

## STRATEGIC ANALYSIS WITH SIMULATION-BASED GAMES

Yevgeniy Vorobeychik

Computer and Information Science  
University of Pennsylvania  
Philadelphia, PA 19104, USA  
yev@seas.upenn.edu

Michael P. Wellman

Computer Science and Engineering  
University of Michigan  
Ann Arbor, MI 48109-2121, USA  
wellman@umich.edu

### ABSTRACT

We present an overview of an emerging methodology for applying game-theoretic analysis to strategic environments described by a simulator. We first introduce the problem of solving a simulation-based game, and proceed to review convergence results and confidence bounds about game-theoretic equilibrium estimates. Next, we present techniques for approximating equilibria in simulation-based games, and close with a series of applications.

### 1 INTRODUCTION

In analyzing economic systems, especially those composed primarily of autonomous computational agents, the complexities must typically be distilled into a stylized model, with the hope that the resulting model captures the key components of the system and is nevertheless amenable to analytics. Increasingly, the boundaries of analytic tractability are being pushed, particularly as the systems, as well as the participating agents, are being more precisely engineered, and as the computational barriers that had once rendered complex mechanisms impractical are now coming down en masse. Consider, for example, combinatorial auctions, which had in the past been shunned because of the complexity of the winner determination problem, but have now become ubiquitous in academic literature, as well as practice ([Cramton et al. 2006](#)). The advantage of the increasingly precise models of Economic micro-systems, such as complex (e.g., combinatorial) auctions, is the ability to model these quite closely in a stylized way. The disadvantage, of course, is the added complexity in analyzing high-resolution models of strategic interactions.

One approach to modeling rich strategic environments is to describe them by a simulator, and generate a game model through controlled sampling over the space of strategic options. For example, [Sureka and Wurman \(2005\)](#) applied this approach to study bidding strategy for combinatorial auctions. Over the past decade, many specific methods have been introduced to analyze game-theoretic models specified using simulations, applied to a variety of multiagent domains ([Greenwald and Kephart 2002](#), [Reeves 2005](#), [Vorobeychik 2008](#), [Vorobeychik et al. 2006](#), [Vorobeychik and Porche 2009](#), [Walsh et al. 2002](#), [Wellman et al. 2005](#), [Wellman et al. 2008](#)). Several of these approaches estimate game models based on a finite number of simulation samples for a collection of joint strategic choices of the players. Game-theoretic analysis of the estimated game form (e.g., identifying Nash equilibria or other solutions) is then extrapolated to draw strategic conclusions about the actual game.

In this paper, we present an overview of the emerging area of analyzing simulation-based games. We first introduce the problem of solving a simulation-based game, and proceed to review convergence results and confidence bounds about game-theoretic equilibrium estimates. Next, we present techniques for approximating equilibria in simulation-based games, and close with a series of application studies.

### 2 GAME THEORETIC PRELIMINARIES

In this section we formalize the notion of strategic interactions (or games) between a set of rational agents. We denote a normal-form game by  $\Gamma_N = [I, R, u(\cdot)]$ , where  $I$  is the (finite) set of players with  $m = |I|$  the number of players (here  $|I|$  denotes the cardinality of a finite set  $I$ ),  $R$  the joint strategy set (pure or mixed, depending on context) with  $R_i$  the set of

strategies of player  $i$ , and  $u(\cdot)$  the function that maps joint strategy profiles  $r \in R$  to a vector of payoff entries for all players, that is,  $u(r) = \{u_1(r), \dots, u_m(r)\}$ , where  $u_i(r)$  denotes the (deterministic) payoff function of player  $i$ . We assume that all payoffs  $u_i(r)$  are finite.

We let  $A_i$  denote the set of player  $i$ 's pure strategies, with  $A = A_1 \times \dots \times A_m$  the joint pure strategy set. One may think of pure strategies as atomic actions of players in the game, such as a choice of a specific bid in an auction or a decision to confess in prisoners' dilemma. Alternatively, a pure strategy can be a complex policy, such as bidding as a function of history of past observed bids and private information, although it remains "atomic" from the perspective of normal-form analysis: a policy can be associated with, say, an index, and selecting a pure strategy would mean selecting the corresponding index. We denote by  $S_i$  the set of  $i$ 's mixed strategies and use  $S$  to denote the joint mixed strategy set. Mixed strategies are probability distributions over pure (atomic) strategies. A key assumption is that each player selects an action to play according to  $s$  *independently* of selections made by all the other players. (For example, players cannot communicate at the time when they randomly draw their actions. An alternative concept of *correlated* strategies has also been studied (Osborne and Rubinstein 1994), although we do not deal with it here.) If  $s \in S$  is a mixed strategy profile, then  $s(a)$  denotes the probability that a pure strategy profile  $a \in A$  is played under  $s$ . Analogously,  $s_i(a_i)$  is the probability that player  $i$ 's mixed strategy selects action  $a_i$  and  $s_{-i}(a_{-i})$  is the probability that  $a_{-i} \in A_i$  is selected under the joint mixed strategy of players other than  $i$ .

Often, we view a strategy (pure or mixed) in a simulation-based game as a black box—for example, a computer program that implements it—and furthermore restrict the space of all strategies to some subset which we can analyze computationally. As such, we use the more abstract notation  $r$  to denote a pure or mixed (and possibly restricted—hence the mnemonic) strategy profile, with  $R$  denoting the set of all (restricted) strategy profiles.  $r_i$  then refers to the strategy of player  $i$  and we use notation  $r_{-i}$  for a profile of strategies other than that of player  $i$ .

An important strategic element of a normal-form game is *game-theoretic regret* (or simply *regret*) of a profile  $r \in R$ , denoted by  $\varepsilon(r)$ , which is the most any player can gain by deviating from  $r_i$  to any strategy in  $R_i$ . Formally,

$$\varepsilon(r) = \max_{i \in I} \max_{r'_i \in R_i} u_i(r'_i, r_{-i}) - u_i(r). \quad (1)$$

In specifying a normal-form model of a strategic setting, the analyst has in mind predicting what players who are faced with decisions in such a strategic context will do. The outcomes of strategic interactions—that is, the ultimate decisions made by the players—are commonly assumed for the purposes of prediction to be *rational* (alternatively, *strategically stable*) in the sense that every player is playing optimally *given* the choices of other players. This notion of strategic stability is formalized as a *Nash equilibrium* solution concept for games.

**Definition 1** A Nash equilibrium of the normal-form game  $\Gamma_N$  is a profile  $r \in R$  such that for every player  $i \in I$ ,

$$u_i(r) \geq u_i(r'_i, r_{-i}) \quad \forall r'_i \in R_i.$$

We can alternatively define the Nash equilibrium concept using the game-theoretic regret:  $r$  is a Nash equilibrium if and only if  $\varepsilon(r) = 0$ .

In an approximation context—for example, when a game is too large to compute a Nash equilibrium exactly, or when only payoff estimates are available—a common concept is that of an  $\varepsilon$ -Nash equilibrium.

**Definition 2** An  $\varepsilon$ -Nash equilibrium of the normal-form game  $\Gamma_N$  is a profile  $r \in R$  such that for every player  $i \in I$ ,

$$u_i(r) + \varepsilon \geq u_i(r'_i, r_{-i}) \quad \forall r'_i \in R_i.$$

We may again relate  $\varepsilon$ -Nash (or approximate Nash) equilibria to the regret function by noting that any profile  $r \in R$  is an  $\varepsilon(r)$ -Nash equilibrium. Henceforth, whenever we talk about solutions or equilibria in games, we mean Nash equilibria or corresponding approximations.

