Choosing Samples to Compute Heuristic-Strategy Nash Equilibrium

William E. Walsh
IBM T. J. Watson Research Center
19 Skyline Dr.,
Hawthorne, NY 10532, USA
wwalsh1@us.ibm.com

David C. Parkes
Division of Engineering and Applied Sciences,
Harvard University
33 Oxford St.
Cambridge, MA 02138, USA
parkes@eecs.harvard.edu

Rajarshi Das
IBM T. J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532, USA
rajarshi@us.ibm.com

ABSTRACT

Auctions define games of incomplete information for which it is often too hard to compute the exact Bayesian-Nash equilibrium. Instead, the infinite strategy space is often populated with *heuristic strategies*, such as myopic best-response to prices. Given these heuristic strategies, it can be useful to evaluate the strategies and the auction design by computing a Nash equilibrium across the restricted strategy space. First, it is necessary to compute the expected payoff for each heuristic strategy profile. This step involves sampling the auction and averaging over multiple simulations, and its cost can dominate the cost of computing the equilibrium given a payoff matrix. In this paper, we propose two information theoretic approaches to determine the next sample through an interleaving of equilibrium calculations and payoff refinement. Initial experiments demonstrate that both methods reduce error in the computed Nash equilibrium as samples are performed at faster rates than naive uniform sampling. The second, faster method, has a lower metadeliberation cost and better scaling properties. We discuss how our sampling methodology could be used within *experimental* mechanism design.

1. INTRODUCTION

Agent-mediated electronic commerce advocates the design of markets in which automated trading agents will engage in dynamic negotiation over the prices and attributes of goods and services. Trading agents promise to remove the monitoring and transaction costs that make dynamic negotiation impractical in traditional commerce. However, before these automated markets are widely deployed, it will be necessary to design trading agents that can follow useful (perhaps even optimal) strategies.

Certainly, careful market design and mechanism design can help, through the design of systems with simple but useful agent strategies (e.g. [7, 8, 11]). However, many real-world problems are simply too complex to be amenable to the theoretical approach of mechanism design (e.g. [13, 20]). First, the problem of optimal mechanism design is often not analytically tractable. Second, the result can be a mechanism that is not practical to deploy, either for reasons of communication complexity (e.g. [12]) or for reasons of computational complexity (e.g. [10]). Simply stated, the cost of simplifying the strategic problem of agents through careful mechanism design is often too high.

For this reason we should expect electronic markets in which the equilibrium trading strategy for an agent is not a straightforward strategy, such as myopic best-response to prices or truthful bidding. As an example, consider a continuous double auction (CDA) in which agents dynamically enter a marketplace and trade goods over time. Computing the Bayesian-Nash equilibria directly for interesting auctions (e.g. the CDA) often proves to be impossible with current methods. Instead, a typical approach is to compute a Nash equilibrium across a space of *heuristic* trading strategies[4, 22]. Heuristic strategies define the actions an agent will take within the auction, e.g. "bid b at time t." For example, in an earlier study of CDAs, we generated a *heuristic payoff table*–an analog of the usual payoff table, except that the entries describe expected payoffs to each agent as a function of the strategies played by all other agents [19]. The heuristic payoff table was then used as the basis for several forms of analysis, including computation of the Nash equilibria with respect to the restricted strategy space, and the market efficiency at those equilibria.

Common across previous work on the evaluation of heuristic trading strategies is the problem of measuring the expected payoff to each agent in the auction for all strategy profiles and populating a payoff table. This step involves sampling from a distribution of agent preferences, and then running the auction mechanism with a particular profile of heuristic strategies, in order to generate an additional sample in one cell of the payoff matrix. This step can be *much* more costly than computing the equilibrium. In the 2000 Trading Agent Competition [23], for instance, each run of the game requires 15 minutes for play and data collection, and the game must be run many times to fill out a payoff table. In contrast, it is possible to compute all equilibria within minutes [19]. Yet, simulations are typically performed statically and with the same number of samples for each entry in the payoff table.

In this paper, we address this problem of selecting simulations more intelligently. We describe methods to *interleave* the sampling of the payoff in the underlying market game with the calculation of Nash equilibrium, and present an information-theoretic methodology to the sampling problem. The high-level idea is quite standard. The methods are designed to sample the strategy profile that is expected to provide the most value of information, measured in terms

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
AAMAS-03 Workshop on Agent-Mediated Electronic Commerce V (AMEC-V) 2003 Melboune, Australia.

of beliefs about the effect that one more sample might have on the current decision about the equilibrium of the system. The difficulty in applying the framework is in the development of appropriate models with which to base these information-theoretic decisions, that are both principled yet fast enough to be useful in practice.

It is useful to place this work in the context of a wider agenda of an *experimental* approach to computational mechanism design. Just as *experimental economics* [15] provides a "wind tunnel" to measure the performance of auctions with human participants, we need our own wind tunnel for an experimental approach to the design of agent-mediated mechanisms. Central to this experimental approach to agent-mediated mechanism design is the ability to compute the performance of a particular market-based system, given a realistic population of trading agents. There is already a rich tradition in performing experiments with automated trading agents, and more recently in using methods from evolutionary computing to compute approximate equilibrium strategies [6, 19, 22]. We believe that sophisticated methods to effectively sample the underlying heuristic strategy space provide one component in developing a framework for effective experimental mechanism design for automated agents.

Following is an outline for the rest of this paper. Section 2 delineates how the heuristic payoff table is created, the costs involved, and how the table is used it to compute the underlying Nash equilibria. Section 3 discusses our information-theoretic approach to the problem of determining how to sample the payoffs of the heuristic strategy space. In Section 4 we present a more computationally efficient method to sample the strategy space. Section 5 provides empirical results of our sampling methods as applied to two specific games, and compares the performances with those from a uniform sampling approach. We conclude with a discussion of how our methodology could be used to aid experimental mechanism design.

