Stochastic Learning of Strategic Equilibria for Auctions

Samy Bengio
Microcell Labs
1250, René-Lévesque West, suite 400
Montréal, Québec, Canada, H3B 4W8
samy@labs.microcell.ca

Yoshua Bengio
CIRANO and *Dept. IRO*
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
bengioy@iro.umontreal.ca

Jacques Robert
CIRANO and *Dept. Sciences*
Economiques, Université de Montréal
Montréal, Québec, Canada, H3C 3J7
robertj@cirano.umontreal.ca

Gilles Bélanger
Dept. Sciences Economiques
Université de Montréal
Montréal, Québec, Canada, H3C 3J7
belangerg@cirano.umontreal.ca

October 7, 1998

Abstract

This paper presents a new application of stochastic adaptive learning algorithms to the computation of strategic equilibria in auctions. The proposed approach addresses the problems of tracking a moving target and balancing exploration (of action space) versus exploitation (of better modeled regions of action space). Neural networks are used to represent a stochastic decision model for each bidder. Experiments confirm the correctness and usefulness of the approach.

1 Introduction

This paper presents a new application of stochastic adaptive learning algorithms to the computation of strategic equilibria in auctions. Game theory has become a major formal tool in economics. A game specifies a sequence of decisions leading to different possible outcomes. Each player or participant is attached to some decision contexts and information sets, and is provided with preferences over the set of possible outcomes. A game provides a formal model of the strategic thinking of economic agents in this situation. An *equilibrium* characterizes a stable rule of behavior for rational players in the game. A strategy for a player is the decision-making rule that he follows in order to choose his actions in a game. A strategic equilibrium (or Nash equilibrium) for a game specifies a strategy for all players which is a best-response against the strategies of the others. Let S_i denote the set of strategies for player i in $N = \{1, 2, \ldots, n\}$ and let $U_i : S_1 \times S_2 \ldots \times S_n \rightarrow R$ represent i's real-valued preference for a strategy s_i given the strategies of the other players, over the set of all outcomes of the game. A vector of strategies $s^* = \{s_1^*, s_2^*, \ldots, s_n^*\}$ forms a *strategic equilibrium* for the n-player game if for all $i \in N$:

$$s_i^* \in \mathrm{argmax}_{s_i \in S_i} U_i(s_1^*, \ldots, s_{i-1}^*, s_i, s_{i+1}^*, \ldots, s_n^*) \qquad (1)$$

At strategic equilibrium, no player wishes to change its strategy given the strategies of the others. In zero-sum games or games in which there is no possibility for cooperation (the latter being the case in auctions), this is a point where a group of rational players will converge and therefore it would be very useful to characterize strategic equilibrium. The approach proposed here to approximate strategic equilibria is quite general and can be applied to many game-theoretical problems. A lot of research has been done in the field of stochastic learning automata applied to game problems. A good review can be found in (Narendra and Thathachar, 1989). We will explain in section 3 the main differences between our approach and others.

In this paper, we focus on the application to auctions. An auction is a market mechanism with a set of rules that determine who gets the goods and at what price, based on the bids of the participants. Auctions appear in many different forms (McAfee and McMillan, 1987). Auction theory is one of the applications of game theory which has generated considerable interest (McMillan, 1994). Unfortunately theoretical analysis of auctions has some limits. One of the main difficulties in pursuing theoretical research on auctions is that all but the simplest auctions are impossible to solve analytically. Whereas previous work on the application of neural networks to auctions focused on emulating the behavior of human players or improving a decision model when the other players are fixed (Dorsey, Johnson and Van Boening, 1994), the objective of this paper is to provide new numerical techniques to search for strategies which appear to correspond to strategic equilibria in auctions, i.e., take empirically into account the feedback of the actions of one player through the strategies of the others. This will help predict the type of strategic behavior induced by the rules of the auctions, and ultimately make predictions about the relative performance of different auction rules.

