

Resource Allocation Games with Changing Resource Capacities

Aram Galstyan
USC Information Sciences
Institute
4676 Admiralty Way
Marina del Rey, CA 90292
galstyan@isi.edu

Shashikiran Kolar
USC Information Sciences
Institute
4676 Admiralty Way
Marina del Rey, CA 90292
skolar@isi.edu

Kristina Lerman
USC Information Sciences
Institute
4676 Admiralty Way
Marina del Rey, CA 90292
lerman@isi.edu

ABSTRACT

In this paper we study a class of resource allocation games which are inspired by the El Farol Bar problem. We consider a system of competitive agents that have to choose between several resources characterized by their time dependent capacities. The agents using a particular resource are rewarded if their number does not exceed the resource capacity, and punished otherwise. Agents use a set of strategies to decide what resource to choose, and use a simple reinforcement learning scheme to update the accuracy of strategies. A strategy in our model is simply a lookup table that suggests to an agent what resource to choose based on the actions of its neighbors at the previous time step. In other words, the agents form a social network whose connectivity controls the average number of neighbors with whom each agent interacts. This statement of the adaptive resource allocation problem allows us to fully parameterize it by a small set of numbers. We study the behavior of the system via numeric simulations of 100 to 5000 agents using one to ten resources. Our results indicate that for a certain range of parameters the system as a whole adapts effectively to the changing capacity levels and results in very little under- or over-utilization of the resources.

1. INTRODUCTION

The problem of coordination in multi-agent systems (MAS), where agents have to achieve a consensus in their actions to receive maximum reward, is an important problem that has attracted much interest recently [27, 23, 9, 13]. Reinforcement learning has been shown to be a general and robust method for achieving coordination in MAS, even when agents are not directly communicating or sharing information [24]. Game dynamics offers a rich foundation [10] for studying learning in multi-agent systems. In the game theory formalism, each agent is characterized by a set of strategies and it seeks to maximize its payoff (*i.e.*, utility or profit).

Game dynamics studies the behavior of agents in response to games that are played many times successively. Over the course of the games, the winning strategies are rewarded, losing ones are penalized, and the agents maximize their profit or utility by choosing the best performing strategies. It is the extra degree of freedom, characterized by the agents' strategies, that allows the system to adapt in dynamic environments. Game dynamics has a number of appealing properties as a control mechanism for multi-agent systems: it is distributed, flexible and scalable. The agents may vary in complexity from the very simple agents who do not have any information about other players or rules of the game, or even be aware of their existence, to more complex deliberative agents who can strategize and reason about their opponents beliefs and actions. The agents may act independently of one another, or jointly as in some cooperative agent systems [27, 9]; they may cooperate or act competitively. Game dynamics also has a long illustrious history in economics and mathematics [29]. It has been studied extensively both theoretically and in simulation in a wide range of domains and has been used to gain insight into the behavior of markets and resolve such conundrums as the emergence of cooperation among selfish agents [2]. Interest from other fields, such as statistical physics [6, 12], has added to the body of work in game dynamics literature.

In this paper we present a study of the minority games (MG) as a model for resource allocation/load balancing problem in a large scale MAS. In a minority game, the payoff each agent receives depends on its action as well as the number of other agents that chose the same action. Minority games are, therefore, a version of congestion games [18]. While MG have been intensively studied for single choice games [16], its application to multiple choice games with time-dependent capacities is novel.

We consider a system of N agents, that have to choose between Q available resources. Each resource is characterized by its capacity which is allowed to change with time, *i.e.*, $L_j = L_j(t)$, $j = 1, ..Q$. The agents in our system form a social network, and each agent's decision as to what resource to choose at time t is based on the actions of its neighbors at time $t - 1$. The efficiency of resource allocation is described by how closely the number of agent choosing a particular resource matches that resource's capacity. We are mainly interested in the case when the resources are moderately scarce, $\sum L_i \sim N$. Thus, an efficient system will have little under-utilization or over-utilization of the resources. We

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 publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

stress again that the agents do not have any explicit knowledge of the capacity levels of resources or the number and actions of other agents, besides the neighbors with whom it is directly communicating. Any coordination observed in the system arises from individual adaptations during the repeated games. Two important questions we are interested in are the following:

- Can the system adapt to changes in capacity levels?
- How effective is this adaptation?