2. HEURISTIC-STRATEGY NASH EQUILIBRIUM

We start with a game, such as an auction, that may include complex, repeated interactions between N agents. The underlying rules of the game are well-specified and common knowledge, but each agent has uncertain information about the types of the other agent. The rules specify particular *actions* that agents may take as a function of the state of the game. The *type* of an agent specifies individual, private characteristics of the agent, which, in the games we consider, specifies the agent's payoff for different outcomes in the game. It is instructive to consider an *ascending-price auction*: the *rules* specify closing rules and price dynamics; the *actions* allow an agent to bid at or above the current price; the *type* of an agent specifies its value for the item.

Many interesting and important games are too complex to compute Nash equilibria on the atomic actions with current game theory tools (at least no-one has yet been able too). This has led a number of researchers to devise innovative heuristic strategies—typically employing economic reasoning, optimization, and artificial intelligence techniques—to complex games including the Trading Agent Competitions [5, 17, 21], market-based scheduling [22], and the continuous double auction [2, 3, 18]. The key point is that the heuristic strategies are a subset of the (generally enormous) space of all strategies, and the set of heuristic strategies do not necessarily contain strategies that constitute an equilibrium of the underlying game (hence heuristic).

Thus, diverging from standard Bayesian-Nash equilibrium analysis, we can assume that each of the N agents has a choice of the *same M* exogenously specified, *heuristic strategies*, and compute a Nash equilibrium across this restricted strategy space. Given the heuristic strategies, we transform the underlying game to one in which the agents' payoffs are the expected payoffs obtained by the heuristic strategies in the underlying game, computed with respect to the distribution of utility functions (or types) in the underlying game [19, 22].

A heuristic strategy is simply an action policy specifying (generally complex) behavior for atomic actions in an underlying game. To give an example, in a continuous double auction (CDA), an agent's type specifies its value for the goods in the market. The underlying rules of the CDA allow agents to take actions of the form form "bid b at time t", while the heuristic strategies can be complex functions, expressed in hundreds or thousands of lines of code, that specify what bids are placed over the course of trading. One component of a CDA strategy could specify, for instance, to "place buy bid $s+\varepsilon$ when the lowest sell bid s is at most δ greater than the highest buy bid b.

Let $\mathcal{H}$ denote the space of heuristic strategies. A pure *strategy profile* $a = (a^{j_1}, \cdots, a^{j_N})$ specifies, for each agent i, the pure strategy $a^{j_i} \in \mathcal{H}$ played by the agent. A *payoff matrix* specifies the payoff to each agent for all possible strategy profiles. The standard payoff table requires M^N entries which can be extremely large, even when M and N are moderate. To mitigate this problem, we restrict our analysis to symmetric games in in which each agent has the same set of strategies and the same distribution of types (and hence payoffs). Hence, we can merely compute the payoff for each strategy as a function of the *number* of agents playing each strategy, without being concerned about the individual identities of those agents. This gives us a much smaller payoff matrices of size $\binom{N+M-1}{N}$. Standard auction analyses often rely on this symmetry assumption to simplify the problem. With the symmetry assumption, we generally dispense with the agent index.

Given a *heuristic strategy payoff matrix* mapping joint heuristic strategy choices to agent payoffs, we then compute a Nash equilibrium in the restricted space of heuristic strategies. Goldman et al. [4] have referred to this as an *experimental equilibrium*. We allow agent i to play a mixed strategy, and choose to play pure strategy $a^j \in \mathcal{H}$ according to probability x_{ij}. Let $x_i = (x_{i1}, \ldots, x_{iM})$ denote the complete mixed strategy, with $x_{ij} \in [0,1]$ and $\sum_{j=1}^{M} x_{ij} = 1$. The vector of all agents' mixed strategies is denoted x and the vector of mixed strategies for all agents except i is denoted x_{-i}. We indicate by $x_i = a^j$, the special case when agent i plays pure strategy j with probability one.

We denote by $u(a^j, x_{-i})$ the expected payoff to an agent i for playing pure strategy j, given that all other agents play their mixed strategies x_{-i}. The expected payoff to agent i with mixed strategy x_i is then $u(x_i, x_{i-1}) = \sum_{j=1}^{M} u(a^j, x_{i-1}) x_{i,j}$. In a Nash equilibrium,no one agent can receive a higher payoff by unilaterally deviating to another strategy, given fixed opponents' strategies. Formally, probabilities x^* constitute a *Nash equilibrium* iff for all agents i, and all $x_i \neq x_i^*$, $u(x_i, x_{-i}^*) \leq u(x_i^*, x_{-i}^*)$. To simplify the computation of equilibrium, in the remainder of this paper, we restrict our attention to symmetric mixed strategy equilibria, whereby $x_i^* = x_k^* = x^*$ for all agents i and k. It is known Nash equilibria of symmetric strategies always exist for symmetric games.

A Nash equilibrium computed with respect to expected payoffs is an *ex ante* Nash equilibrium for the restricted game, given the strategy space specified by the heuristic strategies. The equilibrium is not a Bayesian-Nash equilibrium (BNE) in the restricted strategy space because a full BNE would allow an agent to *choose* a different heuristic strategy for different realizations of its own type. Instead we require an agent to adopt the same heuristic strategy (e.g. "always bid at the price if the price is below my value and

I am not currently winning") whatever its actual value. Thus, an agent plays an *ex ante* Nash equilibrium instead of an *interim*, or Bayesian-Nash, equilibrium. As the heuristic strategy space $\mathcal{H}$ becomes rich and contains arbitrarily complex strategies this distinction disappears because a heuristic strategy can simulate the effect of mapping from the multiple possible types of an agent in multiple different underlying strategies (e.g. "if my value is less than \$5, then always bid at price if the price is below my value; otherwise, wait until the end of the auction and then snipe at 80% of my value.")[1]

To reiterate, the heuristic strategy approach is an approximation in which the designer of an auction consider only a very small subset of all possible strategies. As such, a Nash equilibrium in heuristic strategy space is *not* guaranteed to constitute an equilibrium in the underlying game.