### 3 SIMULATION-BASED GAMES

In this paper we are interested in analyzing game-theoretic models in which player payoff functions are specified using simulations; we refer to these models as *simulation-based games*. A simulation-based game retains all the basic elements of the normal-form game, but makes the notion of utility functions  $u(\cdot)$  somewhat more precise in a way pertinent to analysis. Specifically, a utility function in a simulation-based game is represented by an *oracle*,  $\mathcal{O}$ , which can be queried with any strategy profile  $r \in R$  to produce a possibly noisy sample payoff vector  $U$ . In simulation-based games, we presume to

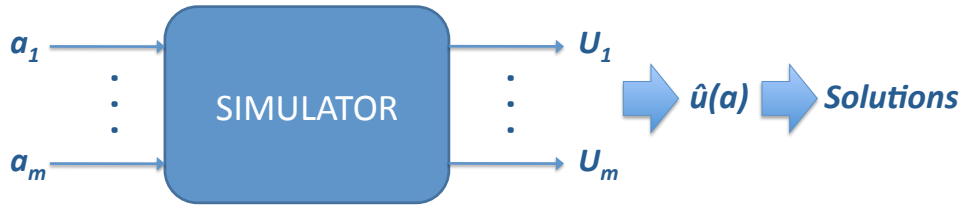


Figure 1: A schematic of simulation-based game analysis.

have a simulation model of the player payoffs, and we must run this simulation to obtain a noisy sample of player payoffs for any fixed joint strategy profile. Figure 1 presents a schematic view of simulation-based game analysis. For example, we may imagine an agent-based combat simulation in which the analyst can set strategic parameters of the adversaries and run the simulation to obtain a sample outcome of a battle or a campaign. As another example, consider a one-item first-price sealed-bid auction game with two players, represented by a simulation. Suppose that each of two players has some valuation  $(v_1, v_2)$  respectively) for the item that is being auctioned off, but the valuations is randomly chosen according to some distribution (each player does get to observe his own valuation, but not that of his opponent). Suppose that the bidders submit bids which are a fixed fraction of their value, that is,  $b_i = \alpha_i v_i$ . If  $b_i > b_j$ , bidder  $i$  gets the good and pays his bid, for a utility of  $(1 - \alpha_i)v_i$ . Otherwise, he does not receive the good and pays nothing (utility of zero). For a fixed strategy profile  $\{\alpha_1, \alpha_2\}$ , a simulator samples values  $v_1, v_2$ , computes the corresponding bids, and then determines the allocation (who receives the good) and payments, returning the corresponding payoff samples to both players.

Formally, a simulation-based game is denoted by  $\Gamma_S = [I, R, \mathcal{O}]$ , where the oracle (simulator)  $\mathcal{O}$  produces a sample vector of payoffs  $U = (U_1, \dots, U_m)$  to all players for a (typically pure) strategy profile  $r$ . We assume throughout that  $E[U] = u(r)$ . We denote an estimate of a payoff to player  $i$  for profile  $r$  based on  $n(r)$  i.i.d. samples (that is, the number of samples  $n$  can be different for different strategy profiles  $r$ ) from  $\mathcal{O}$  by

$$\hat{u}_{i,n}(r) = \frac{1}{n(r)} \sum_{j=1}^{n(r)} U_i(r)^j,$$

where each  $U(r)^j$  is generated by invoking the oracle with profile  $r$  as input. The vector of payoff estimates for all players we denote by  $\hat{u}_n(r)$ , where  $n$  is the number of payoff samples taken for the profile  $r$  (or just  $\hat{u}(r)$  where we would like to talk about any estimate of the payoff at  $r$ ).

Using the estimates  $\hat{u}(r)$ , we can construct an estimated payoff matrix for a finite game from simulation data. We often call this estimated game an *empirical game* to reflect the empirical source of its payoff estimates. Since an *empirical* or estimated game can be viewed as a normal-form game in its own right, all the concepts defined on normal-form games apply. Thus, we obtain *empirical regret* as

$$\hat{\epsilon}(r) = \max_i \max_{r'_i \in R_i} \hat{u}_i(r'_i, r_{-i}) - \hat{u}_i(r),$$

and, similarly, an *empirical* ( $\epsilon$ -)Nash equilibrium is just a ( $\epsilon$ -)Nash equilibrium of the empirical game.

One key distinction between  $\mathcal{O}$  and  $u(\cdot)$  is that  $u(\cdot)$  provides direct access to exact payoff evaluations, whereas  $\mathcal{O}$  evaluates payoffs with noise. Another distinction, no less vital but somewhat more subtle, is that by specifying the payoff function as an oracle, we resign ourselves to the fact that payoffs are not available in any easily tractable form and focus our attention to controlling the simulation experiments.

Implicit to the discussion of both the simulation-based games and empirical games is that they are defined with respect to an *underlying* game  $\Gamma_{N \leftarrow S}$  characterized by the set of players  $I$ , a set of strategy profiles  $R$ , and the payoff function  $u(\cdot)$  from which the oracle, in effect, takes noisy samples. Given  $\Gamma_{N \leftarrow S}$ , the *true* regret of a profile  $r \in R$ ,  $\epsilon(r)$ , in both the simulation-based and the empirical game is evaluated with respect to  $u(\cdot)$  of this underlying game.

## 4 NASH EQUILIBRIUM ESTIMATION IN FINITE GAMES

We begin our discussion with some results and techniques applicable only (or primarily) to games that are finite (i.e., have finite sets of pure strategy profiles  $A$ ). In much of this section, we assume that samples have been taken for every pure strategy profile  $a \in A$ , in which case the game needs to actually be small enough to allow this (what “small enough” means depends on the running time of simulations; if simulations are fast, the size of strategy space can be rather large, whereas when simulations that take hours to run, even a game with relatively few strategies and players will be large).

### 4.1 Baseline (Direct) Nash Equilibrium Estimation

The most direct—although not necessarily optimal—method for estimating Nash equilibria in finite games in which we have at least some samples for every  $a \in A$  is the following:

1. Generate an *empirical game*  $\Gamma_{\mathcal{E}}$  by sampling the simulator (oracle)  $\mathcal{O}$  to obtain a data set  $D$  of profile-payoff tuples  $(r, U)$ .
2. Apply a game solver (e.g., the GAMBIT toolbox (McKelvey et al. 2005)) to  $\Gamma_{\mathcal{E}}$  to obtain Nash equilibrium estimates.

A level of sophistication can be added if we use variance reduction techniques rather than sample means to estimate payoffs of the underlying game (Reeves 2005, Wellman et al. 2005). For example, control variates, conditioning, and quasi-random sampling can achieve a considerable increase in sampling efficiency (Ross 2001, L’Écuyer 1994). In this paper, however, we assume that we use sample means as payoff estimates—as described above.

### 4.2 Convergence

An important property of game-theoretic regret function  $\varepsilon(s)$  on a mixed strategy space  $S$  in finite games is that we can extend the Law of Large Numbers to obtain uniform regret convergence. The following theorem states this result formally.

**Theorem 1 (Vorobeychik (2009))**  $\hat{\varepsilon}_n(s) \rightarrow \varepsilon(s)$  *a.s. uniformly on  $S$ .*

Several useful facts, shown in Vorobeychik (2009), follow from this result.