For the purpose of this paper, we shall focus on a simple auction where n (risk-neutral) bidders compete to buy a single indivisible item. Each bidder i is invited to submit a (sealed) bid b_i. The highest bidder wins the item and pays his bid. This is referred to as the first-price sealed bid auction. If i wins, the benefit is $v_i - b_i$, where we call the *valuation* v_i the expected monetary gain for receiving the unit. The bid b_i is chosen in $[0, v_i]$. In this auction, the only decision context that matters is this valuation v_i. It is assumed to be information private to the bidder i, but all other bidders have a belief about the distribution of v_i. We let $F_i(\cdot)$ denote the cumulative distribution of i's valuation v_i . A strategic equilibrium for this auction specifies for each player i a monotonic and invertible bidding function $b_i(v_i)$ which associates a bid to each possible value of v_i. At the strategic equilibrium, one's bidding strategy must be optimal given the bidding strategies of all the others. Since each bidder's v_i is chosen independently of the others, and assuming that $b_i(v_i)$ is deterministic, the probability that bid b is winning for player i is $G_i(b) = \prod_{j \neq i} F_j(b_j^{-1}(b))$, i.e., the product of the probabilities that the other players bids are less than b.

Therefore the optimal bidding strategy for risk-neutral bidders is

$$b_i(v_i) \in \text{argmax}_b (v_i - b) G_i(b). \quad (2)$$

If the distributions F_i's are the same for all bidders, the strategic equilibrium can be easily obtained analytically. The symmetric bidding strategy is then given by:

$$b(v) = \int_{p_0}^{v} s \frac{dF(s)^{n-1}}{F(v)^{n-1}} \quad (3)$$

where p_0 is the lowest price acceptable by the auctioneer. However, if the F_i's differ the strategic equilibrium can only be obtained numerically. Further, if we consider auctions where multiple units are sold, either sequentially or simultaneously, finding the strategic equilibria is infeasible using the conventional techniques.

The numerical procedure introduced in this paper is general and can be used to approximate strategic equilibria in a large set of auctions. We hope that this particular application of well-known stochastic optimization methods will ultimately lead to breakthroughs in the analysis of auctions and similar complex games.

2 Preliminary Experiments

In preliminary experiments, we tried to infer a decision function $b_i(v_i)$ by estimating the probability $G_i(b)$ that the ith player will win using the bid b. The numerical estimate $\hat{G}_i$ is based on simulated auctions in which each bidder acts as if $\hat{G}_i$ was correct. This probability estimate is then updated using the result of the auction. The maximum likelihood estimate of $G_i(b)$ is simply the relative frequency of winning bids below b.

Two difficulties appeared with this approach. The first problem is that of *mass points*. Whenever $\hat{G}_i$ is not smooth, the selected bids will tend to focus on some particular points. To see this, suppose that the highest bid from all but i is always b^* then i will always bid a hair above b^* whenever $v_i > b^*$. Since this is true for all i, $\hat{G}_i$ will persist with a mass point around b^*. A way to avoid such mass points is to add some noise to the behavior: instead of bidding the (supposedly) optimal strategy, the bidder would bid some random point close to it. This problem is related to the famous *exploration vs exploitation* dilemma in reinforcement learning (Barto, 1992; Holland, 1975; Schaerf, Yoav and Tennenholtz, 1995).

Another difficulty is that we are not optimizing a single objective function but multiple ones (for each player), which interact. The players keep getting better so the optimization actually tries to track a *moving target*. Because of this, "old" observations are not as useful as recent ones. They are based on sub-optimal behavior from the other players. In preliminary experiments, we have found that this problem makes the algorithm very slow to approach a strategic equilibrium.

3 Proposed Approach

To address the above problems and extend the numerical solution to finding strategic equilibria in more complex games, we propose a new approach based on the following basic elements:

- Each player i is associated with a *stochastic decision model* that associates to each possible decision context C and strategy s_i, a probability distribution $P(a_i|C, s_i)$ over possible actions. A context C is an information available to a player before he chooses an action.

- The stochastic decision models are represented by flexible (e.g., non-parametric) models. For example, we used artificial neural networks computing $P(a_i|C, s_i)$ with parameters s_i.

- An on-line Monte-Carlo learning algorithm is used to estimate the parameters of these models, according to the following iterative procedure:

  1. At each iteration, simulate a game by (1) sampling a context from a distribution over decision contexts C, (2) sampling an action from the conditional decision models $P(a_i|C, s_i)$ of each player.

  2. Assuming the context C and the actions a_{-i} of the other players fixed, compute the expected utility $W_i(s_i|a_{-i}, C) = \int U_i(a_i|a_{-i}, C) dP(a_i|C, s_i)$, where $U_i(a_i|a_{-i}, C)$ is the utility of action a_i for player i when the others play a_{-i} in the context C.

