightforward manner to handle role symmetry.

must evaluate all profiles of the form  $(s, s')$ , with  $s'$  an other-agent profile over  $supp(\sigma)$ . We then compare  $\mathbb{E}[\hat{u}(s, \sigma)]$  with  $\mathbb{E}[\hat{u}(\sigma, \sigma)]$ , and if the difference exceeds  $\epsilon$ , then  $\sigma$  is **refuted** by  $s$  and no longer considered a candidate. In other words,  $\sigma$  is by definition a candidate solution iff it is an  $\epsilon$ -NE of  $\hat{\Gamma}_{\downarrow supp(\sigma)}$  and there is no  $s$  that refutes  $\sigma$ . If  $\sigma$  is a candidate solution with all deviation profiles evaluated, then  $\sigma$  is confirmed.

Note that if a profile  $\sigma$  has regret bounded by  $\epsilon$  in empirical game  $\hat{\Gamma}$ , then it will also have regret at most  $\epsilon$  in any complete subgame  $\hat{\Gamma}_{\downarrow X}$  for which  $X$  contains  $supp(\sigma)$ .<sup>10</sup> The game analysis algorithm therefore simply runs standard equilibrium-finding methods for each maximal complete subgame to identify potential candidates. It then filters any that are refuted in the broader strategy space, and classifies the remaining as confirmed or unconfirmed: candidate sets  $\Psi_C$  and  $\Psi_U$  respectively.

Figure 10 shows how the algorithm sketched above (hexagon in the figure) can be incorporated within an iterative search for confirmed solutions. If the game analysis reveals unconfirmed candidates, we attempt to confirm or refute them by testing deviations. Specifically, we evaluate by simulation the profiles  $\bigcup_{\sigma \in \Psi_U} UD(\sigma)$ , where  $UD(\sigma)$  denotes the unevaluated deviation profiles of  $\sigma$ . Once these are all evaluated, each of the unconfirmed candidates is either refuted or confirmed.

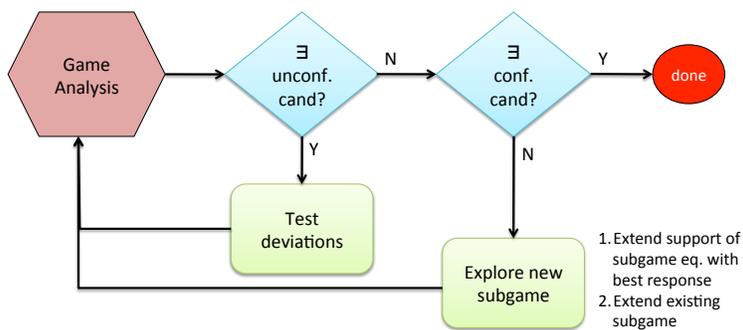


Figure 10: Search over complete subgames to identify and confirm solution candidates.

If at any point we have at least one confirmed candidate and no unconfirmed candidates, the procedure terminates and returns  $\Psi_C$ . If there are no candidate solutions, then we need to further explore the profile space. Evaluating additional profiles cannot affect results of the game analysis algorithm unless the additional profiles lead to completion of a new maximal subgame. Figure 10 lists two approaches for extending subgames. The first aims to complete subgames that appear promising based on existing results. Specifically, for any case where profile  $\sigma$  is an  $\epsilon$ -NE of a maximal subgame  $\hat{\Gamma}_{\downarrow S'}$ , but the best evaluated response to  $\sigma$  is some strategy  $s \in S \setminus S'$ , we consider the subgame defined by  $supp(\sigma) \cup \{s\}$ . If all such subgames have already been evaluated yet still no confirmed equilibria have been found,<sup>11</sup> the second approach nondeterministically chooses to extend one of the current

10. More generally, we observe that regret is monotone in strategy sets, in the following sense. Recall that  $\epsilon^\Gamma(\sigma)$  denotes the regret of profile  $\sigma$  in game  $\Gamma$ . Then  $\forall \sigma. supp(\sigma) \subseteq X \subset X' \implies \epsilon^{\Gamma_{\downarrow X}}(\sigma) \leq \epsilon^{\Gamma_{\downarrow X'}}(\sigma)$ .

11. It may seem counterintuitive, but this can happen. For example, consider a three-player symmetric game with three strategies  $\{A, B, C\}$  such that all profiles except  $(A, B, C)$  have been evaluated. We could have a situation exhibiting a non-transitive response pattern reminiscent of rock-paper-scissors, such as the following:  $(A, A, A)$  is the equilibrium of subgame  $\{A, B\}$ , with best-response  $C$ ;  $(B, B, B)$  is the equilibrium of subgame  $\{B, C\}$ , with best-response  $A$ ; and  $(C, C, C)$  is the equilibrium of subgame  $\{A, C\}$ , with best-response  $B$ . All best-response-enhanced subgames have been evaluated, but to find the true equilibrium (which has full support), we need the missing profile.

maximal subgames. In the worst case, this process can lead to exhaustive exploration of the profile space, which necessarily contains an  $\epsilon$ -NE, which would be confirmed at that point.<sup>12</sup>

## 6. Statistical Reasoning in Empirical Games

A defining feature of empirical game models is that they are estimated or induced from simulation data. Simulating a complex game typically involves stochastic factors, embedded in the game environment or in the players' strategies (i.e., exogenous or endogenous randomization, respectively). In our running example of sequential auctions, the main stochastic element in the environment is the initial random draw of player valuations. Notationally, we mark empirical game models  $\hat{\Gamma}$  with a hat as a reminder that their associated utility functions  $\{\hat{u}_i\}$  may differ from the *true game* utility functions  $\{u_i\}$  due to error inherent in the process of inducing the model from a sample of data. Many EGTA techniques operate on the game model as if it were perfectly accurate. For example, the revealed payoff model discussed in Section 3.2.2 treats each strategy profile as either completely unknown or exactly evaluated. In this section, we discuss methods that recognize the noise inherent in actual payoff samples, and explicitly consider the statistical character of empirical game models.

The first question we address is how to exploit statistical properties of simulator-generated data to improve the quality of payoff-function estimates. The next subsequent sections examine statistical questions about the game models themselves. That is, given a game model generated by simulation according to a sampling process, what kinds of statistical claims can we make about properties of the game and its solutions? Further, how might we design a process that interleaves sampling and game reasoning to produce an empirical game model supporting the sharpest possible conclusions about the true game?

### 6.1 Variance Reduction Methods

Estimating summary statistics from stochastic processes is a well-studied problem in simulation [Ross, 2002], and techniques from that field are directly applicable to estimation tasks in EGTA. In particular, methods that reduce variance in the sampling distribution for payoff estimates can substantially improve the reliability of game models from a given corpus of simulation data.

One variance-reduction technique that has been extensively applied in EGTA and other strategic analysis contexts is the method of *control variates* [Lavenberg and Welch, 1981, L'Ecuyer, 1994, McGeoch, 2012]. The key idea of this technique is to exploit correlation between the random variable of interest (e.g., payoffs of a strategy profile) and other observable variables (the control variates). In our sequential auction example, potential control variates might be some summary statistics on one's own and/or other-agent valuations. Payoffs for typical strategies are likely to be positively correlated with one's own valuation (representing potential for profit), and negatively correlated with others' (representing degree of competition). A given sample may be more or less favorable, and over a course of sampling it might be that some strategies tended to be simulated in

---

12. To guarantee that the procedure returns a confirmed solution candidate requires that the equilibrium-finding procedure applied to subgames is itself assured to return a solution. In particular, RD alone is not sufficient.

more favorable instances than others. The control variates method effectively adjusts estimates for this kind of luck, based on the observed favorability of sampled instances compared to expectation.

An early example of the use of control variates in EGTA was in a study of strategic procurement in a supply chain game [Wellman et al., 2005a]. In this game—part of the Trading Agent Competition series—agents representing computer manufacturers competed to buy parts and assemble computers for sale over the course of a simulated year. Simulating this environment is expensive (seven CPU-hours per sample), and samples are quite noisy. The noise is due to high variance in consumer demand, which played a significant role in payoffs for many strategies. To apply demand level as a control variate, the simulation data was employed to estimate a linear model of payoff as a function of demand, reflecting the underlying correlation between these variables. This linear model was then combined with the known demand distribution to yield an improved estimate of expected payoff, decreasing the amount of simulation required by up to 50%, compared to estimating payoffs directly.

Control variates and related variance reduction methods have also been applied in analyses of other TAC games [Jordan et al., 2010b, Sodomka et al., 2007, Wellman et al., 2007]. Similar techniques were developed to evaluate AI poker strategies [Burch et al., 2018], which proved essential for deriving confident statistical comparisons among strategies that are quite close in strength.

## 6.2 Statistical Characterization of Empirical Games

Analysts may be interested in various characteristics of a game: its equilibria or other solution concepts, welfare properties such as price of anarchy, etc. Since an empirical game model is induced from simulation data drawn from the true game, its accuracy is subject to sampling error, and results from analyzing the empirical model are related only probabilistically to properties of the true game. Researchers have investigated this probabilistic relationship both theoretically and experimentally. Theoretical studies have sought to derive probabilistic guarantees relating the results from empirical game analysis to properties of the true game, given characteristics of the sampling process or model accuracy. Experimental studies have sought ways to measure statistical performance of EGTA techniques, or to estimate the reliability of model results deriving from statistical observations.

In one of the first statistical EGTA investigations, Vorobeychik [2010] showed that as we approach infinitely many i.i.d. simulation queries, estimation by averaging sample payoffs converges to an empirical game reflective of the true game. In particular, the set of equilibria of a true game and its empirical counterpart coincide. He further noted that with only finitely many samples, spurious equilibria (i.e., false positives) can arise in the empirical game.

To state this more precisely, let  $\mathcal{E}(\Gamma)$  denote the set of equilibria (according to some specified solution concept) of game  $\Gamma$ . We say that a profile  $\sigma$  is a *spurious equilibrium* of the empirical game  $\hat{\Gamma}$  if  $\sigma \in \mathcal{E}(\hat{\Gamma})$  but  $\sigma \notin \mathcal{E}(\Gamma)$ . We can quantify the prevalence of spurious solutions using the language of *precision* and *recall*, in the sense of information retrieval [Salton and McGill, 1983]: if all equilibria of the true game are “recalled” as equilibria in the empirical game—that is, if  $\mathcal{E}(\Gamma) \subseteq \mathcal{E}(\hat{\Gamma})$ —this constitutes perfect recall. More generally, the degree of recall is measured by the fraction

$$\frac{|\mathcal{E}(\Gamma) \cap \mathcal{E}(\hat{\Gamma})|}{|\mathcal{E}(\hat{\Gamma})|}.$$

Conversely, if  $\mathcal{E}(\hat{\Gamma}) \subseteq \mathcal{E}(\Gamma)$  (i.e., there are no spurious equilibria in the empirical game), precision is perfect, and the degree of precision can be measured as above, but with  $|\mathcal{E}(\hat{\Gamma})|$  as the denominator.

To illustrate this phenomenon, we conducted a simple experiment using our running sequential-auctions example. The game has three players and four goods, and three available bidding strategies:  $\rho \in \{0, 0.5, 1\}$ . Valuations are based on the homogeneous-good model of Wellman et al. [2017], with maximum values drawn from player-specific distributions. We generated 100 random games, and for each game evaluated the corresponding empirical game after various numbers of samples (100, 200, 500, and 1000) per profile.

The results are shown in Figure 11. For each empirical game, we tested whether each of the 27 pure profiles was an approximate PSNE, with approximation threshold  $\epsilon$  set based on an empirical version of Bennett’s inequality [Cousins et al., 2022] for the given number of samples. We then plot, for each sample level and profile, in how many of the 100 games that profile was deemed an  $\epsilon$ -PSNE in the empirical game after that many samples. As we can see, at 100 samples, there are many spurious equilibria, which are steadily eliminated as the number of samples increases.