Looking ahead to the sampling problem, before equilibrium can be computed we require information about the payoffs of heuristic strategy profiles. But, because the underlying game is complex the *ex ante* payoffs are not analytically derivable and must be computed instead as average payoffs over repeated simulations. It can be necessary to perform a large number of simulations to obtain sufficiently accurate *ex ante* payoff estimates to calculate the Nash equilibria accurately. This is particularly expensive for games that must be run in real time. For instance, the Trading Agent Competition requires 15 minutes for each game, which practically limits the number of payoff samples that can be collected. However, it is not generally necessary to estimate all payoffs to the same degree of accuracy. For instance, if the one equilibrium is for all agents to play the pure strategy a_j, then we will not need to accurately estimate the payoff for all agents playing any other pure strategy to determine that a_j is the equilibrium. Instead, we only need to *bound* the payoffs available from alternative heuristic strategies $a_{j'} \neq a_j$ when every other agent plays a_j. Thus, taking samples uniformly for all strategy profiles may not be the most efficient method to arrive at an accurate equilibrium.

In the remainder of this paper we present information-theoretic methods for selecting which strategy profiles to sample from.

3. AN INFORMATION-THEORETIC APPROACH

In this section, we outline an information-theoretic approach to the problem of determining how to sample the underlying space of heuristic strategies to build the payoff matrix and compute equilibrium strategies.

Let S denote the space of all sample actions, and let $\theta \in S^L$ denote a sequence of sample actions of length L. An example of a sample action could be "perform 10 experiments in which agents follow (pure) strategy profile a", where $a = (a^{j_1}, \ldots, a^{j_N})$ with $a^{j_i} \in \mathcal{H}$ to indicate the heuristic strategy selected by agent i. We find it convenient to overload θ, and use θ to also denote the *information* that results from the new samples.

An information-theoretic approach to sampling requires three modeling assumptions. First, we need a *decision model*, $x(\theta)$, to denote the equilibrium selected, given information θ. Second, we need a *future information model* to predict the cumulative information that will be available after additional samples s, given current information θ. Third, we need an *error model* to estimate the error of the equilibrium selected due to current beliefs about the payoffs, with respect to the true equilibrium in the auction.

Looking ahead, we will choose to define the error in our setting in terms of the gain in expected payoff that an agent can receive by deviating from the current equilibrium decision to a pure strategy, summed across all pure strategies. Clearly, in equilibrium this payoff from deviation is zero and the agent has the same expected payoff for all pure strategies in the support of its mixed strategy and less for any other strategy. Let $\hat{f}_\theta(x)$, denote the estimated error from decision x, as estimated with respect to information θ. Notice that we can only estimate the true error, which we denote as $f_\pi(x)$, where π are the true payoffs.

The framework introduced by Russell & Wefald [16] for metadeliberation in time-critical decision problems with bounded-rational agents proposes to predict the *value of information*, $\mathrm{EVI}(s|\theta)$, for sampling strategy s given information θ as:

$$\mathrm{EVI}(s|\theta) = \mathrm{E}_{s|\theta}\left[\hat{f}_{\theta.s}(x(\theta)) - \hat{f}_{\theta.s}(x(\theta.s))\right]$$

where $\mathrm{E}_{s|\theta}$ takes the expectation with respect to a model of the future samples s given current information θ. Intuitively, $\mathrm{EVI}(s|\theta)$ measures, in expectation, the degree to which further samples s will reduce the estimated error in the equilibrium choice. Notice that the first term is $\hat{f}_{\theta.s}(x(\theta))$ and not $\hat{f}_\theta(x(\theta))$, so that any effect that the information has on refining the *accuracy with which the error is computed* is factored out of this analysis. As observed by Russell & Wefald, this is important to maintain the useful property that the estimated value of information is positive for all possible sampling strategies.

In our model, the information θ that has already been accumulated through current samples provides a set of samples for each entry a in the payoff matrix. With this, the maximum likelihood estimator (MLE) for the *true mean*, $\mu(a)$, of strategy profile a, written $\mu_\theta(a)$ is computed as the sample mean. By the central limit theorem,[2] with sufficient number of samples (generally, 30 is considered sufficient) from any distribution, the true mean becomes normally distributed, with mean $\mu_\theta(a)$, and standard deviation $\sigma_{\mu_\theta(a)} = \sigma(a)/\sqrt{t_a}$, where $\sigma_\theta(a)$ is the standard deviation of the samples collected for a and t_a is the number of samples collected for a. We find it useful to refer to $\mu_\theta(a)$ and $\sigma_{\mu_\theta}(a)$ as the *observed mean*, and the *standard deviation over the observed mean*, given information θ. In the sequel, we drop the a indicator when the specific profile is understood or immaterial to the discussion.

An optimal sampling strategy would take this definition of the expected value of information, and formulate an optimal *sequential* sampling strategy with future decisions contingent on the information returned from earlier decisions, and for a fixed number of total samples. The objective would be to select a sequential sampling strategy to maximize the expected total decrease in decision error by the end of the sampling period. Only the first step of the contingent sequential sampling strategy would be executed (e.g. *perform one experiment in which agents follow strategy profile a*), at which point the complete sequential strategy would be reoptimized based on the new information. Berger [1, chapter 7] provides an extensive discussion of this methodology, which is central to statistical decision theory.