**Theorem 2** *For every  $\gamma > 0$ , there is  $M$  such that  $\forall n \geq M$ ,  $\mathcal{N} \subset \mathcal{N}_{n,\gamma}$  a.s., where  $\mathcal{N}_{n,\gamma} = \{s \in S : \hat{\varepsilon}_n(s) \leq \gamma\}$ .*

Informally, Theorem 2 assures us that when enough samples of all profiles have been taken, the set of *all* Nash equilibrium solutions (which could be infinite in degenerate cases, but would normally be finite) is contained in the set of all approximate equilibrium solutions (for a fixed approximation constant) of the corresponding empirical game. This result can be extremely useful if we are interested in evaluating, for example, worst-case equilibrium outcomes in a game. In such case, a conservative approach of using a worst-case approximate equilibrium outcome of a game produced by simulations is, thus, asymptotically sound.

**Theorem 3**  $h_D(\mathcal{N}_n, \mathcal{N})$  *converges to 0 a.s., where  $h_D(X, Y) = \sup_{x \in X} \inf_{y \in Y} d(x, y)$ .*

The meaning of Theorem 3 is that every Nash equilibrium in an empirical game becomes arbitrarily close to *some* equilibrium of the underlying game, as long as enough samples are taken. Note that we do not claim the converse: indeed, it is not difficult to construct an example with a set of all Nash equilibria of the underlying game which has points that are never close to any estimated Nash equilibria of an empirical game, no matter how many samples have been taken.

### 4.3 Confidence Bounds

We now turn our attention to exploring confidence bounds for the estimates of the regret function  $\hat{\varepsilon}(s)$  on the (infinite) mixed strategy space  $S$ .

**Theorem 4 (Vorobeychik (2009))** *Suppose that simulation samples  $U_i(a)$  are bounded between 0 and 1 and we have taken at least  $n$  samples for every strategy profile  $a \in A$ . Then, for  $s \in S$ ,*

$$\Pr\{|\varepsilon(s) - \hat{\varepsilon}(s)| \geq \gamma\} \leq m|A|(K+1)e^{-\gamma^2 n/2},$$

where  $K = \max_{i \in I} |A_i|$ .

Thus, for example, if  $\hat{s}$  is a Nash equilibrium of a game obtained using simulations with at least  $n$  samples taken for every pure strategy profile  $a \in A$ ,

$$\Pr\{\varepsilon(\hat{s}) \geq \gamma\} \leq m|A|(K+1)e^{-\gamma^2 n/2}.$$

Note that the confidence bound above required rather minimal assumptions on the nature of randomness in the simulation. We now make, instead, a very strong assumption, and use it to derive a tight bound.

To begin, suppose that we have an improper prior on  $u_i(a)$  for all  $a \in A$ , and the sampling noise is Gaussian with known variance  $\sigma_i^2(a)$  (i.e., variance may be different for different players and strategy profiles). The result below builds on the derivation of the distribution of the maximum of  $k$  variables based on samples of these with zero-mean additive Gaussian noise by [Chang and Huang \(2000\)](#), who demonstrate that if we start with an improper prior over the actual payoffs and observe samples distorted by Gaussian noise, the posterior distributions of  $u_i(\cdot) \mid \hat{u}_i(\cdot)$  are Gaussian random variables with mean  $\hat{u}_i(\cdot)$  and variance  $\sigma_i^2(a)$ . Furthermore, if payoffs are sampled independently, the actual (posterior) payoffs  $u_i(\cdot)$  are also independently distributed, conditional on  $\hat{u}_i(\cdot)$ . Given these assumptions, the following confidence bound can be derived for a mixed strategy profile  $s \in S$  (in the sequel we omit conditioning on  $\hat{u}_i(\cdot)$  for brevity):

**Theorem 5 (Vorobeychik (2009))** *Suppose that payoffs for every  $a \in A$  are sampled independently with zero-mean additive Gaussian noise, and suppose that we have an improper prior on  $u_i(a)$ . Let  $s \in S$  be a mixed strategy profile. Then,*

$$\Pr(\varepsilon(s) \leq \varepsilon) = \prod_{i \in I} \int_{\mathbb{R}} \prod_{b \in A_i} [\Pr(W_i(b) \leq u + \varepsilon)] f_{W_i^*}(u) du,$$

where

$$\Pr(W_i(b) \leq u + \varepsilon) = 1 - \Phi \left[ \frac{\sum_{c \in A_{-i}} \hat{u}_i(b, c) s_{-i}(c) - u - \varepsilon}{\sqrt{\sum_{c \in A_{-i}} \frac{\sigma_i^2(b, c) s_{-i}^2(c)}{n(b, c)}}} \right]$$

and

$$W_i^* \sim N \left( \sum_{a \in A} \hat{u}_i(a) s(a), \sum_{a \in A} \frac{\sigma_i^2(a) s^2(a)}{n(a)} \right).$$

#### 4.4 Exploring Large Finite Games

In our results above, we spent no time discussing a process by which profiles are selected to be sampled. Indeed, our results do not very much depend on that. However, it is rather intuitive that intelligent allocation of simulation samples (if these are somewhat expensive) can considerably improve Nash equilibrium estimates. [Walsh et al. \(2003\)](#) present an approach to selecting profiles to query from the simulator at each step of the exploration process which uses a measure, dubbed *EVI* or *expected value of information*. Their first step is to define the information state to be an empirical game,  $\Gamma_{\mathcal{E}}$ , and utility estimate to be a measure of Nash equilibrium approximation error with respect to the empirical game,  $\hat{f}_{\Gamma_{\mathcal{E}}}(\cdot)$ . *EVI* is then defined in this setting as follows:

$$EVI(r \mid \Gamma_{\mathcal{E}}) = E_{\Gamma_{\mathcal{E}.r} \mid \Gamma_{\mathcal{E}}} [\hat{f}_{\Gamma_{\mathcal{E}.r}}(x(\Gamma_{\mathcal{E}})) - \hat{f}_{\Gamma_{\mathcal{E}.r}}(x(\Gamma_{\mathcal{E}.r}))],$$

where  $\Gamma_{\mathcal{E}.r}$  designates an empirical game obtained by adding to  $\Gamma_{\mathcal{E}}$  a payoff sample (or a constant number of samples) for the profile  $r$ . Now, define the error function  $f$  to be game-theoretic regret with respect to *every* Nash equilibrium estimate in  $\Gamma_{\mathcal{E}}$ , that is, with respect to the set  $\mathcal{N}(\Gamma_{\mathcal{E}})$  of equilibrium estimates. First, recall the definition of game-theoretic regret above. Generalizing the notion of regret in empirical games, define *empirical regret* of a profile  $r$  to be

$$\varepsilon_{\Gamma_{\mathcal{E}}}(t) = \max_{i \in I} \max_{b \in D_i(t)} [u_i(b, r_{-i}) - u_i(r)]. \quad (2)$$

In this expression,  $D_i(t)$  is the set of profiles in the data set of the empirical game  $\Gamma_{\mathcal{E}}$  with only player  $i$  deviating from  $t$ . Let the set of solutions to the empirical game  $\mathcal{N}(\Gamma_{\mathcal{E}})$  to be the set of regret-minimizing profiles on the restricted strategy space  $R$ . Define the error of solutions to empirical game  $\Gamma_{\mathcal{E}}^1$  with respect to another empirical game (perhaps, one with a

better estimate of the payoff function),  $\Gamma_{\mathcal{E}}^2$ , to be

$$e(\Gamma_{\mathcal{E}}^1, \Gamma_{\mathcal{E}}^2) = \frac{1}{|\mathcal{N}(\Gamma_{\mathcal{E}}^1)|} \sum_{t \in \mathcal{N}(\Gamma_{\mathcal{E}}^1)} \epsilon_{\Gamma_{\mathcal{E}}^2}(t).$$