Framing the multi-agent learning in the terms of minority games on networks allows us to fully parameterize the problem by a small set of numbers. These parameters are the size of the system N , number of resources Q , connectivity of the network K , and two numbers that describe the complexity of the agent’s decision: the number of strategies the agent uses and a strategy bias (described below in more detail). We show that for some parameters the system is extremely efficient and adapts quickly to capacity changes. Surprisingly, as the number of neighbors an agent interacts with increases, the overall behavior of the systems is degraded, indicating that in this case limited local communication is preferable over global communication.

The rest of the paper is organized as follows: In Section 2 we introduce Minority Games and discuss previous work in the field. In Section 3 we describe our past work on minority games with time dependent capacities, where we introduced local communication among agents based on random NK networks. We present the extension of this work to multi-choice games in Section 4 and results of numeric simulations in Section 5.

2. EL FAROL BAR PROBLEM AND MINORITY GAMES

Since its introduction in 1994, Arthur’s El Farol Bar problem [1] has been one of the most widely studied examples of complex adaptive systems. The model consist of N individuals who have to decide independently whether to attend the El Farol bar in Santa Fe on a given night. The bar has a limited capacity, and people try to avoid attending it when it is overcrowded. There is no explicit communication between individuals, and the only information available to them is the time series of past attendance numbers. Since no deductively rational solution is possible, Arthur suggested to use inductive reasoning instead: Each agent has a set of “predictors” (strategies) that predict next week’s attendance given the history of past attendance. Agents keep track of the performance of their predictors, and reinforce them according to their reliability. Numerical simulations of this simple model showed that the system self-organizes so that attendance fluctuates around the bar capacity.

The Minority Game [6] (MG) was introduced by Challet and Zhang as a simplified version of the El Farol Bar problem, the main difference being that instead of the actual history the agents are provided only with a binary string indicating whether the bar was overcrowded or not. More precisely, let us consider N agents with bounded rationality that repeatedly choose between two alternatives labelled 0 and 1 (*e.g.*, staying at home or going to the bar). If at a given time step the bar was undercrowded (overcrowded) then the winning choice (signal) is 1 (0). In the original model, the number of agent is taken to be odd, and the

capacity was fixed to $(N - 1)/2$, so that the agents who made the minority decision won (hence the name Minority Game). In the Generalized Minority Game [12], the winning group is 1 (0) if the fraction of the agents who chose “1” is smaller (greater) than the capacity level η , $0 < \eta < 1$. As in the bar problem, each agent uses a set of S strategies to decide its next move and reinforces strategies that would have predicted the winning group. The main advantage of the MG model is that strategies can be easily parameterized: a strategy is simply a lookup table that prescribes a binary output for all possible inputs, where the input is a binary string containing the last m outcomes of the game. Thus, for each choice of m , there are $P = 2^m$ possible histories (inputs), and $\Omega = 2^P$ strategies. Note that in this model the agents interact by sharing the same global signal. Despite its simplicity MG has been demonstrated to have a very rich and complex dynamics. The most interesting phenomenon of the minority model is the emergence of a coordinated phase, where the standard deviation of attendance, the volatility, becomes smaller than in the random choice game, where each agents makes either choice with probability $1/2$.¹ Coordination is achieved for memory sizes for which the dimension of the reduced strategy space is comparable to the number of agents in the system, $2^m \sim N$ [7, 22]. It was later pointed out that the dynamics of the game remains mostly unchanged if one replaces the string with the actual histories with a random one [4], provided that all the agents act on the same signal. Analytical studies based on this simplification has revealed many interesting properties of the minority model[5, 16].

3. MINORITY GAMES WITH CHANGING CAPACITY ON NETWORKS

In a previous study [11] we demonstrated that if one introduces time dependent capacities to the MG model defined in the previous section, the system does not adapt well. We also showed that one can achieve adaptation if instead of the interaction via a global signal agents are allowed to interact locally. Below we illustrate our model and recapitulate main results.