3. Change s_i in the direction of the gradient $\frac{\partial W(s_i|a_{-i},C)}{\partial s_i}$.

Let us now sketch a justification for the proposed approach. At strategic equilibrium, the strategies of the players would be stationary, and convergence proofs of stochastic gradient descent would apply (Benveniste, Metivier and Priouret, 1990; Bottou, 1998). We do not have a proof in the general case before a strategic equilibrium is reached, i.e. in the non-stationary case. However, we have observed apparent local convergence in our experiments. When and if the stochastic (on-line) learning algorithm converges for all of the players, it means that the average gradients cancel out. For the ith player,

$$\frac{\partial E(W_i(s_i|a_{-i},C))}{\partial s_i} = 0 \tag{4}$$

where the expectation is over contexts C and over the distribution of decisions of the other players. Let s_i^* be the strategies that are obtained at convergence. From properties of stochastic (on-line) gradient descent, we conclude that at this point a local maximum of $E(W_i(s_i|a_{-i},C))$ with respect to s_i has then been reached for all the players. In the deterministic case ($P(a_i|C,s_i) = 1$ for some $a_i = a_i(s_i)$), the above expectation is simply the utility $U_i(a_1,\ldots,a_n)$. Therefore, a local strategic equilibrium has been reached (see eq. 1) (no local change in any player's strategy can improve his utility). If a global optimization procedure (rather then stochastic gradient descent) was used (which may however require much more computation time), then a global strategic equilibrium would be reached. In practice, we used a finite number of random restarts of the optimization procedure to reduce the potential problem of local maxima. The stochastic nature of the model as well as of the optimization method prevent mass-points, and we conjecture that the on-line learning ensures that each player's strategy tracks a locally optimal strategy (given the other players strategies).

Using a stochastic decision rule in which the dispersion of the decisions (the standard deviation of the bids, in our experiments) is learned appears in our experiments to yield to decreasing *exploration* and increasing *exploitation* as the players approach a local strategic equilibrium for a set of pure (deterministic) strategies. As the strategies of the players become stationary, this dispersion was found to converge to zero, i.e. a set of pure strategies. In other experiments not described here, our approach yielded a set of mixed (non-deterministic) strategies at the apparent strategic equilibrium (so both mixed and pure strategies can in general be the outcome, when a strategic equilibrium is approached).

To understand this phenomenon, let us consider what each player is implicitly maximizing when it chooses a strategy s_i by stochastic gradient descent at a given point during learning (i.e. maybe before a strategic equilibrium is reached, and a pure or mixed strategy may be chosen). It is the expectation over the other players actions of the expected utility $W(s_i|a_{-i}, C)$:

$$E_i = \int dP(a_{-i}) W(s_i|a_{-i}, C)$$

$$
\begin{aligned}
&= \int dP(a_{-i}) \int U(a_i|a_{-i}, C) dP(a_i|C, s_i) \\
&= \int dP(a_i|C, s_i) \int U(a_i|a_{-i}, C) dP(a_{-i}) \\
&= \int dP(a_i|C, s_i) u(a_i|C) \qquad (5)
\end{aligned}
$$

where we have simply switched the order of integration (and $U(a_i|a_{-i}, C)$ is the utility of action a_i when the other players play a_{-i}, in context C). If $P(a_{-i})$ is stationary, then the integral over a_{-i} is simply a function $u(a_i|C)$ of the action a_i. In that case, and if $u(a_i|C)$ has a single global maximum (which corresponds to a pure strategy), the distribution over actions which maximizes the expected utility E_i is the delta function centered on that maximum value, $\mathrm{argmax}_{\mathrm{a_i}} u(a_i|C)$, i.e., a deterministic strategy is obtained and there is no exploration. This happens at a strategic equilibrium because the other players actions are stationary (they have a fixed strategy).