Using this notation, *EVI* becomes

$$EVI(r | \Gamma_{\mathcal{E}}) = E_{\Gamma_{\mathcal{E}.r} | \Gamma_{\mathcal{E}}} [e(\Gamma_{\mathcal{E}}, \Gamma_{\mathcal{E}.r}) - e(\Gamma_{\mathcal{E}.r}, \Gamma_{\mathcal{E}.r})].$$

Now, assume that every  $t \in \mathcal{N}(\Gamma_{\mathcal{E}})$  has zero game theoretic regret evaluated with respect to the game  $\Gamma_{\mathcal{E}}$ , that is,  $\epsilon_{\Gamma_{\mathcal{E}}}(t) = 0$  (this would be tantamount to assuming that there is a profile  $t$  which is an equilibrium of  $\Gamma_{\mathcal{E}}$ ). If we impose no restrictions on our strategy set  $R$ , allowing it to be the set of all mixed strategy profiles, all solutions in  $\mathcal{N}(\Gamma_{\mathcal{E}})$  are guaranteed to have zero regret and, thus, the assumption will necessarily hold. Under this assumption,  $e(\Gamma_{\mathcal{E}.r}, \Gamma_{\mathcal{E}.r}) = 0$  for any  $\Gamma_{\mathcal{E}.r}$ , and *EVI* simplifies to

$$EVI(r | \Gamma_{\mathcal{E}}) = E_{\Gamma_{\mathcal{E}.r} | \Gamma_{\mathcal{E}}} [e(\Gamma_{\mathcal{E}}, \Gamma_{\mathcal{E}.r})] = \frac{1}{|\mathcal{N}(\Gamma_{\mathcal{E}})|} \sum_{t \in \mathcal{N}(\Gamma_{\mathcal{E}})} E_{\Gamma_{\mathcal{E}.r}} [\epsilon_{\Gamma_{\mathcal{E}.r}}(t)].$$

Now, consider a profile  $r$  which is neither an equilibrium estimate (i.e.,  $r \notin \mathcal{N}(\Gamma_{\mathcal{E}})$ ) nor a neighbor to an equilibrium estimate (i.e.,  $r \notin D_i(t) \forall i \in I, \forall t \in \mathcal{N}(\Gamma_{\mathcal{E}})$ ). By sampling this profile,  $E_{\Gamma_{\mathcal{E}.r}}[\epsilon_{\Gamma_{\mathcal{E}.r}}(t)]$  will remain unchanged for every  $t \in \mathcal{N}(\Gamma_{\mathcal{E}})$ , since  $r$  has no bearing on it. Consequently, *EVI* = 0 for every such profile  $r$ . Since  $EVI \geq 0 \forall r \in R$ , we need only to consider profiles  $r$  that are either current solutions or their neighbors. Generically, the number of solutions to  $\Gamma_{\mathcal{E}}$  is very small compared to the size of the game, and the number of neighbors is linear in the size of  $\mathcal{N}(\Gamma_{\mathcal{E}})$ , in the number of players  $I$  and in the number of player strategies  $R_i$  (if finite).

Note that *EVI* is highly computationally intensive, requiring the computation of the entire set of Nash equilibria many times during the evaluation sequence for a single profile  $r$ . One alternative (cheaper) approach, described by [Jordan et al. \(2008\)](#), is to select a profile that maximizes a Kullback-Leibler divergence, which is based on the distribution of minimum regret before and after sampling a chosen profile. They begin by presuming that the sampling action will take  $k$  samples, that is, each profile  $s \in S$  is sampled  $k$  times, and then a profile is selected that yields the highest value of information gain. The core of the approach to computing information gain, based on Kullback-Leibler divergence, is very general in that it can use any prior distribution on profiles obtained based on the current empirical game,  $p_s(\Gamma_{\mathcal{E}})$ . Thus, [Jordan et al. \(2008\)](#) first develop it for an arbitrary distribution, and then specialize to use the one of particular interest to empirical game search.

First, define the *entropy* of a profile  $s$ ,  $\mathcal{H}(s; \Gamma_{\mathcal{E}})$ :

$$\mathcal{H}(s; \Gamma_{\mathcal{E}}) = -p_s(\Gamma_{\mathcal{E}}) \log_2 p_s(\Gamma_{\mathcal{E}}) - (1 - p_s(\Gamma_{\mathcal{E}})) \log_2 (1 - p_s(\Gamma_{\mathcal{E}}))$$

The standard definition of *cross entropy* of  $s$ , denoted here by

$\mathcal{H}(s; \Gamma_{\mathcal{E}}, \hat{\Gamma}_{\mathcal{E}})$ , is then

$$\mathcal{H}(s; \Gamma_{\mathcal{E}}, \hat{\Gamma}_{\mathcal{E}}) = -p_s(\Gamma_{\mathcal{E}}) \log_2 p_s(\hat{\Gamma}_{\mathcal{E}}) - (1 - p_s(\Gamma_{\mathcal{E}})) \log_2 (1 - p_s(\hat{\Gamma}_{\mathcal{E}}))$$

Based on these, define the *information gain* for a profile  $s$  from taking  $k$  additional samples of  $\hat{s}$ , denoted  $\mathcal{G}(s; \Gamma_{\mathcal{E}}, \Gamma_{\mathcal{E}.s^k})$ , to be

$$\mathcal{G}(s; \Gamma_{\mathcal{E}}, \Gamma_{\mathcal{E}.s^k}) = \mathcal{H}(s; \Gamma_{\mathcal{E}}, \Gamma_{\mathcal{E}.s^k}) - \mathcal{H}(s; \Gamma_{\mathcal{E}})$$

Finally, the aggregate information gain from sampling a profile  $s$  a total of  $k$  times, denoted  $\mathcal{G}(\Gamma_{\mathcal{E}}, \Gamma_{\mathcal{E}.s^k})$ , is

$$\mathcal{G}(\Gamma_{\mathcal{E}}, \Gamma_{\mathcal{E}.s^k}) = \sum_{\hat{s} \in D(s)} \mathcal{G}(\hat{s}; \Gamma_{\mathcal{E}}, \Gamma_{\mathcal{E}.s^k})$$

In developing their assessment of likely strategic outcomes based on the evidence encompassed by the empirical game, [Jordan et al. \(2008\)](#) posit that players are most likely to play a profile with the lowest regret. Since they restrict the search space to pure strategy profiles, such profiles need not constitute Nash equilibria, although often they will (particularly in very large games), and even more often the smallest regret will be indeed quite low to justify the belief. Thus, they define

the information gain with respect to the distribution  $p_s(\Gamma_{\mathcal{E}})$  which assigns probabilities to profiles  $s$  in proportion to their likelihood of having the smallest regret. To develop these distributions formally, let us begin with the definition of the highest payoff a player  $i$  can obtain by deviating from  $s$  to another strategic option:

**Definition 3 (Maximum deviation payoff)** For a given player  $i$  and profile  $s$ , the maximum deviation payoff is

$$\delta_i(s) = \max_{\hat{s} \in S_i \setminus s_i} u_i(\hat{s}, s_{-i}).$$

The distribution of  $\delta_i(s)$ , denoted by  $F_{\delta_i(s)}(\delta)$ , is the  $n^{\text{th}}$  order statistic (maximum) over the mean payoffs of the deviations, given by

$$F_{\delta_i(s)}(d) = \prod_{\hat{s} \in S_i \setminus s_i} F_{u_i(\hat{s}, s_{-i})}(d).$$