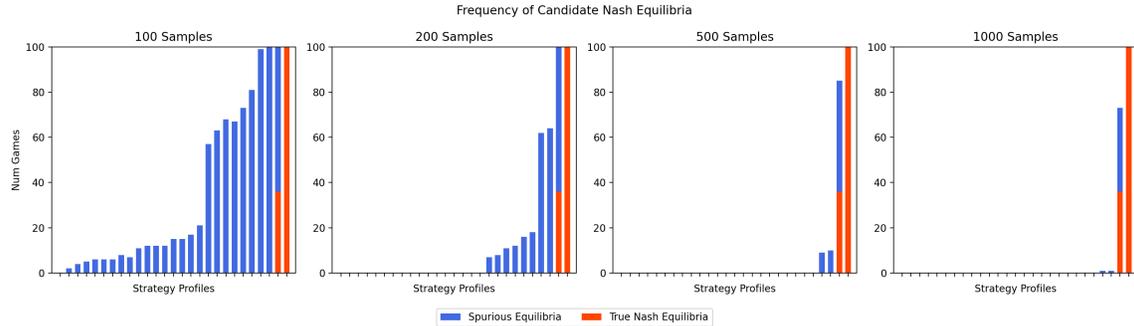


Figure 11: Spurious and true equilibria as a function of the number of samples. Bar height represents frequency of identification as empirical  $\epsilon$ -PSNE in 100 random games, with blue indicating spurious and red, true equilibria. Strategy profiles whose bars are both blue and red were spurious in some games, and true in others, in the proportions shown. In each plot, the order of the 27 strategy profiles on the  $x$ -axes is fixed, from highest to lowest regret, computed using 10,000 samples. “True” equilibria were likewise found using 10,000 samples.

### 6.3 Sampling and Approximation Bounds

A *uniform approximation* of a game is one in which all utilities are estimated to within the same error, simultaneously. At well-defined regret levels, uniform approximations of games support perfect recall and approximate precision: perfect recall, because the set of approximate equilibria of  $\hat{\Gamma}$  contains all true positives—that is, all  $\Gamma$ ’s equilibria; and approximate precision, because all false positives (i.e., spurious equilibria) are approximate equilibria in  $\Gamma$ . Stated more precisely,  $\hat{\Gamma}$  with utility  $\hat{u}$  is said to be an  $\epsilon$ -*uniform approximation* of  $\Gamma$  with utility  $u$  iff  $\|u - \hat{u}\|_\infty = \max_{i \in N, s \in S} |u_i(s) - \hat{u}_i(s)| \leq \epsilon$ . Then, letting  $\mathcal{E}_\epsilon(\Gamma)$  denote the set of  $\epsilon$ -Nash equilibria, if  $\hat{\Gamma}$  is an  $\epsilon$ -uniform approximation of  $\Gamma$ , then  $\mathcal{E}(\Gamma) \equiv \mathcal{E}_0(\Gamma) \subseteq \mathcal{E}_{2\epsilon}(\hat{\Gamma}) \subseteq \mathcal{E}_{4\epsilon}(\Gamma)$  [Tuyls et al., 2020, Areyan Viqueira et al., 2020].

It follows that any algorithm that learns an  $\epsilon$ -uniform approximation of a game also learns the equilibria of that game, up to an accuracy that depends on  $\epsilon$ . A standard approach to learning uniform approximations uses concentration inequalities to first establish high-probability confidence intervals around each individual parameter (in our case, a player’s utility at a strategy profile), and to then apply a statistical correction (e.g., Bonferroni or Šidák (1967)) to bound the probability that all approximation guarantees hold simultaneously.

Tuyls et al. [2020] were the first to try this approach, using a *global sampling* (GS) algorithm where the empirical game model is estimated from batches of  $m$  samples per profile. To derive their guarantee, the authors used Hoeffding’s inequality, a sub-Gaussian tail bound for averages of  $m$   $c/2$ -bounded random variables that yields a supremum deviation bound  $\|u - \hat{u}\|_\infty \leq \epsilon \in \Theta\left(c\sqrt{\frac{\ln(2/\delta)}{2m}}\right)$  with probability at least  $1 - \delta$ . For a constant failure probability  $\delta$ , rather than compute the accuracy  $\epsilon$  given sample size  $m$ , it is often useful to calculate the sample size  $m$  required to achieve a target error  $\epsilon$ . Applying this logic, and a Bonferroni correction, GS requires  $m \geq \frac{c^2 \ln(2|\Gamma|/\delta)}{2\epsilon^2}$  samples per profile to guarantee an  $\epsilon$ -uniform approximation of the game with probability at least  $1 - \delta$ , where  $|\Gamma|$  is the number of game parameters (e.g., utility values).

As Hoeffding’s inequality is sensitive only to the range  $c$  of the random variables, not their variance  $\sigma^2$ , Areyan Viqueira et al. [2020] analyze GS with an alternative concentration inequality. Bennett’s inequality [1962] relaxes the dependency on  $c$ , replacing it with a dependency on  $\sigma^2 < c^2$ , yielding  $\epsilon \in \Theta\left(\frac{c \ln(1/\delta)}{m} + \sqrt{\frac{\sigma^2 \ln(1/\delta)}{m}}\right)$ , which in this case implies that  $m \geq 2 \ln \frac{2|\Gamma|}{\delta} \left(\frac{c}{3\epsilon} + \frac{\|v\|_\infty}{\epsilon^2}\right)$  samples per profile are required to guarantee an  $\epsilon$ -uniform approximation of the game with probability at least  $1 - \delta$ . This bound depends on the so-called *wimpy variance*, the maximum variance across all game parameters, denoted by  $\|v\|_\infty$  [Boucheron et al., 2013]. Using an upper bound on the wimpy variance in terms of its empirical counterpart [Cousins and Riondato, 2020], one can derive a supremum deviation bound that does not depend on any *a priori* variance knowledge.

Figure 12 depicts the sample complexity  $m$  of GS as a function of  $1/\epsilon$ , for  $\delta = 0.05$ , calculated according to four methods. The methods employ Hoeffding’s bound (GS-H), as well as three variants of Bennett’s bound: Bennett’s original bound assuming known wimpy variance (GS-B); Areyan Viqueira et al.’s empirical Bennett bound based on the actual empirical wimpy variance realized in the experiments (GS-EB); and a third upper bound derived by Cousins et al. [2022] (GS-EB Upper Bound). The first and third of the aforementioned curves are smooth, because they are obtained by plugging a “known” variance (determined by 30,000 samples) and a target error  $\epsilon$  into a formula that produces  $m$ , a sample complexity. In contrast, in the GS-EB experiments, the independent variable is the number of samples, while the dependent variable is  $\epsilon$ , as per Areyan Viqueira et al.’s empirical Bennett bound.

These sample complexities were calculated in four bidding games, each with three bidders, one good, and six shading factors ( $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ ). The games varied in that the bidders’ values were drawn from four different beta distributions, which yield four different “known” wimpy variances, as indicated in the figures. When wimpy variance is low (left), Bennett’s bound significantly outperforms Hoeffding’s; when wimpy variance is high (right), Bennett’s bound still performs within a small constant factor of Hoeffding’s.

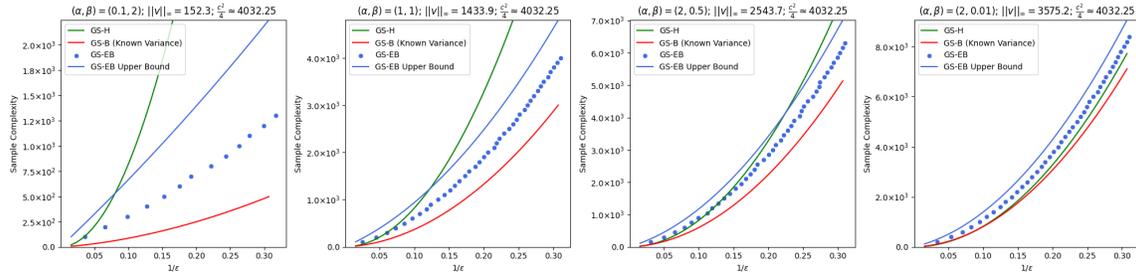


Figure 12: Sample complexity  $m$  of GS as a function of  $1/\epsilon$ , calculated using four bounding methods. Variance increases across the instances plotted moving left to right. Algorithms based on Bennett’s bound are most effective in the leftmost plot, and least in the rightmost. The algorithm based on Hoeffding’s bound is independent of variance; the GS-H curve appears to shift due to compression of the y-axis scale from left to right.

	0.3	0.7		0.3	0.7		0.3	0.7
0.3	$\begin{pmatrix} 5.68 & 1.71 \end{pmatrix}$		0.3	$\begin{pmatrix} 5.50 & 1.96 \end{pmatrix}$		0.3	$\begin{pmatrix} 5.85 & 1.61 \end{pmatrix}$	
0.7	$\begin{pmatrix} 4.40 & 2.43 \end{pmatrix}$		0.7	$\begin{pmatrix} 4.20 & 2.66 \end{pmatrix}$		0.7	$\begin{pmatrix} 3.93 & 3.24 \end{pmatrix}$	

Figure 13: Statistical analysis of an empirical game for a simple sequential auction example. The game is symmetric, and payoffs are shown for the row player. Left: true game  $\Gamma$ ; Middle: empirical game  $\hat{\Gamma}$  ( $m = 20$ ); Right: resampled version  $\hat{\Gamma}'$  of empirical game.

### 6.4 Measuring Uncertainty in Empirical Game Analysis

Once an empirical game model is constructed and analyzed, we often wish to quantify the reliability of results, accounting for sampling error. Even if the model was estimated based on GS or other approaches that provide *a priori* approximation bounds, an *a posteriori* analysis may yield further precision about the uncertainty attached to game-theoretic conclusions.

For example, let us define a simple version of our sequential auction game, with  $n = g = 2$ , two available strategies ( $\rho \in \{0.3, 0.7\}$ ), and three possible valuations drawn uniformly and independently for each player. Thanks to its simplicity, we can work out the exact normal-form game  $\Gamma$ , shown in Figure 13(left). The middle payoff matrix in Figure 13 presents an empirical game  $\hat{\Gamma}$  for these settings, estimated from  $m = 20$  samples per profile.

This game has  $c = 28$  and a known wimpy variance of 13.7, so by the GS-B bounds of the previous section an empirical game generated with  $m = 20$  samples/profile has at least 0.95 probability of being an  $\epsilon$ -approximation with  $\epsilon \approx 5.63$ . Such a bound is not very helpful given the payoff scale of the empirical game generated, and as we can see, it is actually a much better approximation than that.<sup>13</sup> So after we build an empirical game, what can we say about the accuracy of its conclusions?

First, let’s solve  $\hat{\Gamma}$ . It has a unique symmetric NE  $\sigma_{\hat{\Gamma}}^*$ , which plays  $\rho = 0.3$  with probability 0.35. As discussed above, given sampling error, we do not expect empirical-game solutions to be exact

13. Of course, in a real application we would not know the true game. In this example, where the number of samples  $m$  is tiny, so we can expect weak bounds. But the bounds are nonetheless quite conservative; so we may have been lucky, but not extraordinarily so.

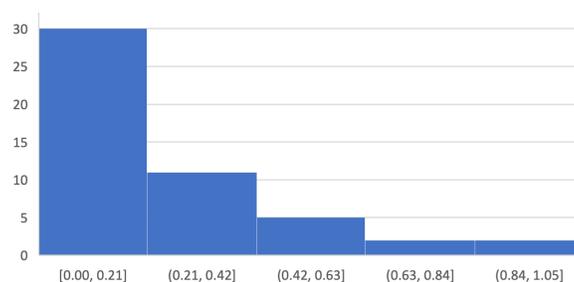


Figure 14: Bootstrapped regret distribution for the empirical game  $\hat{\Gamma}$  of Figure 13(middle).

solutions of the true game. The question of interest, rather, is how likely is it to be an approximate equilibrium, at various levels of approximation? In other words, what is the probability distribution over  $\epsilon^{\Gamma}(\sigma_{\hat{\Gamma}}^*)$ , viewed as a random variable conditional on the samples that produced  $\hat{\Gamma}$ ?

Wiedenbeck et al. [2014] proposed a bootstrap approach for addressing such statistical questions about empirical games. The basic idea of bootstrapping is to use the sampling data itself as the basis for modeling uncertainty in data generation. This technique has been applied to a broad variety of statistical questions [Davison and Hinkley, 1997], and is amenable to straightforward implementation [Shasha and Wilson, 2011]. To bootstrap the sampling distribution for regret, we *resample* with replacement from the payoff data to construct new game models, a process that we assume captures the distribution of empirical games that would arise from random sampling.