On the contrary, if $P(a_{-i})$ is not stationary (i.e., the above integral changes as this distribution changes), then it is easy to show that a deterministic strategy can be very poor, which therefore requires the action distribution $P(a_i|C, s_i)$ to have some dispersion (i.e., there is exploration). Let us for instance take the simple case of an auction in which the highest bet $b^*(t)$ of the other players is steadily going up by Δ after each learning round t: $b^*(t) = b^*(t-1) + \Delta$. The optimal deterministic strategy always chooses to bid just above the previous estimate of b^*, e.g., $b(t) = b^*(t-1) + \epsilon$ where ϵ is very small. Unfortunately, since $\epsilon < \Delta$, this strategy *always loses.* On the other hand, if b was sampled from a normal distribution with a standard deviation σ comparable to or greater than Δ, a positive expected gain would occur. Of course, this is an extreme case, but it illustrates the point that a larger value of σ optimizes E_i better when there is much non-stationarity (e.g., Δ is large), whereas a value of σ close to zero becomes optimal as Δ approaches zero, i.e., the strategic equilibrium is approached.

The approach proposed in this paper builds upon a rich literature in stochastic learning and reinforcement learning. In the stochastic learning automata (SLA) of (Narendra and Thathachar, 1989), one generally considers a single context, whereas we focus on cases with multiple contexts (C can take several values). SLAs have usually a finite set of actions whereas we consider a continuous range of actions. SLAs are usually trained in the more general setting when only sample rewards can be used, whereas in the application to auctions the expected utility can be directly optimized. Gullapalli (Gullapalli, 1990) and Williams (Williams, 1992) also used a probability distribution for the actions. In (Gullapalli, 1990), the parameters (mean, standard deviation) of this distribution (a normal) were not trained using the expected utility. Instead a reinforcement algorithm was used to estimate the mean of the action and a heuristic (with hand-chosen parameters) is used to select standard deviation, i.e., obtain the exploration/exploitation tradeoff as learning progresses. In (Williams, 1992), the mean and variance are both optimized with a gradient descent algorithm, and as in our case, no proof of convergence was provided.

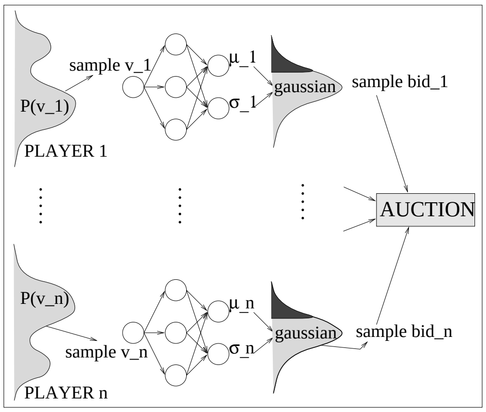

Figure 1: Illustration of the Monte-Carlo simulation procedure for an auction.

4 Applications to Auctions

In the experiments, the stochastic decision model for each bidder is a multi-layer neural network with a single input (the decision context C in this case is the valuation v_i), three hidden units[1], and two outputs, representing a truncated Normal distribution for b_i with parameters μ_i (mean) and σ_i (standard deviation). In the case of single-unit auctions, the normal distribution is truncated so that the bid b_i is in the interval $[0, v_i]$. The case of multi-units auctions is discussed in section 4.3. The Monte-Carlo simulation procedure is illustrated in Figure 1. The valuations v are sampled from the valuation distributions. Each player's stochastic decision model outputs a μ and a σ for its bid(s). Bids are sampled from these distributions, and ordered to determine the winner(s). Based on these observations, the expected conditional utility $W(s_i|a_{-i}, C)$ is

[1]different numbers were tried, without significant differences, as long as there are hidden units

computed: here it is the expectation of $v_i - b_i$ over values of b_i distributed according to the above defined truncated Normal. This integral can be computed analytically and its derivatives $\frac{\partial W(s_i|a_{-i},C)}{\partial s_i}$ with respect to the network parameters are used to update the strategies. It is interesting to note that, as explained in the previous section, in the experiments σ starts out large (mostly exploration of action space) and gradually converges to a small value (mostly exploitation), even if this behavior was not explicitly programmed.

In the following subsections, we will consider different types of valuation probability distributions F_i, as well as the single-unit and multi-units cases.

4.1 Symmetric Auctions with Known Solutions

We consider first a single-unit symmetric auction, i.e., there is only one good to sell and all players share the same probability distribution F over their valuations. As stated in the introduction, the (unique) strategic equilibrium is known analytically and is given by equation 3. In the experiments presented here, we tried two different valuation probability distributions: uniform U[0,1] and Poisson $F(v_i) = \exp(-\lambda \cdot (1 - v_i))$.