In practice, the best one can hope for is an approximation to this approach. Clearly metareasoning is valuable only to the extent that the time spent in metareasoning is less than the time saved through making better-informed sampling decisions. Russell & Wefald make a number of assumptions to keep metareasoning tractable for their setting. Most importantly, they make a *single-step* assumption, which in our setting is equivalent to assuming that only one more experiment will ever be performed. This reduces the sampling

[1] We thank Michael Wellman for explaining this latter point.

[2] The central limit theorem assumes that samples are independent and of finite variance. The first assumption holds if strategies are static (they do not adapt) and the second assumption holds if payoffs in the underlying game are bounded.

decision to choosing the single best sampling action, to maximize $\mathrm{EVI}(s|\theta)$.

The main problem with the single-step assumption in our setting is that it is quite possible that no single additional sample will be expected to have an effect on the decision made about the equilibrium. Yet, we want to avoid deciding to commit to a long sequence of additional samples without a chance to reconsider. To resolve this dilemma we: (1) assume that the space of sampling actions, S, allows the system to choose sample strategy $s \in S$ a total of K times, for some fixed K; and (2) within metareasoning we consider the informational effect for a *long* sequence s^∞ of the same sampling action, s, but then only execute the first of the sequence with the greatest EVI before repeating the calculation of EVI.

Given this, we define the three components of our information-theoretic model as follows:

decision model. The equilibrium $x(\theta)$, given information θ, is computed as *one* of the mixed equilibria given the *mean* payoffs μ_θ in each entry in the payoff matrix. In particular, we select the equilibrium with the lowest error estimated from current information $\hat{f}_\theta(x)$.

future information. Given the current information $(\mu_\theta, \sigma_\theta)$, we need a model for the effect that a large number of additional samples s^∞ on profile a will have on the future observed mean payoff, $\mu_{\theta.s^\infty}(a)$, and the future standard deviation on observed mean, $\sigma_{\mu_{\theta.s^\infty}}(a)$. We adopt two models for the future observed mean: (a) a *point*-estimate, with $\mu_{\theta.s^\infty}(a) = \mu_\theta(a)$; and (b) a *distributional*-estimate, with $\mu_{\theta.s^\infty}(a) \sim N(\mu_\theta(a), \sigma_{\mu_\theta}(a))$. We model the future standard deviation on the observed mean as $\sigma_{\mu_{\theta.s^\infty}}(a) = \sigma_{\theta(a)}/\sqrt{t_a + |s^\infty|})$, where t_a is the number of samples collected for a so far and $|s^\infty|$ is the number of samples in s^∞.

error. We define the true error function $f_\pi(x)$ with respect to payoffs π as $f_\pi(x) = \sum_{j=1}^{M} \max(0, u_i(a^j, x_{-i}) - u_i(x))$. That is, $f_\pi(x)$ is the sum of incentives for an agent to unilaterally deviate from the mixed strategy x to any pure strategy, given the payoffs. It is well-known that x is a Nash equilibrium (with respect to π) iff $f_\pi(x) = 0$. We compute the estimated error $\hat{f}_\theta(x)$ given information θ from Monte Carlo simulations on the actual error, as described below.

Looking at the models of future information, the point-estimate of the future observed mean reflects the fact that our estimate of the true mean will remain the same in expectation. In comparison, the distributional-estimate considers that we expect the observed mean will converge to the true mean after many additional samples, and recognizes that the current information $(\mu_\theta, \sigma_{\mu_\theta})$ is the best current estimate for the true mean. The model for the future standard deviation on observed mean reflects an assumption that the standard deviation on the underlying payoff samples will remain the same as that for the current samples.

The complete algorithm for selecting the next samples is as follows. First, compute the set of Nash equilibria NE given θ, and choose the $x(\theta) \in \mathrm{NE}$ that minimizes estimated error $\hat{f}_\theta(x(\theta))$. Then choose the s that maximizes $\mathrm{EVI}(s|\theta)$. If $\mathrm{EVI}(s|\theta)$ is greater than the cost of performing the first K samples in s, perform those sample simulations and continue the process. Otherwise, stop and return x as the chosen equilibrium. The model of cost and information value will depend heavily on the particular details of the problem, with the cost depending on the run time of simulations and the value of information depending on the importance of making an accurate decision. For example, in the context of an experimental approach to mechanism design we can interpret the value of a decision with respect to the goals of a mechanism designer, such as allocative efficiency. In this paper, we sidestep this issue and compare the decision accuracy across alternative sampling methods for the same total number of samples.

To compute the estimated error $\hat{f}_{\theta.s}(x(\theta))$ for the current decision after additional information, we adopt the point-estimate model for the future observed mean after a large number s^∞ of additional experiments, together with the point-estimate model for the future standard deviation on observed mean. We average the results from C_f Monte Carlo simulations, with each simulation computing $f_{\hat{\pi}}(x(\theta))$ for a draw $\hat{\pi}$ of specific payoffs from the distribution $N(\mu_{\theta.s^\infty}, \sigma_{\mu_{\theta.s^\infty}})$ on true payoffs, with $\mu_{\theta.s^\infty} = \mu_\theta$ and $\sigma_{\mu_{\theta.s^\infty}} = \sigma_\theta/\sqrt{t_a + |s^\infty|})$.

To compute the estimated error $\hat{f}_{\theta.s}(x(\theta.s))$ for the new and improved decision after additional information, we must first estimate the future decision. For this, we adopt the distributional-estimate, $\mu_{\theta.s^\infty} \sim N(\mu_\theta, \sigma_{\mu_\theta})$, for the future observed mean. We sample C_E mean payoffs π' from this distribution,and compute a new equilibrium $x(\pi')$ for each. Then, we measure the estimated error for each of these decisions using the same model for future information as was adopted to compute the estimated error $\hat{f}_{\theta.s}(x(\theta))$ for the current decision, taking an additional C_f samples for each π'. Finally, we average this estimated error for future decision $x(\pi')$ across all C_E equilibrium samples.