In our example, empirical game  $\hat{\Gamma}$  was estimated from  $m = 20$  samples for each cell of the game matrix. For the resampled version  $\hat{\Gamma}'$  shown in Figure 13 (Right), each cell is the average of 20 payoffs resampled with replacement from the original data set. Next, we calculate the regret of the profile of interest with respect to the resampled game model. In this instance,  $\epsilon^{\hat{\Gamma}'}(\sigma_{\hat{\Gamma}'}^*) = 0.14$ . This value represents one data point in the sampling distribution for regret. We resampled 49 more times to build the histogram shown in Figure 14. According to this histogram, the expected (mean) regret is 0.24, the median is 0.16, and the 95% confidence level is 0.72. In this example, the true-game regret turns out to be at the optimistic extreme of the bootstrap distribution,  $\epsilon^{\Gamma}(\sigma_{\hat{\Gamma}}^*) = 0.01$ . Nonetheless, Wiedenbeck et al. [2014] found that bootstrapping regret from the empirical game was often well-calibrated with regret in the true game.

The bootstrapping approach described above attempts to quantify uncertainty using the data already generated for estimating the empirical game. It may also be valuable to perform additional sampling expressly for the purpose of assessing reliability in game-theoretic solutions. Jecmen et al. [2020] propose a bandit-based algorithm for sampling profiles and deviations in order to bound regret estimates for profiles in an empirical game model. Rowland et al. [2019] design sampling strategies to achieve confidence guarantees in orderings produced by  $\alpha$ -Rank.

## 6.5 Dynamic Sampling Algorithms

Measurements of statistical confidence in EGTA can be useful not just for assessing conclusions, but also to guide the collection of samples during the EGTA process. The idea of statistical sample control for EGTA was first pursued by Walsh et al. [2003]. In that early study, where the goal

was to learn Nash equilibria, the authors proposed selecting strategy profiles to sample based on *expected confirmational value of information*, a prediction of the degree by which sampling would decrease estimated error in the current solution candidate. Jordan et al. [2008] framed the statistical sample control task within EGTA based on a *noisy payoff model*, by contrast with the revealed payoff model discussed in Section 3.2.2. These authors proposed an algorithm based on maximizing information gain, defined relative to a mapping from empirical games to beliefs about strategies played [Vorobeychik and Wellman, 2006]. Using the method for bootstrapping regret distributions described above, Wiedenbeck et al. [2014] proposed using confidence intervals on these distributions for sample control. In particular they introduced heuristics based on bootstrap estimates that guide how to allocate and when to stop sampling, and found that they outperformed more common rules of thumb.

Several works frame EGTA as a *black-box optimization* problem, where the objective function is given not analytically, but in the form of a simulator [Audet and Hare, 2017]. More specifically, the *Bayesian optimization* approach to black-box optimization [Garnett, 2023] employs an *acquisition function* to guide sampling from the simulator. To apply Bayesian optimization to game-solving, a natural acquisition function would be based on regret (1), though as the regret function  $\epsilon$  is not available analytically, a surrogate must be used. Al-Dujaili et al. [2018] use Gaussian process (GP) regression to build a probabilistic game model, from which they estimate  $\epsilon_i(s)$  as the maximal difference across all strategies  $s'_i$  between  $i$ 's mean utility  $\bar{u}_i(s'_i, s_{-i})$  plus one standard deviation, and the mean utility  $\bar{u}_i(s)$  of  $s$ . They then propose the minimum of the maximum of this regret across all players as the next sample with probability  $1 - \eta$ , for some  $0 < \eta \ll 1$ , choosing at random according to the uncertainty of the GP, otherwise. Tay et al. [2023] take a similar GP approach, also basing their acquisition function on maximum regret, but in a worst-case sense using confidence bounds: the difference between the lower limit of the confidence bound for a given strategy (a pessimistic outcome) and the upper limit across all alternatives (an optimistic one). Picheny et al. [2019] likewise employ GP, and propose two acquisition functions. Their first is designed to maximize the probability that each player's strategy is part of an equilibrium (i.e., yields no regret), while their second is designed to reduce uncertainty in this measure.

Dynamic sampling methods like these can improve over global sampling (GS) by allocating sampling effort to profiles based on relevance as determined through intermediate analysis. For example, if the goal is to find equilibria, then we can avoid sampling strategy profiles which we deem sufficiently unlikely to affect equilibrium determination. More generally, if it can be established that a strategy profile's variance is sufficiently lower than another's, it may not be necessary to query the first as often as the second. These observations, respectively, form the basis of *progressive sampling with pruning* (PSP) algorithms [Areyan Viqueira et al., 2020, Cousins et al., 2022].

Using GS as a subroutine, a progressive sampling algorithm builds an empirical game iteratively, based on progressively larger samples, and hence more refined utility estimates, until a desired accuracy is achieved. PSP builds on this idea by pruning game parameters between iterations (i.e., ceasing to estimate them) if a certain pruning criterion is met. Cousins et al.'s algorithm prunes low-variance parameters before high-variance parameters, as soon as it establishes that they have been sufficiently well estimated (PSP-WE). With the goal of learning equilibria, Areyan Viqueira et al.'s algorithm prunes the utilities of strategy profiles once it establishes that they are not likely an approximate equilibrium (PSP-REG).

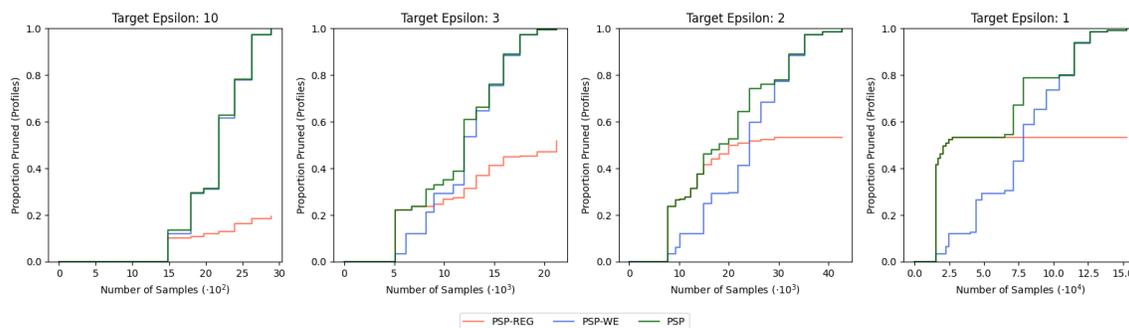


Figure 15: The proportion of strategy profiles pruned by each algorithm for each target  $\epsilon$ . The red curves correspond to PSP-REG; the blue curves, to PSP-WE; and the green curves, to PSP, which prunes using both criteria. The white space between the red and blue curves increases as the target error decreases, corresponding to the increase in significance of regret vs. well-estimated pruning.

Figure 15 depicts the proportion of strategy profiles pruned, hence the work saved, for three PSP variants on an instance of the sequential auctions game. Our setting employed homogeneous valuations [Wellman et al., 2017] with three bidders, four goods, and six shading factors ( $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ ). We used Cousins et al.’s PSP-WE sampling schedule, with  $\beta = 1.1$ , which is tailored to achieve a desired accuracy, for which we set four targets, namely  $\epsilon \in \{10, 3, 2, 1\}$ . We report the proportion of strategy profiles pruned by each algorithm for each target, averaged across five sample games.

We observe that as desired accuracy increases (i.e., as target  $\epsilon$  decreases), the PSP-REG curves eventually stabilize, whereas the PSP-WE curves maintain their basic shape, stretched across the number of samples. These observations are in line with intuitions, as regret pruning is independent of the desired accuracy (it depends only on the number of samples), whereas well-estimated pruning depends on the target. Indeed, Cousins et al. provide a lower bound on number of samples required to prune a strategy profile, which is inversely proportional to the target. As a result, PSP-REG’s contribution to PSP relative to PSP-WE’s increases in significance as the desired accuracy increases.

## 6.6 Game Properties beyond Equilibrium

Whereas the property of greatest interest for EGTA has typically been identification of solutions (e.g., Nash equilibria), statistical reasoning is relevant for other game properties as well, such as social welfare of solution profiles. A property  $f$  is called  $\lambda$ -Lipschitz if  $\|f(\cdot; u) - f(\cdot; u')\|_\infty \doteq \sup_{x \in \mathcal{X}} |f(x; u) - f(x; u')| \leq \lambda \|u - u'\|_\infty$ .<sup>14</sup> For example, common variants of social welfare—utilitarian, egalitarian, and Gini—are all  $\lambda$ -Lipschitz with  $\lambda = 1$  [Beliakov et al., 2009, Cousins et al., 2023]. Regret is  $\lambda$ -Lipschitz with  $\lambda = 2$  [Areyan Viqueira et al., 2020].

Observe the following: if  $f$  is  $\lambda$ -Lipschitz and  $\|u - u'\|_\infty \leq \epsilon$ , then  $\|f(\cdot; u) - f(\cdot; u')\|_\infty \leq \lambda\epsilon$ . Equivalently, if  $f$  is  $\lambda$ -Lipschitz and  $\|u - u'\|_\infty \leq \epsilon/\lambda$ , then  $\|f(\cdot; u) - f(\cdot; u')\|_\infty \leq \lambda(\epsilon/\lambda) = \epsilon$ . Thus, as regret is 2-Lipschitz, it can be  $\epsilon$ -approximated from an  $\epsilon/2$ -uniform approximation of a

14. Here,  $\mathcal{X}$  is usually the set of strategy profiles, but it need not be.

game. Likewise, utilitarian, egalitarian, and Gini social welfare can all be  $\epsilon$ -approximated directly from an  $\epsilon$ -uniform approximation [Cousins et al., 2023].

Based on this observation, Cousins et al. [2023] derive two-sided approximation bounds on the *extrema* of  $\lambda$ -Lipschitz game properties (e.g., maximum utilitarian welfare). Specifically, they derive a two-sided bound on their values, and a dual containment (recall and approximate precision) result characterizing their witnesses. This theorem implies that global sampling can be used to learn any  $\lambda$ -Lipschitz properties of games beyond regret/Nash equilibrium. Pruning algorithms based on game properties other than regret have not yet been fully explored in EGTA, although welfare-based pruning has been analyzed to learn competitive equilibria [Areyan Viqueira et al., 2021].

In summary, statistical tools offer EGTA practitioners guidance in tackling questions about the sample complexity required to achieve a desired accuracy when estimating a game’s properties, and how to distribute the requisite number of queries across the game’s various strategy profiles.

## 7. Strategy Exploration

The EGTA methods discussed in this survey thus far take the empirical game strategy sets,  $X_i$ , as given and fixed. Techniques for dealing with incomplete game models, statistical (Section 6) or otherwise (Section 3.2.2), allow for partial evaluation over the space of profiles induced by these strategy sets, iteratively extended through simulation queries. The subgame search methods (Section 5.2) also iteratively extend the game model, reasoning across different strategy-set restrictions. Nevertheless, all methods discussed to this point treat the base set of strategies defining the overall profile space as constant.

### 7.1 Automated Strategy Generation

Whereas manual specification of heuristic strategies often provides a good starting point for game-theoretic analysis, restricting attention to these sets a priori fundamentally limits the scope of the analysis. One way to address this limitation is through *automated strategy generation*: a process of search through the unrestricted sets  $S_i$  of possible strategies for new strategies to add to the restricted sets  $X_i$  delimiting the empirical game model. An iterative process for EGTA with automated strategy generation is illustrated in Figure 16. The box on the left, termed *inner loop* by Wellman et al. [2013] (essentially corresponding to the blue-arrow cycle of Figure 1), iteratively performs simulation, game model induction (estimation or learning), and game analysis, within a fixed and restricted space of strategies. For example, the inner loop might implement the process of Figure 10, and might include the sample-control methods of Section 6. An *outer loop* (incorporating the green-arrow path of Figure 1) employs results from analyzing the restricted empirical game  $\hat{\Gamma}_{\downarrow X}$  to generate new strategies from the base game for inclusion.