Table 1 summarizes results of experiments performed using the proposed method to find a strategic equilibrium for symmetric auctions with uniform valuation distribution. There were 8 players in these experiments. Since all players share the same probability distribution, we decided to share parameters of the 8 neural networks to ease the learning. We also tried with non-shared

	Symmetric Auction with Uniform Distribution			
	Avg. over 10 runs		Std. dev. over 10 runs	
	Avg. over 1000 bids	Std. dev.	Avg. over 1000 bids	Std. dev.
$\frac{bid}{bid_0}$	1.016	0.01	0.004	0.00003
$\frac{\mu}{bid_0}$	1.019	0.003	0.004	0.000004
σ	0.001	0.0001	0.000001	0.0

Table 1: Result of the symmetric auction with uniform distribution experiments.

parameters, and found almost the same results (but more learning iterations were required). Each experiment was repeated 10 times with different initial random conditions in order to verify the robustness of the method. After 10000 learning iterations (simulated auctions), we fixed the parameters and played 1000 auctions. We report mean and standard deviation statistics over these 1000 auctions and 10 runs. Let bid_0 be the bid that would be made according to the analytical solution of the strategic equilibrium. $\frac{bid}{bid_0}$ is the ratio between the actual bid and the analytical bid if the system was at strategic equilibrium. When this ratio is 1, it means that the solution found by the learning algorithm is identical to the analytical solution. It can be seen from the values of $\frac{\mu}{bid_0}$ at equilibrium that μ and the analytical bid are quite close. A small σ means the system has found a deterministic strategic equilibrium, which is consistent with the analytical solution, where bid_0 is a deterministic function of the valuation v.

Table 2 summarizes results of experiments done to find strategic equilibria for symmetric auctions with a Poisson ($\lambda = 7$) valuation distribution. Again,

	Symmetric Auction with Poisson Distribution			
	Avg. over 10 runs		Std. dev. over 10 runs	
	Avg. over 1000 bids	Std. dev.	Avg. over 1000 bids	Std. dev.
$\frac{bid}{bid_0}$	0.999	0.02	0.00004	0.0001
$\frac{\mu}{bid_0}$	0.999	0.02	0.00004	0.0004
σ	0.0002	0.00002	0.0	0.0

Table 2: Result of the symmetric auction with Poisson distribution experiments.

we can see that the system was always able to find the analytical solution.

4.2 Asymmetric Auctions

An auction is asymmetric when players may have a different probability distribution for their valuation of the goods. In this case, it is more difficult to analytically derive the solution for strategic equilibria. We thus developed an empirical method to test if the solution obtained by our method was indeed a strategic equilibrium. After learning, we fixed the parameters of all players except one. Then we let this player learn again for another 10000 auctions. This second learning phase was tried (1) with initial parameters starting at the point found after the first learning, or (2) starting with random parameters. In order to verify that the equilibrium found was not constrained by the capacity of the model, we also let the free player have more capacity (by doubling his hidden layer size). We tried this with all players. Table 3 summarizes these experiments. Since σ is small, the equilibrium solution corresponds to a deterministic decision function. Since the average gain of the free player is less than

	Asymmetric Auction with Poisson Distribution			
	Avg. over 10 runs		Std. dev. over 10 runs	
	Avg. over 1000 bids	Std. dev.	Avg. over 1000 bids	Std. dev.
σ	0.0016	0.00001	0.0	0.0
G_f	-0.0386	0.0285	0.0	0.0
G_r	-0.0385	0.0284	0.0	0.0
G_d	-0.0381	0.0254	0.0	0.0

Table 3: Result of the asymmetric auction with Poisson distribution experiments, number of players = 8, $\lambda = 7$ for first 4 players, $\lambda = 4$ for last 4 players. $G_{\{f,r,d\}}$ are the excess gain of the free player starting to learn from (f) a strategic equilibrium, (r) random point, and (d) random point with a double capacity model.

the average gains of the fixed players, we conclude that a strategic equilibrium had probably been reached (up to the precision in the model parameters that is allowed by the learning algorithm).