We note that an alternative model to compute the estimated error $\hat{f}_{\theta.s}(x(\theta.s))$ for the new decision after additional information would use a hybrid of the two models for future information. We could adopt the *same* sample π' that is used to compute a future equilibrium decision $x(\pi')$ to model the future observed mean for the purposes of computing the estimated error on that decision, but continue to adopt $\sigma_{\mu_{\theta.s^\infty}} = \sigma_\theta/\sqrt{t_a + |s^\infty|})$ to generate C_f samples for this error calculation. We plan to investigate this hybrid model in future work.

4. A FASTER APPROACH

As we demonstrate in Section 5, $\mathrm{EVI}(s|\theta)$ is an effective criterion for selecting a pure strategy profile to sample next. Unfortunately, $\mathrm{EVI}(s|\theta)$ is slow to compute for even very small games, and impractically slow for moderately sized games. The problem lies in the fact that $\mathrm{EVI}(s|\theta)$ must perform multiple equilibrium computations for each possible sample sequence. For each s we must compute multiple sample future equilibria to estimate the distribution of future equilibria. Although we have tricks to fairly quickly compute a future equilibrium based on the current equilibrium (see Section 5), the computational cost can be prohibitively high, given that we perform C_E equilibrium computations for each strategy profile.

We have developed a much faster method for computing the value of performing a set of further samples that requires no additional equilibrium computations. The algorithm is the same as before, except that, instead of $\mathrm{EVI}(s|\theta)$, we use the estimated *confirmational value of information*, $\mathrm{ECVI}(s|\theta)$, of sampling policy s given information θ, defined as:

$$\mathrm{ECVI}(s|\theta) = \mathrm{E}_{s|\theta}\left[\hat{f}_\theta(x(\theta)) - \hat{f}_{\theta.s}(x(\theta))\right].$$

Intuitively, $\mathrm{ECVI}(s|\theta)$ measures, in expectation, the degree to which further samples s would decrease the estimated error of the current equilibrium choice. Put another way, the s that maximizes $\mathrm{ECVI}(s|\theta)$ provides the best evidence to confirm our current equilibrium choice.

We need not compute any future equilibria with this approach, but need only perform Monte Carlo simulations to estimate the expected error of the current equilibrium. Furthermore, we need perform significantly fewer of these Monte Carlo simulations than for $\text{EVI}(s|\theta)$. If we perform C_f Monte Carlo simulations to estimate the error for each future equilibrium, then $\text{EVI}(s|\theta)$ requires $C_E C_f$ simulations of $f_\pi(x)$, while $\text{ECVI}(s|\theta)$ requires only C_f.

Using $\text{ECVI}(s|\theta)$ appears to run counter to the methodology of Russell and Wefald, who argue that further samples are valuable only to the extent to which they may *change* the equilibrium choice. Still, based on some informal arguments, we can view the approach as a reasonable heuristic to indirectly choose the s that will likely most change the decision. That is, $\text{ECVI}(s|\theta)$ can be viewed as a fast approximation of Russell and Wefald's approach for our problem.

Recall that there is a continuum of mixed strategies that could potentially comprise a Nash equilibrium. Thus, so long as there remains "sufficient" uncertainty in the value of the payoffs in a cell c of the game matrix, we should expect that further samples of c will change the chosen equilibrium, however slightly. Thus, although we choose s to confirm that x is the correct equilibrium, sampling s will, in fact, generally change our decision.

Why should we expect that choosing s to increase our confidence in x would best help change the decision for the better? Since s has the highest direct impact on reducing error for x, we should expect that it also reduces error for some mixed strategies in the probability region around x. And since x is our best estimate of the correct equilibrium, given current information, then we should expect the true equilibrium x^* to lie in the probability region around x. Thus, s can be considered our best guess for reducing the error in x^*, hence making it a more promising candidate for equilibrium selection by our method.

Ultimately, the value of using $\text{ECVI}(s|\theta)$ is clearly borne out in our experiments. It is much faster than $\text{EVI}(s|\theta)$, yet reduces error in the chosen equilibrium at a rate comparable to $\text{EVI}(s|\theta)$.

5. EXPERIMENTS

This section describes empirical results of our methods for choosing samples of payoffs from the space of heuristic strategy profiles, and compares their approach to the standard approach of sampling uniformly. To chart the progress of the approaches as more simulations are performed, we perform the uniform samples in a round-robin fashion. We report on results from two artificial games: (i) one with $N = 1$ agent and $M = 5$ strategies, and (ii) one with $N = 8$ agents and $M = 3$ strategies. So that we could compare the results of the methods with the true equilibria of the games, we generated payoff samples directly from specified distributions, rather than from simulation of a complex auction game.

For each game and each method, $\text{EVI}(s|\theta)$ and $\text{ECVI}(s|\theta)$, we performed 10 initial simulations on each strategy profile to seed the payoff table, before applying the methods. The length of s^∞, which is used to estimate the value of additional sampling within metadeliberation, was 1000, and the length of s, the samples actually performed on one profile, was 10. We perform the same number of total samples with each sampling method and compare the error in the equilibrium decision. For game (i), we performed 1000 total samples, and game (ii) we performed 2000 total samples. We ran each method 200 times on game (i), each time with a different random seed. Computing $\text{EVI}(s|\theta)$ is prohibitively expensive for game (ii), hence we ran only $\text{ECVI}(s|\theta)$ and uniform sampling (both for 200 times) but not $\text{EVI}(s|\theta)$.