The distribution of player regret,  $r$ , denoted by  $F_{\mathcal{E}_i(s)}(r)$  can be obtained by conditioning on the payoff to  $i$  from playing  $s$ :

$$F_{\mathcal{E}_i(s)}(r) = \int_{\mathbb{R}} F_{\delta_i(s)}(u+r) \cdot dF_{u_i(s)}(u). \quad (3)$$

The integral in (3) is estimated using Monte Carlo with importance sampling (Ross 2001). The distribution of regret for a particular profile,  $s$ , is then simply

$$F_{\mathcal{E}(s)}(r) = \prod_{i \in I} F_{\mathcal{E}_i(s)}(r).$$

Now, as the final piece, we can define the actual distribution of minimum regret, that is, we can define, for each profile  $s \in S$ , the probability that  $s$  has minimum regret given the evidence in the empirical game:

$$p_s(\Gamma_{\mathcal{E}}) = \int_{\mathbb{R}} \left[ \prod_{\hat{s} \in S \setminus s} (1 - F_{\mathcal{E}(\hat{s})}(r)) \right] dF_{\mathcal{E}(s)}(r). \quad (4)$$

The estimation of the value of the integral in (4) using Monte Carlo requires generating  $M$  realizations of the random variable. Each of these  $M$  realizations requires in turn computing or estimating the Equation (3) expression  $|S| - 1$  times. Since the resulting algorithm can become quite computationally cumbersome, the authors suggest as an alternative approximating the integral by using point estimates for the mean regret of the remaining profiles:

$$p_s^*(\theta) = F_{\mathcal{E}(s)}(\epsilon_{S \setminus s}^{(1)}), \quad (5)$$

where  $\epsilon_{S \setminus s}^{(1)}$  is the lowest regret over all profiles except  $s$  calculated using the expected mean payoffs given the empirical game  $\Gamma_{\mathcal{E}}$ .

The problem of sample selection is substantially easier if we consider a setting in which the simulator produces samples with no noise. Jordan et al. (2008) present a heuristic, which they call, *min-regret-first search (MRFS)*, for such a setting. The *MRFS* algorithm proceeds from an arbitrary starting point to sample a profile which is a neighbor to some (arbitrary) approximate Nash equilibrium solution of the current empirical game. They find that *MRFS* is very effective in a variety of settings. The reason for this is, perhaps, that in the noise-free setting and given the regret error measure, *MRFS* is nearly equivalent to *EVI*—equivalent if we assume that solutions have zero regret with respect to the empirical game, as above. With this assumption, we already demonstrated that under *EVI* only solutions or neighbors need to be considered as sampling actions. When there is no sampling noise, there is no need to sample solution profiles again, since the values of the players' utilities at these profiles are already known exactly. Furthermore, the neighbors that have already been sampled need not be sampled again for the same reason. Therefore, the only profiles that need to be considered are the neighbors to solutions which have not been added to the data set of the empirical game, and that is precisely what *MRFS* would prescribe.

## 4.5 Hierarchical Reduction for Symmetric Games

Consider a 4-player symmetric game in which each player has 35 strategies (in a symmetric game, all players have identical pure strategy sets ( $A_i = A_j$  for all  $i, j \in I$ ) and symmetric utility functions in the sense that  $u_i(a_i, a_{-i}) = u_j(a_j, a_{-j})$  whenever  $a_i = a_j$  and  $a_{-i} = a_{-j}$  for all  $i, j \in I$ ). The total number of distinct strategy profiles in such a game is 73,815, a large, but, perhaps, manageable number. Now, if we double the number of players, the number of strategy profiles explodes to 118,030,185—over 118 million! Now, if we additionally imagine that such a game is represented using simulations, and, furthermore, that each strategy profile yields a noisy sample payoff, the 8-player variant becomes clearly intractable (even if we run a simulation every second and take 10 samples for each strategy profile, we would need in excess of 13,500 days, or roughly 37 years, to explore the entire game).

The exploration techniques in the previous section suggest that we may be able to avoid exploring the entire game and still arrive at reasonable Nash equilibrium estimates. An alternative (which may well be complementary) to the sophisticated exploration techniques is to shrink the size of the game. Such an idea was explored by Wellman et al. (2005). Specifically, they propose, given an original  $K$ -player game, constructing a smaller game with  $K/q$  players (naturally, assuming that  $K$  is divisible by  $q$ ), where each player “controls”  $c$  others (i.e., groups of  $c$  players form a meta-player with  $c$  “arms”, with the player selecting an identical strategy for each of its arms). For example, the original 8-player game would be reduced to a 4-player game, with each meta-player controlling 2 players. Clearly, such a reduced game will likely have different equilibrium solutions from the underlying game. However, as it may now be explored more completely, better solution estimates could be obtainable for the reduced game. Wellman et al. (2005) provide theoretical and experimental support for this approach in a variety of settings.

## 5 NASH EQUILIBRIUM APPROXIMATION IN INFINITE GAMES

### 5.1 Heuristic Exploration Techniques

Recall our measure of game-theoretic regret,  $\varepsilon(r)$ , of a profile  $r \in R$ . By definition,  $\varepsilon(r) \geq 0$  and  $\varepsilon(r) = 0$  iff  $r$  is a Nash equilibrium. Consequently, the problem of computing a Nash equilibrium is equivalent to global non-linear constrained optimization. When the game is finite, it means that by taking a finite sample of payoff vectors for all pure strategy profiles, we can obtain an estimate of the payoff function and, hence of the regret function, on the mixed strategy space  $S$ , which is strongly consistent (with respect to the actual regret function).

This is, of course, only for (small) finite games. In the case of infinite games, naturally we cannot expect to sample every pure strategy profile even once. Nevertheless, formulating the problem as optimization opens the door for the use of simulation optimization techniques. Specifically, suppose that we have a simulation optimization technique,  $T$ , which is globally convergent as long as estimates of the objective are convergent. Then, if we have a consistent estimator of  $\varepsilon(r)$  for any fixed point  $r$ , we can apply  $T$  to get global convergence to a Nash equilibrium (when one exists on the strategy space  $R$ ). While in our setting we only have access to samples from the payoff function for any fixed  $r$ , we can use the same technique  $T$  to obtain a consistent estimator for  $\varepsilon(r)$ , as required above. Combining these, we can obtain a provably globally convergent algorithm for estimating a Nash equilibrium in infinite games, so long as we can find an appropriate  $T$ . An example of such a globally convergent method is simulated annealing (with appropriate schedules of temperatures) (Ghate and Smith 2008). The formal statement of the Nash equilibrium convergence result is somewhat involved; the details can be found in Vorobeychik and Wellman (2008).

Now, consider the following alternative Nash equilibrium estimation scheme to the convergent algorithm just described. Say that  $T$  is still some technique (perhaps convergent) for (stochastic) simulation optimization, and suppose that we still retain  $T$  for the purpose of estimating a best response vector to a profile  $r$ ,  $br(r)$  (that is, best response choices by all players given that others maintain their strategies in  $r$ ), as well as the corresponding regret  $\varepsilon(r)$ . We can now utilize a well-known best response dynamic, which simply follows a sequence of best responses, with  $r_{k+1} = br(r_k)$ , with  $k$  denoting a current iteration. While not in general convergent, iterative best response has been shown empirically to be very effective as a *search* heuristic (i.e., after the requisite number of best response iterations, we still use profiles with the smallest estimated regret as Nash equilibrium estimates) (Vorobeychik and Wellman 2008).