4.3 Multi-Units Auctions

In the multi-units auction, there are $m > 1$ identical units of a good to be sold simultaneously. Each player can put in his envelope multiple bids if he desires more than one unit. The m units are allocated to those submitting the m highest bids. Each winning buyer pays according to his winning bids. If a bidder i wins k units, he will pay $b_{i,1} + b_{i,2} + \cdots + b_{i,k}$ where $b_{i,1} \geq b_{i,2} \geq \cdots \geq b_{i,k}$. The rules are such that the price paid for the jth unit is no more than the price paid for the $(j-1)$th unit. Hence, $b_{i,j}$ is forced to lie in $[0, \min(v_{i,j}, b_{i,j-1})]$. In this case, no analytic solution is known. The same empirical method was therefore used to verify if a strategic equilibrium was reached. In this case the neural network

	Symmetric Multi-Units Auction with Poisson Distribution			
	Avg. over 10 runs		Std. dev. over 10 runs	
	Avg. over 1000 bids	Std. dev.	Avg. over 1000 bids	Std. dev.
σ of unit 1	0.001	0.0	0.0	0.0
σ of unit 2	0.002	0.0	0.0	0.0
σ of unit 3	0.460	0.0	0.058	0.0
σ of unit 4	0.463	0.0	0.02	0.0
G_f	-0.047	0.042	0.0001	0.0001
G_r	-0.048	0.042	0.0003	0.0
G_d	-0.057	0.034	0.0007	0.0001

Table 4: Result of the symmetric multi-units auction with Poisson(λ) distribution experiments, $\lambda = 7$, number of units = 4, number of players = 8. $G_{\{f,r,d\}}$ are the excess gain of free player starting to learn from (f) strategic equilibrium, (r) random point, and (d) random point with a double capacity model.

has 2 m outputs (μ and σ for each good). Table 4 summarizes the results. It appears that an equilibrium was reached (the free player could not beat the fixed players), but what is interesting to note is that σ for units 3 and 4 is very large. This may be because a player could probably bid anything for units 3 and 4 since he would probably not get more than 2 units at the equilibrium solution.

5 Conclusion

This paper presented an original application of artificial neural networks with on-line training to the problem of finding strategic equilibria in auctions. The proposed approach is based on the use of neural networks to represent a stochastic decision function, and takes advantage of the stochastic gradient descent to track a locally optimal decision function as all the players improve their strategy. Experimental results show that the analytical solutions are well approximated in cases when these are known, and that robust strategic equilibria are obtained in the cases where no analytical solution is known.

Interestingly, in the proposed approach, exploration is gradually reduced as the players converge towards a strategic equilibrium and the distribution of their actions becomes stationary. This is obtained by maximizing (by stochastic gradient descent) the expected utility of the strategy, rather then by fixing heuristically a schedule for reducing exploration.

Future work will extend these results to other (more complex) types of auctions involving sequences of decisions (such as multi-units sequential auctions). The approach could also be generalized in order to infer the valuation distribution of bidders whose bids are observed.

References

Barto, A. G. (1992). Connectionist learning for control: An overview. In Miller, W., Sutton, R., and Werbos, P., editors, *Neural Networks for Control*. MIT Press.

Benveniste, A., Metivier, M., and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, Berlin, New-York.

Bottou, L. (1998). Online algorithms and stochastic approximations. In Saad, D., editor, *Online Learning in Neural Networks*. Cambridge University Press, to appear, Cambridge, UK.

Dorsey, R., Johnson, J., and Van Boening, M. (1994). The use of artificial neural networks for estimation of decision surfaces in first price sealed bid auctions. In *New Directions in Computational Economics*, pages 19–39. Kluwer Academic Publishers.

Gullapalli, V. (1990). A stochastic reinforcement learning algorithm for learning real-valued functions. *Neural Networks*, 3:671–692.

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

McAfee, R. and McMillan, J. (1987). Auctions and bidding. *Journal of Economic Literature*, XXV:699–738.

McMillan, J. (1994). Selling spectrum rights. *Journal of Economic Perspectives*, 8:145–162.

Narendra, K. and Thathachar, M. (1989). *Learning Automata: an introduction*. Prentice Hall.

Schaerf, A., Yoav, S., and Tennenholtz, M. (1995). Adaptive load balancing: a study in multi-agent learning. *Journal of Artificial Intelligence Research*, 2:475–500.

Williams, R. (1992). Simple statistical gradient-following for connectionist reinforcement learning. *Machine Learning*, 8:229–256.