Ultimately, this iterative analysis is still limited by the restriction of the empirical game model to the set of strategies generated. However, by invoking unrestricted search as part of the process, the analyst consults the full sets  $S_i$  repeatedly. This tends to strengthen the analysis, as the intermediate EGTA results can provide an informed basis for selecting strategies to evaluate explicitly. We would generally expect better coverage from strategy sets generated in an informed manner incrementally

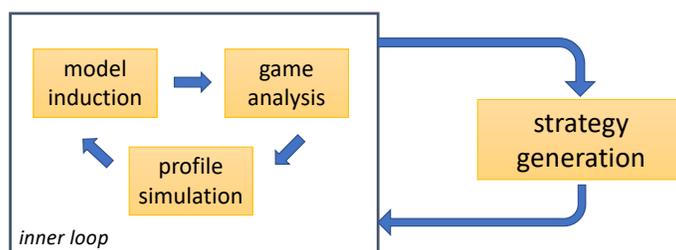


Figure 16: EGTA with automated strategy generation. The inner loop constructs and analyzes empirical game models over restricted strategy sets  $X_i$ . The outer loop searches over  $S_i$  to generate one or more new strategies to include, based on this analysis. These new strategies are then added to the  $X_i$ , and the process iterates.

over multiple (up to  $|X|$ ) iterations, compared to an a priori identification of  $|X|$  strategies from a parameterized heuristic space.

How to incrementally extend a game model in an informed manner is what we call the **strategy exploration problem** [Jordan et al., 2010a]. To illustrate the problem, consider the simple  $4 \times 4$  game shown in Table 6. The strategy exploration problem asks in which order to introduce the strategies to our empirical game analysis. Note that regardless of the ordering, once  $X = S$ , equilibria in the restricted game and base game coincide, so regret is zero. Thus, we might expect that regret would tend to start high, and decrease progressively until reaching zero in the last step. This is not necessarily the case, however. Introducing strategy 1 first, for example, would produce the solution profile  $(1, 1)$  after the first iteration, which has a regret  $\epsilon(1, 1) = 3$ . If we then introduce additional strategies in the order  $(2, 3, 4)$ , the additional sequence of regrets we observe would be  $(4, 5, 0)$ , thus increasing monotonically until inevitably falling to zero at the very end.

	1	2	3	4
1	1,1	1,2	1,3	1,4
2	2,1	2,2	2,3	2,6
3	3,1	3,2	3,3	3,8
4	4,1	6,2	8,3	4,4

Table 6: An example symmetric two-player game of four strategies [Jordan et al., 2010a]. Exploring strategies in the sequence  $(1, 2, 3, 4)$  yields increasing regrets until the last step.

As the example suggests, it will be difficult to *guarantee* progress throughout the strategy exploration process. That does not mean, however, that exploration decisions are arbitrary. Indeed, it can make quite a difference how one selects among strategies to add. To compare alternative exploration policies, we generally evaluate their performance in expectation, with respect to distributions of games and stochastic elements of the EGTA process with automated strategy generation.

One natural approach to strategy exploration is the **double oracle** (DO) method of McMahan et al. [2003]. DO was originally defined for two-player games,<sup>15</sup> though the idea readily generalizes to  $n$

15. The original paper likewise defined and analyzed DO for zero-sum games, but it can be applied without modification in the general-sum case.

players.<sup>16</sup> Let  $X^k = (X_1^k, \dots, X_n^k)$  denote the restricted strategy sets corresponding to iteration  $k$ , with  $\sigma^{*k}$  a NE profile for  $\Gamma_{\downarrow X^k}$ .  $BR_i$  is a best-response oracle for player  $i$ . DO augments the strategy sets for the next iteration with the best responses to the current iteration’s NE:  $X_i^{k+1} = X_i^k \cup \{BR_i(\sigma_{-i}^{*k})\}$ . If the best responses are already contained in the strategy sets,  $BR_i(\sigma_{-i}^{*k}) \in X_i^k$ , a NE for the full game has been found, and the process can terminate. More typically, DO proceeds until the gains fall below a threshold, or until a time or iteration limit has been reached.

The first application of automated strategy generation in EGTA was the use of genetic optimization by Phelps et al. [2006] for bidding in a continuous double auction (CDA). Schwartzman and Wellman [2009, 2010] demonstrated the use of RL (non-deep: with tile-coded Q-functions) for optimization, also in CDA trading environments. The latter studies were essentially instances of DO using RL to approximate the best-response oracle. Across several trading games, these approaches were shown to produce new CDA strategies exceeding the performance of hand-coded predecessors. Automated strategy generation using local search methods also played a role in EGTA applications to protocol compliance [Wellman et al., 2013] and credit network formation [Dandekar et al., 2015].

## 7.2 PSRO: Policy-Space Response Oracles

Lanctot et al. [2017] introduced *policy-space response oracles* (PSRO) as a general procedure combining deep RL with EGTA. Pseudocode for the PSRO algorithm, expressed in the terminology and notation of this survey, is presented as Alg. 1. The algorithm maintains an empirical game,  $\hat{\Gamma} = \langle N, (X_i), (\hat{u}_i) \rangle$ , which models the game  $\Gamma_{\downarrow X}$  over the strategy space  $X$  as of the current PSRO iteration. The empirical payoff matrix  $\hat{u}$  can be computed and updated each iteration using any method for estimating or learning empirical games. From  $\hat{\Gamma}$ , PSRO derives a target profile  $\sigma$ . For each player  $i$ , it then uses deep RL to train a policy  $\pi'_i$  that optimally responds to  $\sigma_{-i}$ . As the *training target*  $\sigma_{-i}$  is a mixed profile, the method operates by sampling during training. This deep RL computation represents the *response oracle* that gives PSRO its name.

PSRO’s signal innovation was introducing the concept of *meta-strategy solvers* (MSSs), a generalized approach to selecting training targets. The MSS selects an other-agent profile to train against, which in this framework is the essential driver of strategy exploration. Double oracle as described above can be viewed as a special case of PSRO where the training target is the other-agent profile in a Nash equilibrium (i.e., the MSS is an NE-finding algorithm). Though DO is often effective, there is ample evidence that best-response to NE is not always the best approach to strategy exploration. Jordan et al. [2010a] demonstrate this for a simple auction game, where even adding random strategies could provide substantial speedups. More generally, Lanctot et al. [2017] argued that best-responding to Nash overfits to the current equilibrium strategies, and thus tends to produce results that may not be generally effective across the relevant strategy space.

This was indeed the motivation for employing a generalized MSS concept in PSRO, maintaining the principle of best-response but allowing the target to vary. Lanctot et al. [2017] propose several alternative MSSs, for example, their *projected replicator dynamics* method constrains the training target to include every strategy generated so far with at least some minimum probability. This ensures that a broad set of opponent strategies are encountered during training. It also limits the

16. Current usage convention retains the “double” part of the name for any  $n$ .

**input** : initial strategy sets  $X_1, \dots, X_n$   
 Estimate empirical payoff functions  $\hat{u}_i$  by simulating profiles in  $X = X_1 \times \dots \times X_n$  ;  
 Initialize training targets  $\sigma_i \leftarrow \text{UNIFORM}(X_i)$  ;  
**while** *within iteration limit* **do**  
   **for** *player*  $i \in N$  **do**  
     **for** *many episodes* **do**  
       Sample  $\pi_{-i} \sim \sigma_{-i}$  ;  
       Train policy  $\pi'_i$  with other agents playing  $\pi_{-i}$  ;  
     **end**  
      $X_i \leftarrow X_i \cup \{\pi'_i\}$  ;  
   **end**  
   Extend  $\hat{u}_i$  to cover newly added strategies ;  
   Compute new training target  $\sigma \leftarrow \text{MSS}(\hat{\Gamma})$  ;  
**end**  
 Extract solution from final  $\hat{\Gamma}$  ;

**Algorithm 1:** PSRO pseudocode [Lanctot et al., 2017].

change in training targets from one round to the next, which serves to inhibit thrashing, in which an agent trained in one round throws away what the previous agent learned, due to drastic changes in the training-target mixed strategy.

The MSS abstraction is quite flexible, allowing PSRO to cover some classic game-solving approaches as special cases. For example, selecting a uniform distribution over strategies  $X^e$  as MSS essentially reproduces the fictitious play algorithm [Brown, 1951].<sup>17</sup> An MSS that simply extracts the most recent strategy corresponds to *iterated best response*, which in a symmetric context is also termed *self play* [Silver et al., 2018].

Any solution concept can readily be adopted as an MSS. Muller et al. [2020] employ an MSS based on  $\alpha$ -Rank, and Marris et al. [2021] define MSSs that correspond to correlated equilibrium concepts or employ solution refinement criteria. Combinations are also possible; for example Wang et al. [2019] employed a mixture of NE and uniform, which essentially randomizes over whether to apply DO or FP on a given PSRO iteration.

Balduzzi et al. [2019] introduced an MSS specifically for zero-sum games, called *rectified Nash*. Rectified Nash includes in the training target the subset of opponent strategies supported by the current equilibrium that the latest strategy beats or ties. The idea behind this approach is to expand the scope of existing strategies by building on their strengths. Dinh et al. [2022] proposed a MSS for two-player zero-sum games that applies online learning to the empirical game and outputs the online profile as a best-response target.

Whereas the MSS abstraction provides significant flexibility in the choice of training target, there may be some benefit to broadening the concept of best-responding to a fixed target. Wright et al.

17. To see this, recall that FP is defined by best response of each player to the others' distribution of play over prior iterations. If we view the strategy introduced on each PSRO iteration as a play, the uniform distribution over these is exactly what FP would respond to.

[2019] suggest a *history-aware* approach, where the BR to the primary training target is adjusted to improve performance with respect to previous targets. Perez-Nieves et al. [2021] propose maximizing a weighted combination of response quality and contribution of diversity to the current empirical game. Yao et al. [2023] likewise define a diversity metric and apply it as a regularizer during the best-response computation.

### 7.3 Frontiers in Strategy Exploration

PSRO is actively being exercised and extended by several research groups [Bighashdel et al., 2024]. Developments include computational enhancements, for example to parallelize best-response training [McAleer et al., 2020] or adaptively optimize hyperparameters [Li et al., 2024]. New MSS ideas are generated on a regular basis, and as yet there is no definitive understanding of which MSS is the best to employ for a given game environment. In general, we cannot even be sure that exploration with a given MSS will produce progress from iteration to iteration (recall the example of Table 6). An exception is to best-respond to the profile in the empirical game  $\hat{\Gamma}_{\downarrow X}$  that minimizes full-game regret, what Jordan et al. [2010a] termed the *minimum regret constrained profile* (MRCP):

$$\text{MRCP}(\Gamma, X) = \arg \min_{\sigma \in \Delta(X)} \epsilon^\Gamma(\sigma).$$

McAleer et al. [2022] proposed using MRCP as an MSS for two-player zero-sum games, observing that this would ensure an *anytime* property of monotone improvement. Wang et al. [2022] likewise considered MRCP for general-sum games, addressing its computational challenges and evaluating it as an MSS for strategy exploration. Despite the anytime property, these authors found that MRCP may not perform as well as alternatives. An explicit regularization approach proposed by Wang and Wellman [2023b], which balances equilibrium and MRCP (i.e., trades off regret in the empirical game for reduced full-game regret), seems to offer robust performance compared to other MSSs.

In the absence of broad theoretical characterization of MSS effectiveness, the literature relies on computational experimentation over a variety of game environments. Conducting these experiments presents subtle issues. As Wang et al. [2022] point out, the object of evaluation is the series of restricted strategy sets produced in the exploration process. These authors propose a consistency condition, which mandates that in comparing trajectories of strategy sets generated by alternative MSSs, the same solver should be employed. For example, to compare DO and FP (implemented by Nash and uniform MSSs, respectively), one should fix a particular solver. When feasible to compute, MRCP can be an appropriate solver for evaluation. If instead one uses Nash for DO and uniform for FP (as in some prior literature), there is a confound between the role of the solver in identifying solutions and in guiding exploration.

Typical evaluation of strategy exploration focuses on how quickly we can construct an empirical game that contains an approximate equilibrium of the full game. For games with multiple equilibria, we generally care about *which* equilibria are found, and may wish to cover a diverse set of equilibria. Wang and Wellman [2024] show that varying the response objective (i.e., beyond maximization of own utility) can effectively direct exploration toward solutions with desired qualities.