For the purpose of the analysis, we graph two error measures of the chosen equilibrium x. Both are calculated with respect to the *true* game information, and as a function of the number of simulations performed and averaged over all runs. The first error measure is $f_\pi(x)$ discussed before, where π are the true payoffs. The second measure is the L2 Norm, defined as $\sum(x_j - x_j^*)^2$, where x_j is the probability of playing heuristic strategy a^j in profile x, and likewise for x^*, the true equilibrium. When multiple true equilibria exist, we compute the minimal L2 Norm across all equilibrium.[3]

Recall that our approaches require repeated computation of Nash equilibrium. Nash equilibria can be formulated and computed in a variety of ways [9]. We use a formulation of Nash equilibrium as a non-linear optimization problem, in which the equilibria lie at the zero points of the objective function. We used amoeba [14], a direct search method to find the zero points. Although amoeba (and indeed any Nash equilibrium solver) is not guaranteed to find all, or even any equilibria, we have found it effective in finding all equilibria for even reasonably large problems (e.g. 20 agents and 3 strategies) [19]. We find that the expensive part is not finding an equilibrium, but verifying to our satisfaction that all equilibria have been found, requiring that we restart at a number of random points. However, in our sample selection algorithms, performing new samples does not move the equilibria too far from their previous location. Thus, we can start the search at the previously computed equilibrium points and the gradient descent most often quickly converges to the nearby new equilibrium (although, at times we do have to recompute the equilibria from scratch).

5.1 1-agent 5-strategies game

This is a degenerate game with only one agent, where each strategy corresponds to a unique profile and the unique equilibrium is simply the (pure) strategy with the highest mean payoff. This feature is computationally advantageous though, for we are able to very quickly compute the equilibrium by bypassing amoeba and simply finding the highest payoff strategy. Since there are only 5 strategy profiles, the cost of computing $\text{EVI}(s|\theta)$ and $\text{ECVI}(s|\theta)$ is further reduced. Furthermore, it is easy to get some informal validation of our methods by inspection in this simple game.

We model the true payoff distribution in each of the 5 entries in the payoff table as normal, with the parameters (μ, σ) as follows:

Strategy	μ	σ
1	1.0	0.5
2	0.9	0.5
3	0.5	0.1
4	0.3	0.3
5	0.35	0.4

We can imagine that the variation in payoffs come from variations in the value of the agent, and or in the agent's environment. It is evident that the expected-utility maximizing strategy is to play strategy 1 (i.e., $x^* = (1,0,0,0,0)$), which has the highest mean. Clearly, strategy 2, with its high mean and large standard deviation,

[3]We note one twist in the implementation of our methods for these experiments. In game (i) methods $\text{EVI}(s|\theta)$ and $\text{ECVI}(s|\theta)$ both almost always predict that additional sampling has zero value long before 1000 simulations are performed (within an average of 163 simulations for $\text{EVI}(s|\theta)$ and 304 simulations for $\text{ECVI}(s|\theta)$). To provide a useful experimental comparison, we choose to continue sampling until 1000 experiments, choosing the profile with the smallest L2 Norm distance to the current equilibrium. In this situation, there other possibilities for allocating the new simulations. For example, the new simulations could be assigned to a randomly chosen profile. In fact, such a randomized approach may help a sampling strategy to avoid premature convergence to a solution which is not a true Nash equilibrium. We note that $\text{ECVI}(s|\theta)$ does not reach zero in game (ii) within 2000 simulations, hence this issue does not arise.

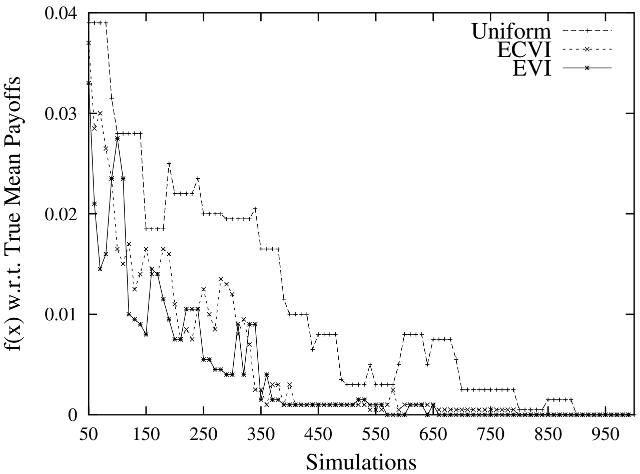

Figure 1: True error $f_\pi(x)$ of the computed equilibrium x with respect to true mean payoffs from the three sampling methods on the 1-agent 5-strategies game. The three sampling methods are uniform, $\mathrm{EVI}(s|\theta)$, and $\mathrm{ECVI}(s|\theta)$. Results shown here are averaged over 200 separate runs, where each run consist of a total of 1000 samples.

is by far the most likely candidate for the Nash equilibrium after strategy 1. A few samples would distinguish quickly that strategy 1 and 2 are the top two strategies, hence we would expect a good sampling method would assign most experiments to these strategies in an effort to distinguish which is clearly the one with the highest payoff.

Figure 1 shows the $f_\pi(x)$ error measure for the three methods. As expected, after the initial 50 experiments, all three methods find equilibria that have roughly the same magnitude of $f_\pi(x)$. After this initial stage, $f_\pi(x)$ of the $\mathrm{ECVI}(s|\theta)$ and $\mathrm{EVI}(s|\theta)$ methods decrease more rapidly and reach zero error faster than the for the uniform sampling method, demonstrating the effectiveness of our approach. Moreover, until the stage at which our sampling methods achieve zero error, our methods achieve significantly lower $f_\pi(x)$ than the uniform sampling approach, for a given number of experiments. In our experiments, method $\mathrm{EVI}(s|\theta)$ has a somewhat lower $f_\pi(x)$ than $\mathrm{ECVI}(s|\theta)$ through most of the simulations, although both methods reach near zero error at near the same point.