Our model consist of a large number of simple autonomous agents that form a social network. Instead of interacting via a global signal of histories, the agents interact locally: each agent gets an input from K neighbors (that are randomly chosen) and maps the input to a new state that prescribes the decision (*e.g.*, attend the bar or not):

$$s_i(t + 1) = F_i^j(s_{k_1}(t), s_{k_2}(t), \dots, s_{k_K}(t)) \quad (1)$$

where $s_{k_i}(t)$, $i = 1, \dots, K$ are the choices made by its neighbors during the previous time step, and F_i^j , $j = 1, \dots, S$ are randomly chosen boolean functions (called strategies hereafter) used by the i -th agent. Both neighbors and strategies are chosen randomly and quenched throughout the game. If $S = 1$ such a network is the well known NK or Kauffman network [14]. As in the traditional MG, the agent keeps a score for each strategy F_i^j that monitors the performance of that strategy, adding (subtracting) a point if the strategy predicted the winning (losing) choice. We let

¹In the random choice game the average number of agents choosing “1” is $(N-1)/2$ with standard deviation $\sigma = \sqrt{N}/2$ in the limit of large N .

the “attendance” $A(t)$ be the cumulative output of the system at time t , $A(t) = \sum_{i=1}^N s_i(t)$. Then the winning choice is “1” if $A(t) \leq N\eta(t)$, and “0” otherwise. Those in the winning group are awarded a point while the others lose one. Agents play the strategies that have predicted the winning side most often, with ties are broken randomly.

As a measure of efficiency we introduce $\delta(t) = A(t) - N\eta(t)$, that describes the deviation from the most optimal resource utilization. We are primarily interested in the cumulative “waste” over a certain time window:

$$\sigma^2 = \frac{1}{T_0} \sum_{t=t_0}^{t_0+T_0} \delta(t)^2 \quad (2)$$

We can compare the performance of our system to a default random choice game, defined as follows: assume that the agents can query the capacity $\eta(t)$ at a given time step, and they choose to go to the bar with probability proportional to $\eta(t)$. In this case the main attendance is close to $\eta(t)N$ at each time step, and the fluctuations around the mean are given by the standard deviation

$$\sigma_0^2 = N \frac{1}{T} \int_{T_0}^{T_0+T} dt' \eta(t') [1 - \eta(t')] \quad (3)$$

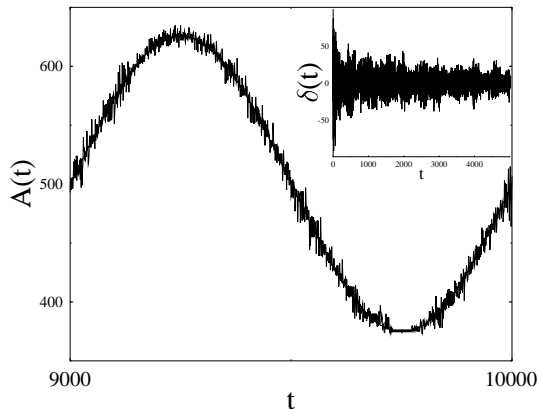


Figure 1: A segment of the attendance time series for $K = 2$, $\eta(t) = 0.5 + 0.12\sin(2\pi t/T)$, $T = 1000$.

The results of our extensive numerical simulation indicate that the networks with $K = 2$ achieve very effective utilization of resources, even when the changes in the capacity level are relatively large. In Fig. 1 we plot the time series of the attendance for a sinusoidal change in capacity. One can see that the system follows the changes in capacity level very effectively. The inset of Fig. 1 shows the time series of the deviation $\delta(t)$ for $K = 2$. Initially there are strong fluctuations, hence poor utilization of the resource, but after some transient time the system as a whole adapts and the strength of fluctuations decreases. In fact, the standard deviation of the fluctuations is considerably smaller than in the random choice game as defined by Eq. (3).

In Fig. 2 we plot the variance per agent versus network connectivity K , for system sizes $N = 100, 500, 1000$. For each K we performed 32 runs and averaged results. Our simulations suggest that the details of this dependence are