Other recent ideas provide potential new directions for strategy exploration. Smith et al. [2023] investigate opportunities for *strategic knowledge transfer*, whereby products of response learning

can be reused or repurposed in subsequent related response computation. One of their proposed methods, *mixed-opponents*, uses the value functions underlying mixed strategies produced by an MSS to construct pure-strategy response targets representing qualitatively different but plausibly relevant behavior. Li et al. [2023b] propose enhancing response policy generation with AlphaZero-style tree search [Silver et al., 2018], producing policies with runtime performance beyond the policy networks represented in the empirical game. Dyna-PSRO [Smith and Wellman, 2024] co-learns a world model (environment transition dynamics and rewards) along with the empirical game, to gain the most leverage from experience accrued in simulation during game estimation and best-response computation.

## 8. Applications

The foregoing sections have cited numerous works that advanced the methodology of EGTA, many of which were driven by demands of particular applications. In this section we focus on the application areas, selectively outlining a few in which EGTA has contributed domain insights.

**Recreational Games** Like for AI in general, game-playing has been a driving application for EGTA advances [Tuyls et al., 2020]. Early studies applied EGTA to games played among heuristic strategies for Texas Hold'em poker [Ponsen et al., 2008, 2009], and Leduc poker has served as a common benchmark for EGTA with RL [Lanctot et al., 2017]. Recent breakthroughs in game-playing have featured empirical game-theoretic reasoning; for example, the development of AlphaStar [Vinyals et al., 2019] included a “Nash league” that tracked equilibria of candidate policies generated over many iterations.<sup>18</sup> Empirical game versions of a wide variety of recreational games were also employed by Czarnecki et al. [2020] to understand common strategic landscapes.

**Economics and Finance** The earliest EGTA developments (see Section 2.1) were motivated by games involving bidding in auctions, and other economic applications. Some agent-based finance studies not explicitly labeled EGTA essentially took this approach in estimating game models from simulation data [Zhan and Friedman, 2007]. Wellman [2020] surveys economic applications of EGTA, including several examples where such methods produced new insights for canonical auction games beyond analytic tractability. Systematic EGTA studies established the centrality of price prediction in bidding heuristics for complex auctions [Wellman et al., 2008, 2017], for example. Other economic domains addressed by EGTA include financial and environmental regulation [Cheng and Wellman, 2017, Cheng et al., 2019], blockchain mechanisms [Wu et al., 2024], simulated economies [Dwarakanath et al., 2024], bank interest-rate risk [Zhao et al., 2023], formation of credit networks [Dandekar et al., 2015], adoption and use of payment mechanisms [Cheng et al., 2016, Mayo et al., 2021], fraud detection [Mayo et al., 2024], and management of debt in financial networks [Mayo and Wellman, 2021, Zhou et al., 2024].

EGTA has been applied perhaps most extensively to strategic scenarios regarding trading in financial markets [Wah and Wellman, 2016, Wah et al., 2017]. Early on, empirical game modeling shed light on heuristic strategies from the agent-based finance literature for trading in continuous double

18. <https://www.deepmind.com/blog/alphastar-mastering-the-real-time-strategy-game-starcraft-ii>

auctions [Kaisers et al., 2008, Schwartzman and Wellman, 2009]. Financial market applications have also considered specialized domains like prediction markets [Wah et al., 2016] and markets for exchange-traded funds [Shearer et al., 2021], and issues like order priority rules [Qi and Ventre, 2022] and strategic choice between market mechanisms [Wah et al., 2015]. EGTA studies have investigated the implications of market manipulation [Liu et al., 2022, Wang et al., 2021], including the prospect for automated learning of manipulative strategies [Shearer et al., 2023].

**Other Applications** Wellman et al. [2019] surveyed applications of empirical game-theoretic techniques to problems in cyber-security. For example, Hutchins et al. [2024] employed EGTA to rank network defense strategies. Other domains subjected to EGTA treatment include pursuit-evasion games [Li et al., 2023a], social dilemmas [Leibo et al., 2017, Phelps, 2016, Pretorius et al., 2020, Willis and Luck, 2023], software development [Gavidia-Calderon et al., 2020], space debris removal [Klima et al., 2016], and team formation [Yang et al., 2021].

## 9. Empirical Mechanism Design

The problem of *mechanism design* is to specify rules of interaction (i.e., the *mechanism*) for a set of agents, based on design objectives. The mechanism together with agent preferences define a game, and so in a sense the mechanism designer is specifying a game. Evaluating a mechanism essentially requires solving the game, so that the properties of the solution can be quantified. With such an evaluator on hand, a designer can—in principle—search for mechanisms that optimize their desiderata. The use of algorithms to design economic mechanisms has been termed *automated mechanism design* (AMD) [Conitzer and Sandholm, 2003]. AMD formulates the design as an optimization problem, with specified objectives and constraints that enforce equilibrium behavior.

In many real-world applications, such as developing tax or climate policies, the game induced by a mechanism is not available in analytic form. If we can simulate agent strategies interacting through a mechanism, then reasoning about the game induced by a particular design is a form of EGTA. Thus, the techniques described here can be used to characterize the strategic behavior induced by a mechanism. In such cases, we refer to the mechanism design problem as *empirical mechanism design* (EMD).<sup>19</sup>

In the first study framed explicitly as EMD, Vorobeychik et al. [2006] used EGTA to evaluate candidate designs to fix a pathology observed in the inaugural TAC supply chain management tournament. Specifically, they investigated whether changing a storage cost parameter would be sufficient to deter excessive procurement of supplies. Their approach was simply to construct empirical game models for a discrete set of parameter settings over a specified interval. They found that while increasing storage costs did indeed decrease procurement levels, no settings in the range considered reasonable were sufficient to remove the pathology.

Jordan et al. [2010b] took a similar approach to EMD in their study of the TAC Ad Auctions (TAC/AA) game. Taking the perspective of the search publisher, they examined the effect of auc-

19. Phelps et al. [2010b] define a related approach termed *evolutionary mechanism design*, which likewise employs simulation over heuristic strategy sets but emphasizes evolutionary search methods and stability concepts. The recent *differentiable economics* approach of Dütting et al. [2024] leverages machine learning methods to solve AMD problems where mechanism operations and equilibrium constraints can be expressed in neural networks.

tion parameters, such as reserve price, on publisher revenue. For each candidate setting, they solved an empirical game, essentially re-equilibrating the play among the top agents from the TAC/AA tournament. In another auction-related application, Brinkman and Wellman [2017] used EMD to determine optimal clearing intervals for call markets.

The preceding studies evaluated a fixed set of candidate mechanisms. Vorobeychik et al. [2012] proposed an approach based on black-box optimization, where candidates are generated by a stochastic search process. This work also incorporated constraints on mechanism properties (e.g., individual rationality). Areyan Viqueira et al. [2019] likewise employed a black-box technique, specifically Bayesian optimization, to search for revenue-maximizing reserve prices in a simultaneous auction scenario based on the TAC Ad Exchange game [Tao et al., 2015]. Each candidate vector of reserve prices (one per auction) defines an empirical game, a model of which they constructed using some of the sampling methods described in Section 6, assuming plausible bidding heuristics. In another distinct approach, Zhang et al. [2023] show how to reformulate mechanism design problems as two-player zero-sum games, amenable to solution using PSRO.

## 10. Conclusion

More than twenty years of research in empirical game-theoretic analysis has produced a large body of concepts, representations, algorithms, and application experience. The enterprise started with the motivating idea that empirical methods such as simulation, machine learning, and statistics could broaden the practical scope of game-theoretic reasoning, beyond what is feasible through deductive analytic techniques alone. The literature surveyed here demonstrates the extent of this broadening.

The core idea of EGTA is to induce a game model from simulation data. Distinguishing the object of deductive reasoning—the *empirical game*—from the source of knowledge about the game (e.g., an agent-based simulation model) has the practical effect of decoupling descriptive complexity from game-theoretic reasoning complexity. We can simulate a complicated world to produce an empirical game that is as simple or complex as we can computationally afford. Any simplification invariably sacrifices fidelity, but the EGTA perspective affords a smoother trade-off than we can typically achieve with analytic modeling.

In reviewing the ideas and techniques introduced over the course of EGTA’s development as a methodology, we aim to provide the reader with an understanding of the state-of-art, as well as a structure for extending the EGTA toolbox. Many of the components presented here correspond to well-defined subproblems, for example what we have labeled *game model learning* (Section 3.2.3), *strategy exploration* (Section 7), or *dynamic sampling* (Section 6.5). Progress on subproblems may be easier to evaluate than entire game reasoning frameworks. Other EGTA advances may operate at the interfaces, or could have cross-cutting effects on multiple subproblems.

For example, most EGTA techniques developed to date assume an empirical game model in normal form. Some works have started to exploit structure in empirical game models, for example, symmetry or interaction sparsity, to support representational scalability [Li and Wellman, 2020]. Capturing extensive-form (tree) structure in an empirical game model [Konicki et al., 2022, 2025, McAleer et al., 2021] can also provide advantages in representation and reasoning, including the ability to consider refined solution concepts based on this structure. In principle, EGTA could be

applied with respect to any special game class or representation, adopting any solution concept deemed appropriate. For example, recent work has shown how to conduct EGTA for *mean-field games* [Muller et al., 2022, Wang and Wellman, 2023a] and team games [McAleer et al., 2023]. In all these works, we see that extending EGTA beyond normal-form—and moreover taking full advantage of the special game features—entails further innovations in game model induction, strategy exploration, or other elements of the EGTA process.

Another path of future work should further develop tools for understanding the applicability of EGTA results. Solving an empirical game gives us a solution for that literal game model, which is a limited representation of the game we actually care about. Statistical techniques like those presented in Section 6 can inform probabilistic statements relating empirical-game solutions to the game without sampling error (i.e., what we have called the *true* game). The true game, however, is typically defined over a restricted strategy space compared to the *base* game. We typically lack strong theoretical connections between restricted-game and base-game solutions (except in the asymptotic limit), and so our level of confidence generally relies on experimental evidence. Moreover, even the base game may be a simplified version of the *underlying* game of interest, limited by the fidelity of simulation modeling. In many contexts, our interest in strategic analysis is not actually for any particular game instance, but rather in a class of strategic situations of which the underlying game is representative. Ultimately, we seek a more precise understanding of how insights from EGTA results (or any form of game-theoretic reasoning) may bear on general strategic situations beyond specific games analyzed.

The development of EGTA methodology has coincided with a significant increase in the adoption of game-theoretic principles throughout AI and computer science, as well as major advances in machine learning methods. These ML advances have contributed to improved EGTA, as well as to game-theoretic reasoning approaches that do not necessarily employ empirical game models. We expect that the use of empirical methods for analyzing games will continue to evolve rapidly, further expanding the scope of principled strategic reasoning.

## Acknowledgments

Karl Tuyls conducted the work for this survey while at Google Deepmind, Paris, France. We thank students and colleagues at the University of Michigan, Google DeepMind, and Brown University for comments and discussions on the content of this survey. We are particularly grateful to Daniel Hennes and Bhaskar Mishra for assistance with computational experiments reported herein.

## References

- Abdullah Al-Dujaili, Erik Hemberg, and Una-May O’Reilly. Approximating Nash equilibria for black-box games: A Bayesian optimization approach. Technical report, MIT CSAIL, 2018. arXiv 1804.10586v2.
- Enrique Areyan Viqueira, Cyrus Cousins, Yasser Mohammed, and Amy Greenwald. Empirical mechanism design: Designing mechanisms from data. In *35th Conference on Uncertainty in Artificial Intelligence*, 2019.