For this particular game the L2-Norm is identical to the $f_\pi(x)$ after scaling. This is because, given the true payoff distributions, the computed equilibria is most likely to be at strategy 1 (i.e., $x^* = (1,0,0,0,0)$), or at strategy 2 (i.e., $x^* = (0,1,0,0,0)$).

We also notice that the sampling methods are selective in assigning new simulations. While the $\mathrm{EVI}(s|\theta)$ method continues to define samples, it assigns on average 50.3% and 49.6% of the total number of experiments allocated up to that point to strategy 1 and strategy 2, respectively. The remaining 0.1% of the experiments are assigned to the other strategies. The $\mathrm{ECVI}(s|\theta)$ method exhibits a similar emphasis by assigning 51.2% and 46.5% of the total number of experiments to strategy 1 and and strategy 2 respectively. Thus, the methods make decisions that correspond to the intuitively correct distribution of simulations.

5.2 8-agents 3-strategies game

This is a more complex game with a total of 45 strategy profiles. The true μ for each payoff was chosen randomly from a uniform distribution [300, 600], with $\sigma = 50$ for all payoffs. Again, the

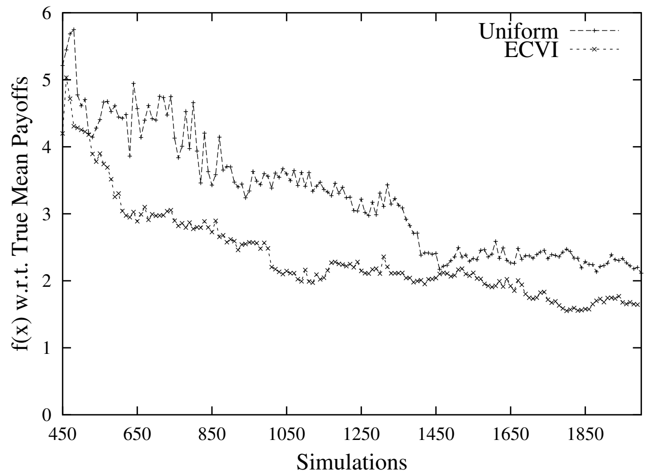

Figure 2: True error $f_\pi(x)$ of the computed equilibrium x with respect to true mean payoffs from sampling methods uniform and $\mathrm{ECVI}(s|\theta)$ on the 8-agents 3-strategies game. Results shown here are averaged over 200 separate runs, where each run consist of a total of 2000 experiments.

distribution over payoff is adopted as a model over the distribution over types of agents. The three mixed-strategy Nash equilibria computed from the true mean payoffs are $\{(0.4114, 0.0000, 0.5886), (0.0000, 0.5213, 0.4787), (0.0268, 0.0877, 0.8856)\}$.

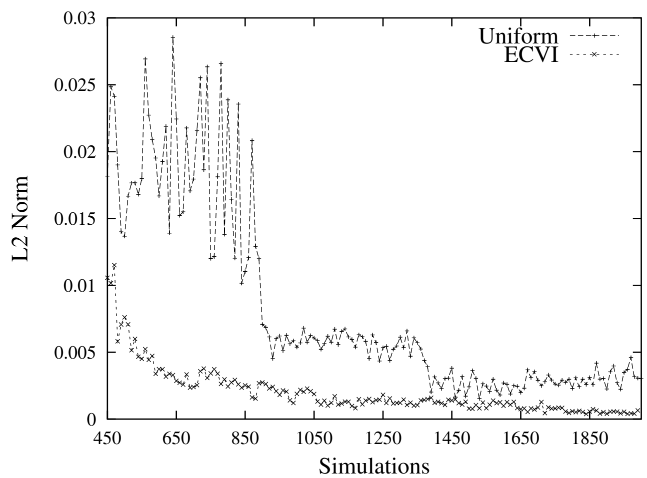

Figure 3: L2-Norm of two sampling methods on the 8-agents 3-strategies game. Results shown here are averaged over the same 200 runs reported in Figure 2.

Figure 2 plots $f_\pi(x)$ for the $\mathrm{ECVI}(s|\theta)$ and uniform sampling methods. As with the 1-agent, 5-strategy game, using $\mathrm{ECVI}(s|\theta)$ gives a smaller error for any number of simulations than does the uniform method.

We see similar, but more pronounced results when comparing with the L2 Norm, as shown in Figure 3. The rather small L2 Norm of ≈ 0.0005 after 1050 simulations with the $\mathrm{ECVI}(s|\theta)$ method indicates that the computed mixed-strategy equilibrium in each run is indeed very close to one of the Nash equilibria determined from the true means. In other words, even though there remains some benefit to an agent for deviation to another strategy, the actual mixed equilibrium is very close to the true equilibrium.

A closer analysis of each run also indicate that for this game, the ECVI$(s|\theta)$ method chooses samples very non-uniformly between the strategy profiles. Strategy profiles that are closer to one of the true Nash equilibria (in terms of L2 Norm) received 5% to 10% of the 2000 experiments, while some profiles were assigned virtually no new experiments after their initial set of 10 experiments. We conclude that we can choose simulations selectively (i.e., non-uniformly) and get lower error using ECVI$(s|\theta)$.

6. DISCUSSION

We have presented two methods for choosing the samples necessary to compute a Nash equilibrium with respect to ex ante payoffs of heuristic strategies. The first method is a direct application of an information-theoretic methodology proposed by Russell and Wefald, but is impractically slow. We were not able to run the method even on our 8-agents and 3-strategies problem. The second method approximates the first method, but at a much lower computational cost. Both methods are designed to select samples to perform that are most likely to improve our estimation of the equilibrium. Our initial experiments suggest that these selective approaches give us an equilibrium with less error from fewer simulations than a uniform sampling method. Moreover, the fast, approximate method reduces error nearly as quickly as the direct information-theoretic method on the small single-agent problem.