### 5.2 Learning Payoff Functions from Exploration Data

No matter what technique we may follow to guide the simulations, we in the end obtain a “trail” of data,  $D = \{(a_j, U_j)_{j=1}^L\}$ , with  $L$  data points featuring strategy profiles and corresponding payoff (sample) tuples. A “direct” (or sample based) way

of using this data to produce a Nash equilibrium estimate would be by using a profile with the smallest empirical regret given data. Vorobeychik et al. (2007) proposed an alternative: treat this as a supervised learning problem for estimating players' payoff functions, and then use the learned payoff functions (and, consequently, the learned game) to obtain solution estimates. Experimental experience suggests that the learning technique can yield considerably better equilibrium candidates than direct estimation of the empirical game. The general problem of inducing a game model from simulation data can be formulated as a loss minimization problem, with cross-validation employed to select among candidate models (Jordan and Wellman 2009).

## 6 APPLICATIONS

Many of the developments in methodology for analysis of simulation-based games have been driven by applications. We present a sample, biased toward our own work, illustrating the nature of some of the games and issues that have been addressed through these methods. The first example presents a case where simulation-based game analysis enabled development and validation of a superior strategy for a common (but analytically intractable) auction mechanism. Subsequent sections describe applications that shed light on incentive properties of established economic mechanisms, and qualitative phenomena observed in supply chain scenarios.

### 6.1 Bidding Strategies for Simultaneous Ascending Auctions

In many trading situations, agents interact with several markets simultaneously, buying and selling multiple goods. Though these markets may operate independently, the agents' values or costs for these goods are often interrelated. This poses a challenging strategic problem, given the pervasive uncertainty about other agents' behavior and market conditions in general.

One canonical model of multiple interdependent markets is the *simultaneous ascending auction* (SAA) mechanism (Cramton 2006). In an SAA, agents decide in each discrete round which goods to bid on, and the price of a good is incremented whenever an agent not currently winning the good bids on it. When a round passes with no new bids, the game ends and goods are allocated to their current winners at current prices.

Despite the simplicity of this mechanism, and ubiquity of SAA variants in practice, auction theory has little to say about equilibrium strategies for simultaneous markets. Simulation-based approaches, in contrast, have fruitfully informed strategic reasoning in this domain (Wellman, Osepayshvili, MacKie-Mason, and Reeves 2008). As typical in empirical game studies, our investigation is based on a reformulation of strategy space based on a basic agent architecture. In the case of SAA bidding, we defined the class of *perceived-price* bidding strategies, in which agents select a set of goods to bid on based on a specification of beliefs about the effective prices of these goods. The method for generating perceived prices provides a convenient way to parametrize this strategy space. For example, bidding based on current prices (i.e., myopic behavior) is called *straightforward bidding*.

Simulation-based experimentation, and experience in related trading games, suggested that generating perceived prices based on predictions could address the key strategic challenges in this domain. There are many potential sources of price predictions, each representing a strategy candidate given the underlying agent architecture. One particularly effective approach is to identify *self-confirming prices*: a set of prices such that, when all agents bid based on price-prediction using these prices, these come out to be correct, on average. Generalization of this approach to probabilistic predictions yields the strategy we label  $PP(F^{SC})$ , which stands for price-prediction (PP) using the self-confirming probability distribution  $F^{SC}$ .

A comprehensive simulation-based game-theoretic analysis for one particular SAA environment (five goods and five agents) revealed that  $PP(F^{SC})$  constitutes a pure-strategy Nash equilibrium, in the empirical game comprising 53 strategies evaluated. Somewhat less extensive analyses confirmed that  $PP(F^{SC})$  supports an equilibrium—exactly or approximately—in a range of different SAA configurations. What these configurations do have in common is that agents' preferences for the goods are complementary. When goods are substitutes, different strategic issues come to the fore, and analyses of simulation-based games under this assumption led to the conclusion that simple demand-reduction strategies fare best.

The SAA game is just one instance of a trading game where simulation-based methods have led to new conclusions about best known strategies. Other examples include the well-studied domain of continuous double auctions (Schwartzman and Wellman 2009), and the Trading Agent Competition travel (Wellman, Greenwald, and Stone 2007) and supply chain management (Jordan, Kiekintveld, and Wellman 2007) games.

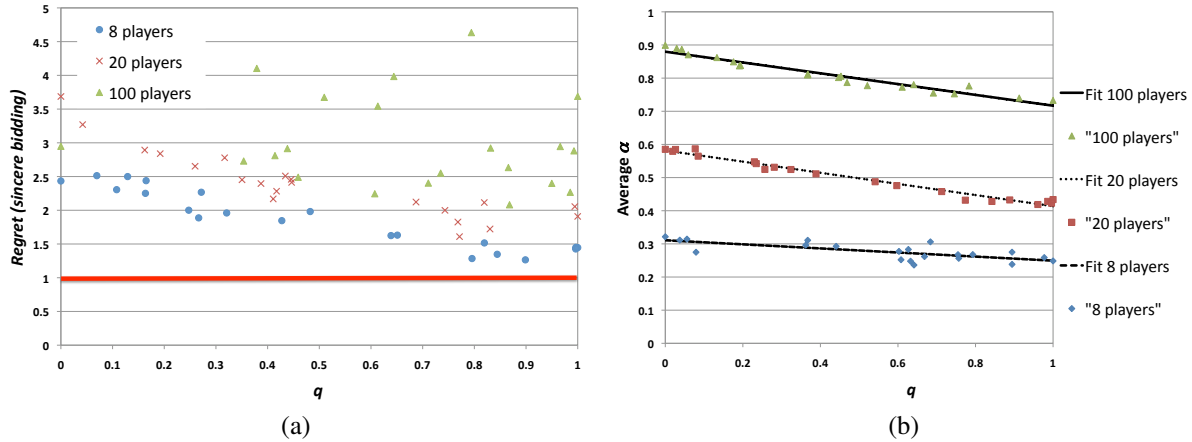


Figure 2: (a) Incentives to deviate from sincere bidding and (b) average fraction of true value bid in equilibrium, when  $v_i$  and  $e_i$  are i.i.d. uniform in  $[0, 1]$  with varying numbers of players.

## 6.2 (Un)Truthful Bidding in Keyword Auctions

Sponsored search auctions (also known as keyword, slot, or ad auctions) have become one of the primary sources of revenue for the major search engines. Since their beginnings these auctions have undergone a series of changes, from a first-price to a next-price format, from rank-by-bid to rank-by-revenue. Roughly, the consensus today stands at the rank-by-revenue schemes, which augment the ranking rule to include advertiser *quality scores* (reflecting the effect of the ad quality on the probability it is clicked) and a next-price (or *generalized second-price (GSP)*) format (under which each advertiser pays the minimum amount sufficient to remain in the currently allocated slot).

Since the *GSP* pricing rule was, as its name suggests, an attempt to generalize a second-price (Vickrey) auction, one may anticipate that incentive compatibility (i.e., bidders have an incentive to be sincere about their actual preferences), being the central property of Vickrey auctions, would carry over. In reality, it is known that *GSP* is not, in fact, incentive compatible. The problem is that it is difficult to quantify analytically exactly to what extent incentives to be insincere exist in actual auctions. These incentives, however, can be quantified using simulation-based game theoretic analysis methods. In Figure 2 (a) and (b) we show incentives to deviate from truthful bidding (i.e., submitting bids that are actual valuations) and the fraction of value that the advertisers actually submit as their (approximate) equilibrium bids. Figure 2 demonstrates that incentives to deviate exceed 100% of bidder profits, and, furthermore, equilibrium bids are well below actual willingness to pay, suggesting that *GSP* is unlikely to elicit bids that are in any sense close to sincere. As a bidder, the results suggest that consideration of these incentives should be a significant component of a bidding strategy.