- Enrique Areyan Viqueira, Cyrus Cousins, and Amy Greenwald. Improved algorithms for learning equilibria in simulation-based games. In *19th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 79–87, 2020.
- Enrique Areyan Viqueira, Cyrus Cousins, and Amy Greenwald. Learning competitive equilibria in noisy combinatorial markets. In *20th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1446–1448, 2021.
- Olivier Armantier, Jean-Pierre Florens, and Jean-Francois Richard. Empirical game-theoretic models: Constrained equilibrium and simulations. Technical report, State University of New York at Stonybrook, 2000.
- Charles Audet and Warren Hare. *Derivative-Free and Blackbox Optimization*. Springer, 2017.
- Tim Baarslag, Katsuhide Fujita, Enrico H. Gerding, Koen Hindriks, Takayuki Ito, Nicholas R. Jennings, Catholijn Jonker, Sarit Kraus, Raz Lin, Valentin Robu, and Colin R. Williams. Evaluating practical negotiating agents: Results and analysis of the 2011 international competition. *Artificial Intelligence*, 198:73–103, 2013.
- David Balduzzi, Karl Tuyls, Julien Pérolat, and Thore Graepel. Re-evaluating evaluation. In *32nd Annual Conference on Neural Information Processing Systems*, pages 3272–3283, 2018.
- David Balduzzi, Marta Garnelo, Yoram Bachrach, Wojciech M Czarnecki, Julien Perolat, Max Jaderberg, and Thore Graepel. Open-ended learning in symmetric zero-sum games. In *36th International Conference on Machine Learning*, pages 434–443, 2019.
- Gleb Beliakov, Tomasa Calvo, and Simon James. Some results on Lipschitz quasi-arithmetic means. In *International Fuzzy Systems Association World Congress and European Society for Fuzzy Logic and Technology Conference*, pages 1370–1375, 2009.
- George Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.
- Martin Bichler, Maximilian Fichtl, Stefan Heidekrüger, Nils Kohring, and Paul Sutterer. Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence*, 3:687–695, 2021.
- Ariyan Bighashdel, Yongzhao Wang, Stephen McAleer, Rahul Savani, and Frans A. Oliehoek. Policy space response oracles: A survey. In *33rd International Joint Conference on Artificial Intelligence*, pages 7951–7961, 2024.
- Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research*, 53:659–697, 2015.
- Tilman Börgers and Rajiv Sarin. Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1–14, 1997.
- Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

- Erik Brinkman. *Understanding Financial Market Behavior through Empirical Game-Theoretic Analysis*. PhD thesis, University of Michigan, 2018.
- Erik Brinkman and Michael P. Wellman. Empirical mechanism design for optimizing clearing interval in frequent call markets. In *18th ACM Conference on Economics and Computation*, pages 205–221, 2017.
- George W. Brown. Iterative solution of games by fictitious play. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, 1951.
- Neil Burch, Martin Schmid, Matej Moravčík, Dustin Morill, and Michael Bowling. AIVAT: A new variance reduction technique for agent evaluation in imperfect information games. In *32nd AAAI Conference on Artificial Intelligence*, 2018.
- Ben-Alexander Cassell and Michael P. Wellman. EGTAOnline: An experiment manager for simulation-based game studies. In Francesca Giardini and Frédéric Amblard, editors, *Multi-Agent-Based Simulation XIII*, volume 7838 of *Lecture Notes in Artificial Intelligence*. Springer, 2013.
- Frank Cheng and Michael P. Wellman. Accounting for strategic response in an agent-based model of financial regulation. In *18th ACM Conference on Economics and Computation*, pages 187–203, 2017.
- Frank Cheng, Junming Liu, Kareem Amin, and Michael P. Wellman. Strategic payment routing in financial credit networks. In *17th ACM Conference on Economics and Computation*, pages 721–738, 2016.
- Frank Cheng, Yagil Engel, and Michael P. Wellman. Cap-and-trade emissions regulation: A strategic analysis. In *28th International Joint Conference on Artificial Intelligence*, pages 187–193, 2019.
- John Collins, Wolfgang Ketter, and Norman Sadeh. Pushing the limits of rational agents: The trading agent competition for supply chain management. *AI Magazine*, 31(2):63–80, 2010.
- Vincent Conitzer and Tuomas Sandholm. Applications of automated mechanism design. In *UAI-03 Bayesian Modeling Applications Workshop*, Acapulco, 2003.
- Cyrus Cousins and Matteo Riondato. Sharp uniform convergence bounds through empirical centralization. In *Advances in Neural Information Processing Systems*, volume 33, pages 15123–15132, 2020.
- Cyrus Cousins, Bhaskar Mishra, Enrique Areyan Viqueira, and Amy Greenwald. Computational and data requirements for learning generic properties of simulation-based games. *arXiv preprint arXiv:2208.06400*, 2022.
- Cyrus Cousins, Bhaskar Mishra, Enrique Areyan Viqueira, and Amy Greenwald. Learning properties in simulation-based games. In *22nd International Conference on Autonomous Agents and Multi-Agent Systems*, pages 272–280, 2023.

- Wojciech M. Czarnecki, Gauthier Gidel, Brendan D. Tracey, Karl Tuyls, Shayegan Omidshafiei, David Balduzzi, and Max Jaderberg. Real world games look like spinning tops. In *34th Annual Conference on Neural Information Processing Systems*, 2020.
- Pranav Dandekar, Ashish Goel, Michael P. Wellman, and Bryce Wiedenbeck. Strategic formation of credit networks. *ACM Transactions on Internet Technology*, 15(1):3:1–41, 2015.
- Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.
- A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, 1997.
- Le Cong Dinh, Stephen Marcus McAleer, Zheng Tian, Nicolas Perez-Nieves, Oliver Slumbers, David Henry Mguni, Jun Wang, Haitham Bou Ammar, and Yaodong Yang. Online double oracle. *Transactions on Machine Learning Research*, October 2022.
- Quang Duong, Yevgeniy Vorobeychik, Satinder Singh, and Michael P. Wellman. Learning graphical game models. In *21st International Joint Conference on Artificial Intelligence*, pages 116–121, 2009.
- Paul Dütting, Zhe Feng, Harikrishna Narasimham, David C. Parkes, and Sai Srivasta Ravindranath. Optimal auctions through deep learning: Advances in differentiable economics. *Journal of the ACM*, 71(1):5:1–5:53, 2024.
- Kshama Dwarakanath, Svitlana Vyetrenko, and Tucker Balch. Empirical equilibria in agent-based economic systems with learning agents. *arXiv preprint arXiv:2408.12038*, 2024.
- John Fearnley, Martin Gairing, Paul Goldberg, and Rahul Savani. Learning equilibria of games via payoff queries. In *14th ACM Conference on Electronic Commerce*, 2013.
- John Fearnley, Martin Gairing, Paul Goldberg, and Rahul Savani. Learning equilibria of games via payoff queries. *Journal of Machine Learning Research*, 16:1305–1344, 2015.
- Sevan G. Ficici, David C. Parkes, and Avi Pfeffer. Learning and solving many-player games through a cluster-based representation. In *24th Conference on Uncertainty in Artificial Intelligence*, pages 187–195, 2008.
- Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*. MIT Press, 1998.
- Xi Alice Gao and Avi Pfeffer. Learning game representations from data using rationality constraints. In *26th Conference on Uncertainty in Artificial Intelligence*, pages 185–192, 2010.
- Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.
- Madelyn Gatchel and Bryce Wiedenbeck. Learning parameterized families of games. In *22nd International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1044–1052, 2023.

- Nicola Gatti and Marcello Restelli. Equilibrium approximation in simulation-based extensive-form games. In *10th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 199–206, 2011.
- Carlos Gavidia-Calderon, Federica Sarro, Mark Harman, and Earl T. Barr. Game-theoretic analysis of development practices: Challenges and opportunities. *Journal of Systems and Software*, 159, 2020.
- Herbert Gintis. *Game Theory Evolving*. Princeton University Press, 2nd edition, 2009.
- Fred Glover. Tabu search: Part I. *ORSA Journal on Computing*, 1(3):190–206, 1989.
- Paul W. Goldberg, Christos H. Papadimitriou, and Rahul Savani. The complexity of the homotopy method, equilibrium selection, and Lemke-Howson solutions. *ACM Transactions on Economics and Computation*, 1(2):9, 2013.
- Amy Greenwald, Jiacui Li, and Eric Sodomka. Approximating equilibria in sequential auctions with incomplete information and multi-unit demand. In *Advances in Neural Information Processing Systems*, 2012.
- Amy R. Greenwald and Jeffrey O. Kephart. Shopbots and pricebots. In *16th International Joint Conference on Artificial Intelligence*, pages 506–511, 1999.
- Amy R. Greenwald, Jeffrey O. Kephart, and Gerald J. Tesauro. Strategic pricebot dynamics. In *1st ACM Conference on Electronic Commerce*, pages 58–67, 1999.
- John C. Harsanyi and Reinhard Selten. *A General Theory of Equilibrium Selection in Games*. MIT Press, 1988.
- J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.
- Jean Honorio and Luis Ortiz. Learning the structure and parameters of large-population graphical games from behavioral data. *Journal of Machine Learning Research*, 16:1157–1210, 2015.
- Charles Hutchins, Leonardo Aniello, Enrico Gerding, and Basel Halak. MANET-Rank: A framework for defence protocols against packet dropping attacks in MANETs. In *IEEE Network Operations and Management Symposium*, 2024.
- Steven Jecmen, Arunesh Sinha, Zun Li, and Long Tran-Thanh. Bounding regret in empirical games. In *34th AAAI Conference on Artificial Intelligence*, pages 4280–4287, 2020.
- Patrick R. Jordan. *Practical Strategic Reasoning with Applications in Market Games*. PhD thesis, University of Michigan, 2010.
- Patrick R. Jordan and Michael P. Wellman. Generalization risk minimization in empirical game models. In *8th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 553–560, 2009.

- Patrick R. Jordan, Christopher Kiekintveld, and Michael P. Wellman. Empirical game-theoretic analysis of the TAC supply chain game. In *6th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 1188–1195, 2007.
- Patrick R. Jordan, Yevgeniy Vorobeychik, and Michael P. Wellman. Searching for approximate equilibria in empirical games. In *7th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 1063–1070, 2008.
- Patrick R. Jordan, L. Julian Schwartzman, and Michael P. Wellman. Strategy exploration in empirical games. In *9th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1131–1138, 2010a.
- Patrick R. Jordan, Michael P. Wellman, and Guha Balakrishnan. Strategy and mechanism lessons from the first ad auctions trading agent competition. In *11th ACM Conference on Electronic Commerce*, pages 287–296, 2010b.
- Michael Kaisers, Karl Tuyls, Frank Thuijsman, and Simon Parsons. Auction analysis by normal form game approximation. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 447–450, 2008.
- Michael Kearns. Graphical games. In Nisan et al. [2007], pages 159–180.
- Jeffrey O. Kephart, James E. Hanson, and Jakka Sairamesh. Price and niche wars in a free-market economy of software agents. *Artificial Life*, 4:1–23, 1998.
- Richard Klima, Daan Bloembergen, Rahul Savani, Karl Tuyls, Daniel Hennes, and Dario Izzo. Space debris removal: A game theoretic analysis. *Games*, 7:20:1–20:18, 2016.
- Christine Konicki, Mithun Chakraborty, and Michael P. Wellman. Exploiting extensive-form structure in empirical game-theoretic analysis. In *18th Conference on Web and Internet Economics*, pages 132–149, 2022.
- Christine Konicki, Mithun Chakraborty, and Michael P. Wellman. Policy abstraction and Nash refinement in tree-exploiting PSRO. In *24th International Conference on Autonomous Agents and Multi-Agent Systems*, 2025.
- Marc Lanctot, Vinicius Zambaldi, Audrūnas Gruslys, Angeliki Lazaridou, Karl Tuyls, Julien Pérolat, David Silver, and Thore Graepel. A unified game-theoretic approach to multiagent reinforcement learning. In *31st Annual Conference on Neural Information Processing Systems*, pages 4190–4203, 2017.
- Stephen S. Lavenberg and Peter D. Welch. A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Management Science*, 27(3):322–335, 1981.
- Pierre L’Ecuyer. Efficiency improvement and variance reduction. In *Twenty-Sixth Winter Simulation Conference*, pages 122–132, 1994.
- Joel Z. Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. In *16th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 464–473, 2017.