The most interesting direction in which to extend and apply this work is to the problem of *experimental* mechanism design, in which computational methods are used in a closed loop to evaluate the designs of alternative electronic markets. In this setting, the relevant question becomes: *how should the system allocate experiments across multiple market designs*, in addition to within a particular market design. Our information-theoretic methodology can be extended quite naturally to this new problem, in which costly simulations help to *choose* a market design. In this setting the goals of the mechanism designer, for example allocative efficiency, provide a compelling method with which to define the value of information with respect to the decision error. For example, rather than measure the *accuracy* of the equilibrium in terms of individual agent utilities, we can measure the *loss in efficiency* implied by making a decision with current information, and weigh the cost of further experiments with the possible efficiency gains of a better design.

Looking ahead, we believe an interesting research agenda for agent-mediated electronic commerce is to relax some of the requirements of mechanism design (in particular *incentive-compatibility* and *direct-revelation*), but still seek to design market-based systems that enjoy desirable economic properties when populated by realistic (and necessarily bounded-rational) trading agents. In particular, notice that it is not necessary to give up on *design* when giving up on the formal mathematics of mechanism design. Instead, one moves from an off-line and analytic approach to mechanism design to an *experimental* approach to mechanism design. Rather than modeling the behavior of agents offline, and designing a mechanism with respect to this model we can design a market mechanism from a space of mechanisms and then *experimentally validate* the performance of the mechanism with respect to a set of desiderata. This is attractive because we can explicitly design for automated trading agents, since these agents can be used *explicitly* to evaluate the performance of a proposed market design.

Acknowledgments

We thank Dan Reeves for his helpful discussions on this topic and comments on this paper. David Parkes was supported in this work by an IBM Faculty Award.

7. REFERENCES

[1] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 1985.

[2] D. Cliff and J. Bruten. Minimal-intelligence agents for bargaining behaviors in market-based environments. Technical Report HPL-97-91, Hewlett Packard Labs, 1997.

[3] S. Gjerstad and J. Dickhaut. Price formation in double auctions. *Games and Economic Behavior*, 22:1–29, 1998.

[4] C. V. Goldman, S. Kraus, and O. Shehory. Agent strategies: for sellers to satisfy purchase-orders, for buyers to select sellers. In *Tenth European Workshop on Modeling Autonomous Agents in a Multiagent World*, 2001.

[5] A. Greenwald. The 2002 trading agent competition: An overview of agent strategies. *AI Magazine*, 2003.

[6] J. Kephart and A. Greenwald. Shopbot economics. *Autonomous Agents and Multiagent Systems*, 2002. forthcoming.

[7] A. Kothari, D. C. Parkes, and S. Suri. Approximately-strategyproof and tractable multi-unit auctions. In *Fourth ACM Conf. on Electronic Commerce (EC'03)*, 2003. To appear.

[8] D. Lehmann, L. I. O'Callaghan, and Y. Shoham. Truth revelation in approximately efficient combinatorial auctions. *Journal of the ACM*, 49(5):577–602, September 2002.

[9] R. D. McKelvey and A. McLennan. Computation of equilibria in finite games. In *Handbook of Computational Economics*, volume 1. Elsevier Science B. V., 1996.

[10] N. Nisan and A. Ronen. Computationally feasible VCG mechanisms. In *Proc. 2nd ACM Conf. on Electronic Commerce (EC-00)*, pages 242–252, 2000.

[11] N. Nisan and A. Ronen. Algorithmic mechanism design. *Games and Economic Behavior*, 35:166–196, 2001.

[12] N. Nisan and I. Segal. The communication complexity of efficient allocation problems. Technical report, Hebrew University and Stanford University, 2002.

[13] D. C. Parkes, J. R. Kalagnanam, and M. Eso. Achieving budget-balance with Vickrey-based payment schemes in exchanges. In *Proc. 17th International Joint Conference on Artificial Intelligence (IJCAI-01)*, 2001.

[14] W. Press, S. Teukolsky, W. Vetterling, and B. . Flannery. *Numerical recipes in C*. Cambridge University Press, 1992.

[15] A. E. Roth. The economist as engineer. *Econometrica*, 70:1341–1378, 2002.

[16] S. Russell and E. Wefald. Principles of metareasoning. *Artificial Intelligence*, 49:361–395, 1991.

[17] P. Stone and A. Greenwald. The first international trading agent competition: Autonomous bidding agents. *Journal of Electronic Commerce Research*, to appear.

[18] G. Tesauro and R. Das. High-performance bidding agents for the continuous double auction. In *Third ACM Conference on Electronic Commerce*, pages 206–209, 2001.

[19] W. E. Walsh, R. Das, G. Tesauro, and J. O. Kephart. Analyzing complex strategic interactions in multi-agent systems. In *AAAI-03 Workshop on Game Theoretic and Decision Theoretic Agents*, pages 109–118, 2002.

[20] W. E. Walsh and M. P. Wellman. Efficiency and equilibrium in task allocation economies with hierarchical dependencies. In *Proc. 16th International Joint Conference on Artificial Intelligence (IJCAI-99)*, pages 520–526, 1999.

[21] M. P. Wellman, A. Greenwald, P. Stone, and P. R. Wurman. The 2001 trading agent competition. *Electronic Markets*, 13(1):4–12, 2003.

[22] M. P. Wellman, J. K. MacKie-Mason, D. M. Reeves, and S. Swaminathan. Exploring bidding strategies for market-based scheduling. In *Fourth ACM Conference on Electronic Commerce, to appear*, 2003.

[23] M. P. Wellman, P. R. Wurman, K. O'Malley, R. Bangera, S. Lin, D. Reeves, and W. E. Walsh. Designing the market game for a trading agent competition. *IEEE Internet Computing*, 5(2):43–51, 2001.