## 6.3 Combinatorial Auctions with Approximate Allocation Algorithms

Whereas prevalent *GSP* mechanism in keyword auctions generalizes the single-item second-price (Vickrey) auction in some respects, it does not preserve the theoretical incentive to report truthful valuations. As we observed in the previous section, analysis of the simulation-based game confirms that the *GSP* mechanism induces rational agents to bid considerably below actual valuations. Another mechanism, called *VCG* (Vickrey-Clarke-Groves), however, does yield a socially optimal allocation (in the sense that it maximizes the sum of bidders' valuations) and is also incentive compatible (Krishna 2002). However, *VCG* requires that the optimal allocation given bids be computed *exactly*—approximation will, in general, result in the loss of incentive compatibility. As in the previous study, our goal is to quantify this loss, as displayed in Figure 3 for a particular kind of combinatorial auction (one in which valuations of bidders for items satisfy a *submodularity* property, that is, a marginal value of a fixed set of items falls as more items are added to the total bundle). In the figure we can see that, while the greedy algorithm in the submodular setting is a 2-approximation, the incentives to deviate are quite low (in fact, likely to be even lower since the greedy, rather than optimal, outcome is used as a measuring stick).

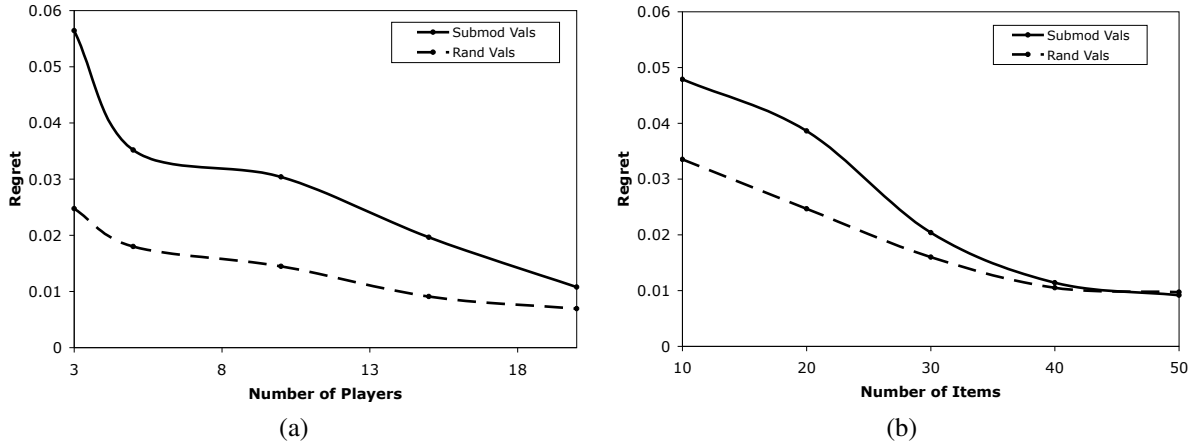


Figure 3: Average game-theoretic regret, with greedy algorithm approximating the allocation in a combinatorial auction with submodular valuations and random marginal valuations (that is, a marginal value of an item given a fixed set of other items is uniformly random over a unit interval), as a function of the number of (a) players (number of items fixed at 20), (b) items (number of players is fixed at 5).

#### 6.4 Strategic Procurement in a Supply-Chain Game

The Trading Agent Competition (TAC) is an annual venue that provides a stochastic market simulation environment in which autonomous trading agents can compete using an API provided by the designers. The first competition, held in the summer of 2000, featured a travel-shopping scenario, in which the trading agents were to act as travel agents for simulated customers (Wellman et al. 2001). Several years later (2003), the supply-chain game (TAC/SCM) was introduced (Arunachalam and Sadeh 2005). Early in the seeding rounds of TAC/SCM 2003, a particular strategic element of the game began quickly to attract considerable attention: all of the high-scoring agents were submitting very large orders for all or nearly all of the input components *on the first day of the simulation*. Once this strategy became widely recognized as yielding a considerable advantage, progressively more agents adopted it. In retrospect, this behavior, termed *aggressive day-0 procurement*, can be understood from studying the supplier pricing model: it turns out that the prices of all inputs are lowest and their availability highest on simulation day 0.

While aggressive day-0 procurement emerged as a prominent strategic aspect of the inaugural TAC/SCM tournament, it is unclear if it was entirely in players' interest to follow it: perhaps it only seemed so upon superficial observation of the games. However, simulation-based game theory provides tools to systematically address questions such as this. The first step is to define a stylized game that captures the relevant aspects of the problem. This was done by Wellman et al. (2005), who distilled the complex space of player strategies into two: *aggressive* (A) and *baseline* (B), characterized by the particular day-0 procurement behavior. Let  $iA$  denote the profile with  $i$  agents playing A (the rest playing B). The unique pure-strategy Nash equilibrium in this game based on the estimated payoff matrix is 4A. That this is an equilibrium can be seen by comparing adjacent columns in the bar chart of Figure 4. Arrows indicate for each column, whether an agent in that profile would prefer to stay with that strategy (arrow head), or switch (arrow tail). Solid black arrows denote statistically significant comparisons. Profile 4A is the only one with only in-pointing arrows.

Although jockeying for day-0 procurement turned out to be an interesting strategic issue in itself, the phenomenon detracted from other important problems, such as adapting production levels to varying demand (since component costs were already sunk), and dynamic management of production, sales, and inventory. In response to the problem, the TAC/SCM designers adopted several rule changes intended to penalize large day-0 orders. These included modifications to supplier pricing policies and introduction of storage costs assessed on inventories of components and finished goods. Despite the changes, day-0 procurement was very high in the early rounds of the 2004 competition. In a drastic measure, the Game Master imposed a fivefold increase of storage costs midway through the tournament. Even this did not stem the tide, and day-0 procurement in the final rounds actually *increased* (by some measures) from 2003 (Kiekintveld et al. 2005).

To address the mechanism design problem involved in appropriately setting a storage cost to stem the high day-0 procurement, we again turn to simulation-based game theory. First, we create a stylized game model as above by abstracting away the complexity of player strategies, this time into actual day-0 procurement decisions (restricted to a closed interval

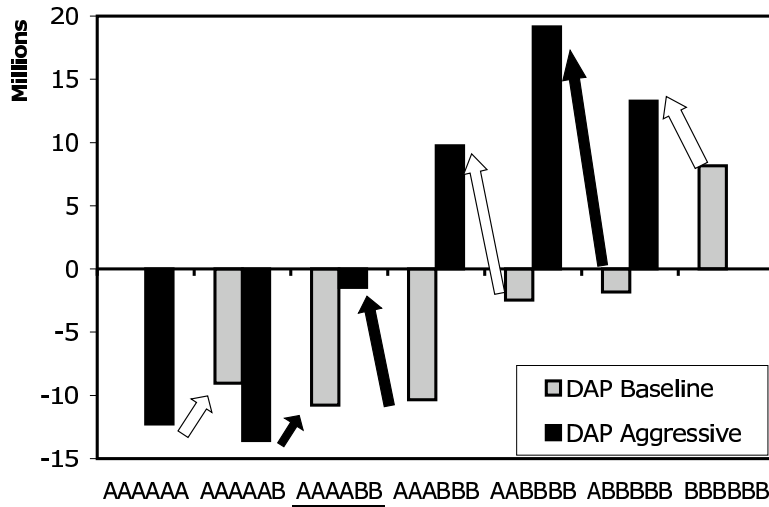


Figure 4: Estimated payoffs for the 6-player 2-strategy TAC/SCM stylized game.