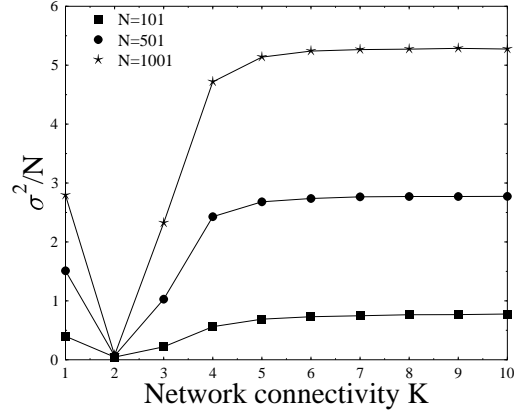


Figure 2: σ^2/N vs the network connectivity for different system sizes

not very sensitive to the particular form of the perturbation $\eta_1(t)$, and the general picture is the same for a wide range of functions, provided that they are smooth enough. The variance attains its minimum for $K = 2$ independent of the number of agents in the system. Note that this is different from the traditional minority game, where the position of the minimum scales logarithmically with N . For bigger K it saturates at a value that depends on the amplitude of the perturbation and on the number of agents in the system. We found that for large K the time series of the attendance closely resembles the time series in the absence of perturbation. This implies that for large K the agents do not “feel” the change in the capacity level. Consequently, the standard deviation increases linearly with the number of agents in the system, $\sigma \propto N$. For $K = 2$, on the other hand, the scaling has the form $\sigma \propto N^\gamma$, where $\gamma < 1$. Our results indicate that the value of the scaling exponent γ is not universal, and depends on the perturbation η .

4. MULTI-CHOICE GAMES

In the previous section we considered a situation where agents face a binary decision whether to use a resource or not (an alternative interpretation is that agents are choosing between two resources the capacities of which sum up to N). In most of the practical situations, however, the number of choices might be greater. In this section we extend our model to account for the multi-resource scenario. Namely, agents have to choose one of Q available resources, characterized by their capacities $L_i(t), i = 1, \dots, Q$. To study this scenario in the context of previous sections we use multi-state Kauffman networks [26] to model inter-agent interactions. The state of agents are given by $s_i = \{0, 1, \dots, Q - 1\}$, $i = 1, \dots, N$; therefore, state corresponds to the choice an agent makes. The dynamics is specified analogously to the binary game, *i.e.*, each agent receives its inputs from K other agents and maps it to one of the Q available states.

$$s_i(t+1) = \Lambda_i^j(s_{k_1}(t), s_{k_2}(t), \dots, s_{k_K}(t)) \quad (4)$$

As before, each agent have S functions (strategies), and keep a virtual score for each of its strategies and plays the strategy with the highest number of accumulated points.

| $S_{k_1}(t)$ | $S_{k_2}(t)$ | $S_i(t+1)$ |
|--------------|--------------|------------|
| 0 | 0 | 0 |
| 0 | 1 | 2 |
| 0 | 2 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |
| 1 | 2 | 0 |
| 2 | 0 | 2 |
| 2 | 1 | 0 |
| 2 | 2 | 2 |

Figure 3: Example of a strategy for $K = 2$

Let $A_i(t)$ be the number of agents who chose i th resource at time t . Then these agents will be rewarded if $A_i(t) \leq L_i(t)$ and punished otherwise. At the start of the games, every agent randomly chooses K neighbors and a set of S strategies, which are fixed throughout the game. As in the binary game, the strategies can be represented as lookup tables, that for each of the Q^K possible inputs assign an output (action) from the set $\{0, 1, \dots, Q-1\}$ (see Fig. 3). Strategy bias P is the parameter that determines the relative homogeneity of the output column of the strategy table. The entries in the output column are chosen as follows: first, with probability $1/Q$ one of the resources is chosen, let us say Q_k . Then, in the output column we choose i) Q_k with probability P ii) $Q_i, i \neq k$ with probability $(1-P)/(Q-1)$. Thus, for a binary choice game ($Q = 2$), for $P = 0.5$ there are (in average) equal numbers of '0's and '1's in the output column; while for $P = 1.0$, the entries are all '0's or '1's, whichever symbol has been picked randomly.

As before, the baseline solution we compare against is the random choice game where agents choose a particular resource with probability $\eta_i(t) = L_i(t)/N$. If the capacities are constant, then the standard deviation of each A_i that characterize the waste of resource is $\sigma_i^0 = N\eta_i(1-\eta_i)$ (the upper index 0 stays for the random choice game). In the case of changing capacities, one has to take an integral over the time window for which waste is calculated as in Eq. 3. In the results presented in the next section we will primarily consider the waste averaged over the resources, i.e.,

$$\sigma_{tot}^2 = \frac{1}{Q} \sum_{i=1}^Q \sigma_i^2 \quad (5)$$

5. RESULTS

We performed extensive numerical simulations for the system described above, with the number of agents ranging from 100 to 5000, and number of choices $Q = 3$ up to $Q = 10$. The number of possible inputs for a fixed K and Q is Q^K , and the number of possible strategies is Q^{Q^K} . Hence, we had to restrict ourselves to small values of K . It is known, on the other hand, that the dynamical properties of multi-state Kauffman networks with given connectivity K can be regulated by tuning the function (strategy) bias P [26]. So instead of studying the properties of the system for various K , we set $K = 2$ and vary P instead. This approach is computationally very efficient and scalable with respect to the number of resources Q and allows us to study system sizes up to $N = 5000$.