- David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2006.
- Kevin Leyton-Brown and Yoav Shoham. *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*. Morgan & Claypool, 2008.
- Pengdeng Li, Shuxin Li, Chang Yang, Xinrun Wang, Xiao Huang, Hau Chan, and Bo An. Self-adaptive PSRO: Towards an automatic population-based game solver. In *33rd International Joint Conference on Artificial Intelligence*, pages 139–147, 2024.
- Shuxin Li, Xinrun Wang, Youzhi Zhang, Wanqi Xue, Jakub Černý, and Bo An. Solving large-scale pursuit-evasion games using pre-trained strategies. In *37th AAAI Conference on Artificial Intelligence*, pages 11586–11594, 2023a.
- Zun Li and Michael P. Wellman. Structure learning for approximate solution of many-player games. In *34th AAAI Conference on Artificial Intelligence*, pages 2119–2127, 2020.
- Zun Li and Michael P. Wellman. Evolution strategies for approximate solution of Bayesian games. In *35th AAAI Conference on Artificial Intelligence*, pages 5531–5540, 2021.
- Zun Li and Michael P. Wellman. A meta-game evaluation framework for deep multiagent reinforcement learning. In *33rd International Joint Conference on Artificial Intelligence*, pages 148–156, 2024.
- Zun Li, Marc Lanctot, Kevin R. McKee, Luke Marris, Ian Gemp, Daniel Hennes, Paul Muller, Kate Larson, Yoram Bachrach, and Michael P. Wellman. Combining tree-search, generative models, and Nash bargaining concepts in game-theoretic reinforcement learning. *arXiv preprint arXiv:2302.00797*, 2023b.
- Buhong Liu, Maria Polukarov, Carmine Ventre, Lingbo Li, Leslie Kanthan, Fan Wu, and Michail Basios. The spoofing resistance of frequent call markets. In *21st International Conference on Autonomous Agents and Multi-Agent Systems*, pages 825–832, 2022.
- Alberto Marchesi, Francesco Trovó, and Nicola Gatti. Learning probably approximately correct maximin strategies in simulation-based games with infinite strategy spaces. In *19th International Conference on Autonomous Agents and Multi-Agent Systems*, 2020.
- Luke Marris, Paul Muller, Marc Lanctot, Karl Tuyls, and Thore Graepel. Multi-agent training beyond zero-sum with correlated equilibrium meta-solvers. In *38th International Conference on Machine Learning*, pages 7480–7491, 2021.
- J. Maynard Smith and G. R. Price. The logic of animal conflicts. *Nature*, 246:15–18, 1973.
- Katherine Mayo and Michael P. Wellman. A strategic analysis of portfolio compression. In *2nd International Conference on Artificial Intelligence in Finance*, 2021.
- Katherine Mayo, Shaily Fozdar, and Michael P. Wellman. An agent-based model of strategic adoption of real-time payments. In *2nd International Conference on Artificial Intelligence in Finance*, 2021.

- Katherine Mayo, Nicholas Grabill, and Michael P. Wellman. Fraud risk mitigation in real-time payments: A strategic agent-based analysis. In *33rd International Joint Conference on Artificial Intelligence*, pages 157–165, 2024.
- Stephen McAleer, John B. Lanier, Roy Fox, and Pierre Baldi. Pipeline PSRO: A scalable approach for finding approximate Nash equilibria in large games. In *34th Annual Conference on Neural Information Processing Systems*, 2020.
- Stephen McAleer, John Lanier, Kevin A. Wang, Pierre Baldi, and Roy Fox. XDO: A double oracle algorithm for extensive-form games. In *35th Annual Conference on Neural Information Processing Systems*, 2021.
- Stephen McAleer, Kevin A. Wang, John Lanier, Marc Lanctot, Pierre Baldi, Tuomas Sandholm, and Roy Fox. Anytime PSRO for two-player zero-sum games. In *AAAI-22 Workshop on Reinforcement Learning in Games*, 2022.
- Stephen McAleer, Gabriele Farina, Gaoyue Zhou, Mingzhi Wang, Yaodong Yang, and Tuomas Sandholm. Team-PSRO for learning approximate TMECor in large team games via cooperative reinforcement learning. In *37th Annual Conference on Neural Information Processing Systems*, 2023.
- Catherine C. McGeoch. *A Guide to Experimental Algorithmics*. Cambridge University Press, 2012.
- H. Brendan McMahan, Geoffrey J. Gordon, and Avrim Blum. Planning in the presence of cost functions controlled by an adversary. In *20th International Conference on Machine Learning*, pages 536–543, 2003.
- John H. Miller and Scott E. Page. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press, 2007.
- Paul Muller, Shayegan Omidshafiei, Mark Rowland, Karl Tuyls, Julien Pérolat, Siqi Liu, Daniel Hennes, Luke Marris, Marc Lanctot, Edward Hughes, Zhe Wang, Guy Lever, Nicolas Heess, Thore Graepel, and Rémi Munos. A generalized training approach for multiagent learning. In *8th International Conference on Learning Representations*, 2020.
- Paul Muller, Mark Rowland, Romuald Elie, Georgios Piliouras, Julien Perolat, Mathieu Lauriere, Raphael Marinier, Olivier Pietquin, and Karl Tuyls. Learning equilibria in mean-field games: Introducing mean-field PSRO. In *21st International Conference on Autonomous Agents and Multi-Agent Systems*, pages 926–934, 2022.
- Kumpati S. Narendra and Mandayam A. L. Thathachar. *Learning Automata: An Introduction*. Prentice Hall, 1989.
- John F. Nash. Equilibrium points in  $n$ -person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.
- Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007.

- Shayegan Omidshafiei, Christos Papadimitriou, Georgios Piliouras, Karl Tuyls, Mark Rowland, Jean-Baptiste Lespiau, Wojciech M Czarnecki, Marc Lanctot, Julien Perolat, and Rémi Munos.  $\alpha$ -Rank: Multi-agent evaluation by evolution. *Scientific Reports*, 9, 2019.
- Shayegan Omidshafiei, Karl Tuyls, Wojciech M. Czarnecki, Francisco C. Santos, Mark Rowland, Jerome Connor, Daniel Hennes, Paul Muller, Bart De Vylder, Audrunas Gruslys, and Remi Munos. Navigating the landscape of multiplayer games. *Nature Communications*, 11(5603), 2020.
- Nicolas Perez-Nieves, Yaodong Yang, Oliver Slumbers, David Henry Mguni, Ying Wen, and Jun Wang. Modelling behavioural diversity for learning in open-ended games. In *38th International Conference on Machine Learning*, pages 8514–8524, 2021.
- S. Phelps, M. Marcinkiewicz, S. Parsons, and P. McBurney. A novel method for automatic strategy acquisition in  $n$ -player non-zero-sum games. In *5th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 705–712, 2006.
- Steve Phelps. An empirical game-theoretic analysis of the dynamics of cooperation in small groups. *Journal of Artificial Societies and Social Simulation*, 19(2), 2016.
- Steve Phelps, Simon Parsons, and Peter McBurney. An evolutionary game-theoretic comparison of two double-auction market designs. In *Agent-Mediated Electronic Commerce VI: Theories for and Engineering of Distributed Mechanisms and Systems*, volume 3435 of *Lecture Notes in Computer Science*, pages 101–114. Springer, 2005.
- Steve Phelps, Peter McBurney, and Simon Parsons. A novel method for automatic strategy acquisition and its application to a double-auction market game. *IEEE Transactions on Systems, Man, and Cybernetics: Part B*, 40:668–674, 2010a.
- Steve Phelps, Peter McBurney, and Simon Parsons. Evolutionary mechanism design: A review. *Autonomous Agents and Multi-Agent Systems*, 21:237–264, 2010b.
- Victor Picheny, Mickael Binois, and Abderrahmane Habbal. A Bayesian optimization approach to find Nash equilibria. *Journal of Global Optimization*, 73:171–192, 2019.
- Marc Ponsen, Jan Ramon, Tom Croonenborghs, Kurt Driessens, and Karl Tuyls. Bayes-relational learning of opponent models from incomplete information in no-limit poker. In *23rd AAAI Conference on Artificial Intelligence*, pages 1485–1486, 2008.
- Marc Ponsen, Karl Tuyls, Michael Kaisers, and Jan Ramon. An evolutionary game-theoretic analysis of poker strategies. *Entertainment Computing*, 1(1):39–45, 2009.
- Arnu Pretorius, Scott Cameron, Elan van Biljon, Thomas Makkink, Shahil Mawjee, Jeremy du Plessis, Jonathan Shock, Alexandre Laterre, and Karim Beguir. A game-theoretic analysis of networked system control for common-pool resource management using multi-agent reinforcement learning. In *34th Annual Conference on Neural Information Processing Systems*, 2020.
- Ji Qi and Carmine Ventre. Accounting for strategic response in limit order book dynamics. In *24th International Conference on Principles and Practice of Multi-Agent Systems*, pages 630–639, Valencia, 2022.

- Sheldon M. Ross. *Simulation*. Academic Press, third edition, 2002.
- Mark Rowland, Shayegan Omidshafiei, Karl Tuyls, Julien Perolat, Michal Valko, Georgios Piliouras, and Remi Munos. Multiagent evaluation under incomplete information. In *33rd Annual Conference on Neural Information Processing Systems*, 2019.
- Gerard Salton and Michael McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, NY, 1983.
- Yuzuru Sato and James P. Crutchfield. Coupled replicator equations for the dynamics of learning in multiagent systems. *Physical Review E*, 67(1), jan 2003.
- Peter Schuster and Karl Sigmund. Replicator dynamics. *Journal of Theoretical Biology*, 100(3): 533 – 538, 1983.
- L. Julian Schvartzman and Michael P. Wellman. Stronger CDA strategies through empirical game-theoretic analysis and reinforcement learning. In *8th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 249–256, 2009.
- L. Julian Schvartzman and Michael P. Wellman. Learning improved entertainment trading strategies for the TAC travel game. In Esther David, Enrico Gerding, David Sarne, and Onn Shehory, editors, *Agent-Mediated Electronic Commerce: Designing Trading Strategies and Mechanisms for Electronic Markets*, volume 59 of *Lecture Notes in Business Information Processing*. Springer-Verlag, 2010.
- Zhengdao Shao, Liansheng Zhuang, Houqiang Li, and Shafei Wang. COPSRO: An offline empirical game theoretic method with conservative critic. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.
- Dennis Shasha and Manda Wilson. *Statistics is Easy!* Morgan and Claypool, second edition, 2011.
- Megan J. Shearer, David Byrd, Tucker Balch, and Michael P. Wellman. Stability effects of arbitrage in exchange-traded funds. In *2nd International Conference on Artificial Intelligence in Finance*, 2021.
- Megan J. Shearer, Gabriel Rauterberg, and Michael P. Wellman. Learning to manipulate a financial benchmark. In *4th International Conference on Artificial Intelligence in Finance*, pages 592–600, 2023.
- 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.
- Max Smith, Thomas Anthony, and Michael P. Wellman. Strategic knowledge transfer. *Journal of Machine Learning Research*, 24(233):1–96, 2023.
- Max Olan Smith and Michael P. Wellman. Co-learning empirical games and world models. *Reinforcement Learning Journal*, 1:1–15, 2024.

- Eric Sodomka, John Collins, and Maria L. Gini. Efficient statistical methods for evaluating trading agent performance. In *22nd National Conference on Artificial Intelligence*, pages 770–775, 2007.
- Samuel Sokota, Caleb Ho, and Bryce Wiedenbeck. Learning deviation payoffs in simulation-based games. In *33rd AAAI Conference on Artificial Intelligence*, pages 2173–2180, 2019.
- Ashish Sureka and Peter R. Wurman. Using tabu best-response search to find pure strategy Nash equilibria in normal form games. In *4th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 1023–1029, 2005.
- Bingyang Tao, Fan Wu, and Guihai Chen. TAC AdX’14: Autonomous agents for realtime ad exchange. In *14th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1111–1119, 2015.
- Sebastian Shenghong Tay, Quoc Phong Nguyen, Chuan Sheng Foo, and Bryan Kian Hsiang Low. No-regret sample-efficient Bayesian optimization for finding Nash equilibria with unknown utilities. In *26th International Conference on Artificial Intelligence and Statistics*, 2023.
- P. Taylor and L. Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156, 1978.
- Gerald Tesauro and Rajarshi Das. High-performance bidding agents for the continuous double auction. In *3rd ACM Conference on Electronic Commerce*, pages 206–209, 2001.
- Leigh Tesfatsion. Agent-based computational economics: A constructive approach to economic theory. In Leigh Tesfatsion and Kenneth L. Judd, editors, *Handbook of Computational Economics*. Elsevier, 2006.
- Karl Tuyls and Simon Parsons. What evolutionary game theory tells us about multiagent learning. *Artificial Intelligence*, 171(7):406–416, 2007.
- Karl Tuyls, Katja Verbeeck, Tom Lenaerts, Sam Maes, and Bernard Manderick. Towards a relation between learning agents and evolutionary dynamics. In *14th Belgium-Netherlands Conference on Artificial Intelligence BNAIC 2002*, pages 315–322, 2002.
- Karl Tuyls, Katja Verbeeck, and Tom Lenaerts. A selection-mutation model for Q-learning in multi-agent systems. In *2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 693–700, 2003.
- Karl Tuyls, Julien Pérolat, Marc Lanctot, Joel Z. Leibo, and Thore Graepel. A generalised method for empirical game theoretic analysis. In *17th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 77–85, 2018a.
- Karl Tuyls, Julien Perolat, Marc Lanctot, Rahul Savani, Joel Leibo, Toby Ord, Thore Graepel, and Shane Legg. Symmetric decomposition of asymmetric games. *Scientific Reports*, 8(1):1015, 2018b.
- Karl Tuyls, Julien Pérolat, Marc Lanctot, Edward Hughes, Richard Everett, Joel Z. Leibo, Csaba Szepesvári, and Thore Graepel. Bounds and dynamics for empirical game theoretic analysis. *Autonomous Agents and Multi-Agent Systems*, 34:7, 2020.

- Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander S. Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom L. Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 2019.
- Yevgeniy Vorobeychik. Probabilistic analysis of simulation-based games. *ACM Transactions on Modeling and Computer Simulation*, 20(3):16:1–25, 2010.
- Yevgeniy Vorobeychik and Michael P. Wellman. Mechanism design based on beliefs about responsive play (position paper). In *ACM EC-06 Workshop on Alternative Solution Concepts for Mechanism Design*, Ann Arbor, MI, 2006.
- Yevgeniy Vorobeychik, Christopher Kiekintveld, and Michael P. Wellman. Empirical mechanism design: Methods, with application to a supply chain scenario. In *7th ACM Conference on Electronic Commerce*, pages 306–315, 2006.
- Yevgeniy Vorobeychik, Michael P. Wellman, and Satinder Singh. Learning payoff functions in infinite games. *Machine Learning*, 67:145–168, 2007.
- Yevgeniy Vorobeychik, Daniel M. Reeves, and Michael P. Wellman. Constrained automated mechanism design for infinite games of incomplete information. *Autonomous Agents and Multi-Agent Systems*, 25:313–351, 2012.
- Zbynek Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.
- Elaine Wah and Michael P. Wellman. Latency arbitrage in fragmented markets: A strategic agent-based analysis. *Algorithmic Finance*, 5:69–93, 2016.
- Elaine Wah, Dylan R. Hurd, and Michael P. Wellman. Strategic market choice: Frequent call markets vs. continuous double auctions for fast and slow traders. In *Third EAI Conference on Auctions, Market Mechanisms, and Their Applications*, 2015.
- Elaine Wah, Sébastien Lahaie, and David M. Pennock. An empirical game-theoretic analysis of price discovery in prediction markets. In *25th International Joint Conference on Artificial Intelligence*, pages 510–516, 2016.
- Elaine Wah, Mason Wright, and Michael P. Wellman. Welfare effects of market making in continuous double auctions. *Journal of Artificial Intelligence Research*, 59:613–650, 2017.
- William E. Walsh. *Market Protocols for Decentralized Supply Chain Formation*. PhD thesis, University of Michigan, 2001.
- William E. Walsh, Michael P. Wellman, and Fredrik Ygge. Combinatorial auctions for supply chain formation. In *2nd ACM Conference on Electronic Commerce*, pages 260–269, 2000.

- William E. Walsh, Rajarshi Das, Gerald Tesauro, and Jeffrey O. Kephart. Analyzing complex strategic interactions in multi-agent systems. In *AAAI-02 Workshop on Game-Theoretic and Decision-Theoretic Agents*, 2002.
- William E. Walsh, David C. Parkes, and Rajarshi Das. Choosing samples to compute heuristic-strategy Nash equilibrium. In *AAMAS-03 Workshop on Agent-Mediated Electronic Commerce*, 2003.
- Xintong Wang, Chris Hoang, Yevgeniy Vorobeychik, and Michael P. Wellman. Spoofing the limit order book: A strategic agent-based analysis. *Games*, 12(2), 2021.
- Yongzhao Wang and Michael P. Wellman. Empirical game-theoretic analysis for mean-field games. In *22nd International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1025–1033, 2023a.
- Yongzhao Wang and Michael P. Wellman. Regularization for strategy exploration in empirical game-theoretic analysis. *arXiv preprint arXiv:2302.04928*, 2023b.
- Yongzhao Wang and Michael P. Wellman. Generalized response objectives for strategy exploration in empirical game-theoretic analysis. In *23rd International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1892–1900, 2024.
- Yongzhao Wang, Qiurui Ma, and Michael P. Wellman. Evaluating strategy exploration in empirical game-theoretic analysis. In *21st International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1346–1354, 2022.
- Yufei Wang, Zheyuan Ryan Shi, Lantao Yu, Yi Wu, Rohit Singh, Lucas Joppa, and Fei Fang. Deep reinforcement learning for green security games with real-time information. In *33rd AAAI Conference on Artificial Intelligence*, pages 1401–1408, 2019.
- Christopher J. C. H. Watkins and Peter Dayan. Technical note: Q-learning. *Machine Learning*, 8: 279–292, 1992.
- Kevin Waugh, Brian D. Ziebart, and J. Andrew Bagnell. Computational rationalization: The inverse equilibrium problem. In *28th International Conference on Machine Learning*, pages 1169–1176, 2011.
- Jorgen Weibull. *Evolutionary Game Theory*. MIT Press, 1997.
- Michael P. Wellman. Methods for empirical game-theoretic analysis (extended abstract). In *21st National Conference on Artificial Intelligence*, pages 1552–1556, 2006.
- Michael P. Wellman. Putting the agent in agent-based modeling. *Autonomous Agents and Multi-Agent Systems*, 30:1175–1189, 2016.
- Michael P. Wellman. Economic reasoning from simulation-based game models. *Æconomia*, 10: 257–278, 2020.
- Michael P. Wellman and Katherine Mayo. Navigating in a space of game views. *Autonomous Agents and Multi-Agent Systems*, 38(31):1–25, 2024.

- Michael P. Wellman, Jeffrey K. MacKie-Mason, Daniel M. Reeves, and Sowmya Swaminathan. Exploring bidding strategies for market-based scheduling. In *4th ACM Conference on Electronic Commerce*, pages 115–124, 2003.
- Michael P. Wellman, Joshua Estelle, Satinder Singh, Yevgeniy Vorobeychik, Christopher Kiekintveld, and Vishal Soni. Strategic interactions in a supply chain game. *Computational Intelligence*, 21:1–26, 2005a.
- Michael P. Wellman, Daniel M. Reeves, Kevin M. Lochner, Shih-Fen Cheng, and Rahul Suri. Approximate strategic reasoning through hierarchical reduction of large symmetric games. In *20th National Conference on Artificial Intelligence*, pages 502–508, 2005b.
- Michael P. Wellman, Amy Greenwald, and Peter Stone. *Autonomous Bidding Agents: Strategies and Lessons from the Trading Agent Competition*. MIT Press, 2007.
- Michael P. Wellman, Anna Osepayshvili, Jeffrey K. MacKie-Mason, and Daniel M. Reeves. Bidding strategies for simultaneous ascending auctions. *B. E. Journal of Theoretical Economics (Topics)*, 8(1), 2008.
- Michael P. Wellman, Tae Hyung Kim, and Quang Duong. Analyzing incentives for protocol compliance in complex domains: A case study of introduction-based routing. In *12th Workshop on the Economics of Information Security*, 2013.
- Michael P. Wellman, Eric Sodomka, and Amy Greenwald. Self-confirming price-prediction strategies for simultaneous one-shot auctions. *Games and Economic Behavior*, 201:339–372, 2017.
- Michael P. Wellman, Mason Wright, and Thanh H. Nguyen. Empirical game-theoretic analysis for adaptive cyber-defense. In Sushil Jajodia, George Cybenko, Peng Liu, Cliff Wang, and Michael P. Wellman, editors, *Adversarial and Uncertain Reasoning for Adaptive Cyber Defense*, volume 11830 of *Lecture Notes in Computer Science*, pages 112–128. Springer, 2019.
- John R. Wicks and Amy Greenwald. An algorithm for computing stochastically stable distributions with applications to multiagent learning in repeated games. In *21st Conference on Uncertainty in Artificial Intelligence*, pages 623–632, 2005.
- Bryce Wiedenbeck and Erik Brinkman. Data structures for deviation payoffs. In *22nd International Conference on Autonomous Agents and Multi-Agent Systems*, pages 670–678, 2023.
- Bryce Wiedenbeck and Michael P. Wellman. Scaling simulation-based game analysis through deviation-preserving reduction. In *11th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 931–938, 2012.
- Bryce Wiedenbeck, Ben-Alexander Cassell, and Michael P. Wellman. Bootstrap statistics for empirical games. In *13th International Conference on Autonomous Agents and Multi-Agent Systems*, pages 597–604, 2014.
- Bryce Wiedenbeck, Fengjun Yang, and Michael P. Wellman. A regression approach for modeling games with many symmetric players. In *32nd AAAI Conference on Artificial Intelligence*, pages 1266–1273, 2018.

- Richard Willis and Michael Luck. Resolving social dilemmas through reward transfer commitments. In *AAMAS-23 Workshop on Adaptive Learning Agents*, 2023.
- James R. Wright and Kevin Leyton-Brown. Predicting human behavior in unrepeated, simultaneous-move games. *Games and Economic Behavior*, 106:16–37, 2017.
- Mason Wright, Yongzhao Wang, and Michael P. Wellman. Iterated deep reinforcement learning in games: History-aware training for improved stability. In *20th ACM Conference on Economics and Computation*, pages 617–636, 2019.
- Fei Wu, Thomas Thiery, Stefanos Leonardos, and Carmine Ventre. To compete or collude: Bidding incentives in ethereum block building auctions. In *5th International Conference on Artificial Intelligence in Finance*, pages 813–821, 2024.
- Fengjun Yang, Negar Mehr, and Mac Schwager. Decentralized role assignment in multi-agent teams via empirical game-theoretic analysis. *arXiv preprint arXiv:2109.14755*, 2021.
- Jian Yao, Weiming Liu, Haobo Fu, Yaodong Yang, Stephen McAleer, Qiang Fu, and Wei Yang. Policy space diversity for non-transitive games. In *37th Annual Conference on Neural Information Processing Systems*, 2023.
- H. Peyton Young. The evolution of conventions. *Econometrica*, 61(1):57–84, 1993.
- E. C. Zeeman. Population dynamics from game theory. In Z. Nitecki and C. Robinson, editors, *Global Theory of Dynamical Systems*, volume 819 of *Lecture Notes in Mathematics*, pages 471–497. Springer, 1980.
- E. C. Zeeman. Dynamics of the evolution of animal conflicts. *Theoretical Biology*, 89:249–270, 1981.
- Wenjie Zhan and Daniel Friedman. Markups in double auction markets. *Journal of Economic Dynamics and Control*, 31:2984–3005, 2007.
- Brian Hu Zhang, Gabriele Farina, Ioannis Anagnostides, Federico Cacciamani, Stephen McAleer, Andreas Haupt, Andrea Celli, Nicola Gatti, Vincent Conitzer, and Tuomas Sandholm. Computing optimal equilibria and mechanisms via learning in zero-sum extensive-form games. In *37th Annual Conference on Neural Information Processing Systems*, 2023.
- Lingxiao Zhao, Maria Polukarov, and Carmine Ventre. Liquidity and solvency risks in financial networks. In *4th International Conference on Artificial Intelligence in Finance*, pages 210–218, 2023.
- Hao Zhou, Yongzhao Wang, Konstantinos Varsos, Niholas Bishop, Rahul Savani, Anisoara Calinescu, and Michael Wooldridge. Selfishly prepaying in financial credit networks. *Journal of Artificial Intelligence Research*, 81:877–906, 2024.