beyond which the choices are unlikely to be reasonable). We proceed to estimate sets of approximate Nash equilibria by applying *MRFS* (described above) to a discretized joint strategy space of players, and taking an average of 5 to 10 samples as the true mean payoff for a strategy profile. The results, shown in Figure 5, suggest that the effect of storage cost, while

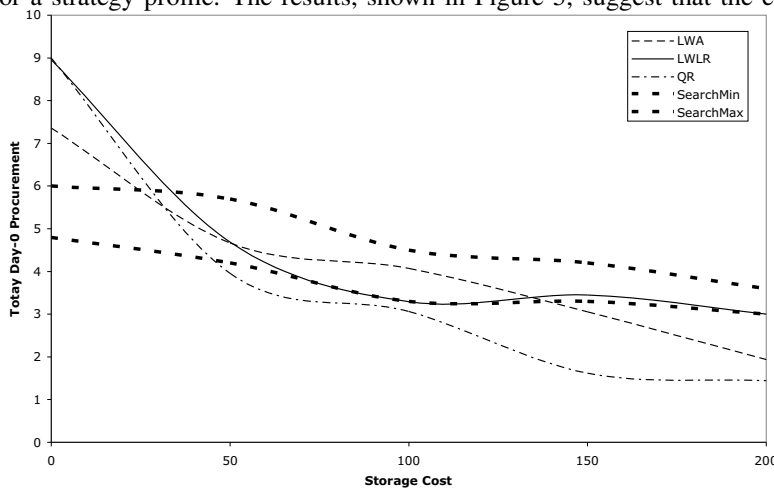


Figure 5: Aggregate day-0 procurement estimates based on search in strategy profile space compared to three machine learning techniques. The equilibrium total procurement correspondence is the interval between “SearchMin” and “SearchMax”.

noticeable in the downward sloping equilibrium correspondence of total day-0 procurement decisions by the players, is not sufficient to stem the tide, and more fundamental adjustments must be (and ultimately were) made to the game simulation.

## REFERENCES

Arunachalam, R., and N. M. Sadeh. 2005. The supply chain trading agent competition. *Electronic Commerce Research and Applications* 4:63–81.

Chang, Y.-P., and W.-T. Huang. 2000. Generalized confidence intervals for the largest value of some functions of parameters under normality. *Statistica Sinica* 10:1369–1383.

Cramton, P. 2006. Simultaneous ascending auctions. See [Cramton, Shoham, and Steinberg \(2006\)](#).

Cramton, P., Y. Shoham, and R. Steinberg. (Eds.) 2006. *Combinatorial auctions*. MIT Press.

- Ghate, A., and R. L. Smith. 2008. Adaptive search with stochastic acceptance probabilities for global optimization. *Operations Research Letters* 36 (3): 285–290.
- Greenwald, A. R., and J. O. Kephart. 2002. Shopbot economics. *Journal of Autonomous Agents and Multiagent Systems* 5 (3): 255–287.
- Jordan, P., Y. Vorobeychik, and M. P. Wellman. 2008. Searching for approximate equilibria in empirical games. In *Seventh International Conference on Autonomous Agents and Multiagent Systems*, 1063–1070.
- Jordan, P., and M. P. Wellman. 2009. Generalization risk minimization in empirical game models. In *Eighth International Conference on Autonomous Agents and Multiagent Systems*.
- Jordan, P. R., C. Kiekintveld, and M. P. Wellman. 2007. Empirical game-theoretic analysis of the TAC supply chain game. In *Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 1188–1195. Honolulu.
- Kiekintveld, C., Y. Vorobeychik, and M. P. Wellman. 2005. An analysis of the 2004 supply chain management trading agent competition. In *IJCAI-05 Workshop on Trading Agent Design and Analysis*.
- Krishna, V. 2002. *Auction theory*. Academic Press.
- L'Écuyer, P. 1994. Efficiency improvement and variance reduction. In *Winter Simulation Conference*, 122–132.
- McKelvey, R. D., A. M. McLennan, and T. L. Turocy. 2005. Gambit: Software tools for game theory, version 0.2005.06.13.
- Osborne, M. J., and A. Rubinstein. 1994. *A course in game theory*. MIT Press.
- Reeves, D. M. 2005. *Generating trading agent strategies: Analytic and empirical methods for infinite and large games*. Ph. D. thesis, University of Michigan.
- Ross, S. M. 2001. *Simulation*. 3rd ed. Academic Press.
- Schwartzman, L. J., and M. P. Wellman. 2009. Stronger CDA strategies through empirical game-theoretic analysis and reinforcement learning. In *Eighth International Conference on Autonomous Agents and Multiagent Systems*. Budapest.
- Sureka, A., and P. R. Wurman. 2005. Applying metaheuristic techniques to search the space of bidding strategies in combinatorial auctions. In *Genetic and Evolutionary Computation Conference*, 2097–2103. Washington, DC.
- Vorobeychik, Y. 2008. *Mechanism design and analysis using simulation-based game models*. Ph. D. thesis, University of Michigan.
- Vorobeychik, Y. 2009. Probabilistic analysis of simulation-based games. Working paper.
- Vorobeychik, Y., C. Kiekintveld, and M. P. Wellman. 2006. Empirical mechanism design: Methods, with an application to a supply chain scenario. In *Seventh ACM Conference on Electronic Commerce*, 306–315.
- Vorobeychik, Y., and I. Porche. 2009. Game theoretic methods for analysis of combat simulations. Working paper.
- Vorobeychik, Y., and M. P. Wellman. 2008. Stochastic search methods for Nash equilibrium approximation in simulation-based games. In *Seventh International Conference on Autonomous Agents and Multiagent Systems*, 1055–1062.
- Vorobeychik, Y., M. P. Wellman, and S. Singh. 2007. Learning payoff functions in infinite games. *Machine Learning* 67 (2): 145–168.
- Walsh, W. E., R. Das, G. Tesauro, and J. O. Kephart. 2002. Analyzing complex strategic interactions in multi-agent systems. In *AAAI-02 Workshop on Game Theoretic and Decision Theoretic Agents*.
- Walsh, W. E., D. C. Parkes, and R. Das. 2003. Choosing samples to compute heuristic-strategy Nash equilibrium. In *Agent Mediated Electronic Commerce*.
- Wellman, M. P., J. J. Estelle, S. Singh, Y. Vorobeychik, C. Kiekintveld, and V. Soni. 2005. Strategic interactions in a supply chain game. *Computational Intelligence* 21 (1): 1–26.
- Wellman, M. P., A. Greenwald, and P. Stone. 2007. *Autonomous bidding agents: Strategies and lessons from the trading agent competition*. MIT Press.
- Wellman, M. P., A. Osepayshvili, J. K. MacKie-Mason, and D. M. Reeves. 2008. Bidding strategies for simultaneous auctions. *B.E. Journal of Theoretical Economics (Topics)* 8 (1).
- Wellman, M. P., D. M. Reeves, K. M. Lochner, S.-F. Cheng, and R. Suri. 2005. Approximate strategic reasoning through hierarchical reduction of large symmetric games. In *Twentieth National Conference on Artificial Intelligence*, 502–508.
- Wellman, M. P., P. R. Wurman, K. O'Malley, R. Banger, S. de Lin, D. M. Reeves, and W. E. Walsh. 2001. The 2001 trading agent competition. *IEEE Internet Computing* 5 (2): 43–51.