First, we examine the resource allocation problem with fixed capacities. The time series of the resource usage for $Q = 3$ choices, and capacities $\eta_0 = 0.5, \eta_1 = 0.3, \eta_2 = 0.2$ is

shown in Fig. 4(a). After a short transient time, the number of agents $A_i(t)$ using the resource i start to fluctuate around the fixed capacity level. The strength of these fluctuations determines the global waste. In Fig. 4(b) we show the dependence of waste (averaged over the resources and normalized to σ_{tot}^0) on the strategy bias. The horizontal line shows the value of σ_0 for the random choice game. Remarkably, for some values of P the efficiency of resource allocation is an order of magnitude better than in the random choice game.

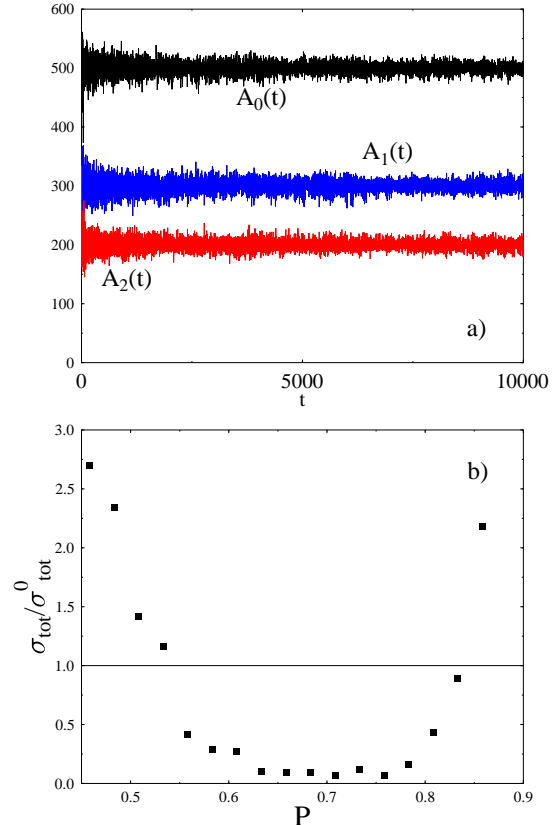


Figure 4: a) Time series for the resource usage for $Q = 3$ resources with capacities $\eta_0 = 0.5, \eta_1 = 0.3, \eta_2 = 0.2$. The simulation parameters are $N = 1000, S = 2, K = 2$, and $P = 0.6$. b) Normalized variance versus the strategy bias P

Now we turn to the case of time-varying capacities. In Fig. 5 we show a typical segment of a time series of resource usage for $Q = 3$ and periodic perturbation of form $\eta_i(t) = 1/3 + \delta\eta_i \sin(2\pi t/T)$, $i = 0, 1, 2$, and with amplitudes $\delta\eta_0 = 1/6, \delta\eta_1 = \delta\eta_2 = -1/12$. We plot the time series for only two choices since the third one is fully determined by the first two. It can be seen from the figure that agents follow the changes in the capacity levels very effectively, even when the capacities change by as much as 50%.

In Fig. 6 we plot standard deviation vs the strategy bias P for $Q = 3, 5, 10$ and number of strategies $S = 2, 3, 5$. Again, an average over the choices has been taken. For each set of parameters we did eight trials, and averaged the results. The duration of simulations is $T = 10^4$ steps, and

we used the data for the last 20% of the time series to calculate the waste. For each Q , the capacities are given by $\eta_i(t) = \eta_Q + \delta\eta_i \sin(2\pi t/T)$, $i = 1, \dots, Q$, where $\eta_Q = 1/Q$ and the amplitudes of the oscillations $\delta\eta_i$ $i = 1, \dots, Q$ were chosen as high as 60% of η_Q . Again, we have normalized the cumulative waste by its value for the random choice game (horizontal line). One can see that for the certain range of the bias the efficiency of the system is better than in the random choice game, and there is a well pronounced minimum in σ_{tot} . For $Q = 3$, the parameter range for which the system is the most efficient (by an order of magnitude when compared to the random choice game) is $P \sim 0.65 \div 0.75$. For $Q = 5$ and $Q = 10$, this region shifts to the higher values of P and gets wider at the same time. Another interesting observation is that for all three cases, the minimum is the deepest for $S = 2$, *i.e.*, two strategies per agent. One of the factors contributing to this is that it takes longer time for the system to achieve the coordinated phase as one increases S , and since we have used the same duration of simulation for each S , this would result in a slightly greater value of σ_{tot} . However, we have verified that even if the simulation are carried out for long enough times, the system with $S = 2$ performs slightly better. Although we do not have a sound explanation for this yet, we believe that this might be due to the fact that large number of strategies allows an agent to switch between resource much more frequently, hence contributing more to the fluctuations.

Finally, we also studied how the standard deviation scales with the number of agents in the system. As in the binary game, we found that there is a scaling of the form $\sigma \propto N^\gamma$, which is however not universal and depends on the amplitude of the perturbation, as well as the number of choices and strategies in play. Still, our results indicate that the standard deviation per agent grows sub-linearly with N , which is a result of coordination.

6. RELATED WORK

Because of the many parallels between the behavior of markets and resource allocation (and the related problem of task allocation) in multi-agent systems, this task was seen as an ideal candidate for a market-based solution. In fact, since Reid Smith proposed his Contract Net Protocol [25], economically-inspired mechanisms for coordination in multi-agent systems have been widely studied and applied to a number of agent-related problem domains [32]. Such mechanisms are attractive because they are inherently distributed, flexible, and in many cases quite scalable. Moreover, they are backed by a rich body of economic theory, so that rigorous results exists for many problems. Of the economically-based control schemes, the one that has received by far the most attention in the multi-agent community was auctions. Algorithms for holding and deciding auctions have been proposed and analyzed [21, 19], both from the perspective of resource allocation [3, 20] and task allocation [30], and applied to task domains such as allocating operating system resources [15], providing library services [17], scheduling jobs on networks of workstations [8], and maintaining optimal office climate [33]. The most substantial difference between auctions and our game dynamics approach is that auction require a central authority, an auctioneer, to determine winners. In our scheme, the payoff an agent receives is local, based on the load on the chosen resource only. Another difference is where the intelligence resides. In auction schemes,

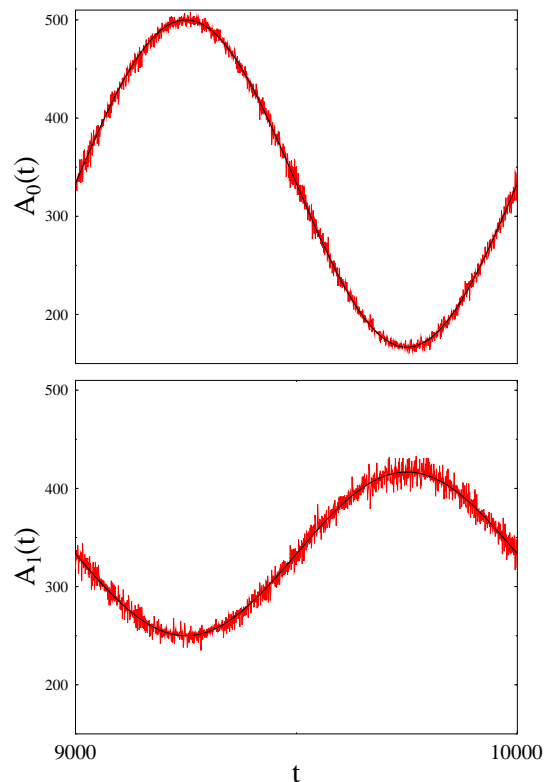


Figure 5: a) Time series for the resource usage for $Q = 3$ resources with time dependent capacities $\eta_i(t) = 1/3 + \delta\eta_i \sin(2\pi t/T)$, $i = 0, 1, 2$, and with amplitudes $\delta\eta_0 = 1/6$, $\delta\eta_1 = \delta\eta_2 = -1/12$. The simulation is done for $N = 1000$, $S = 2$, $K = 2$, and $P = 0.68$.

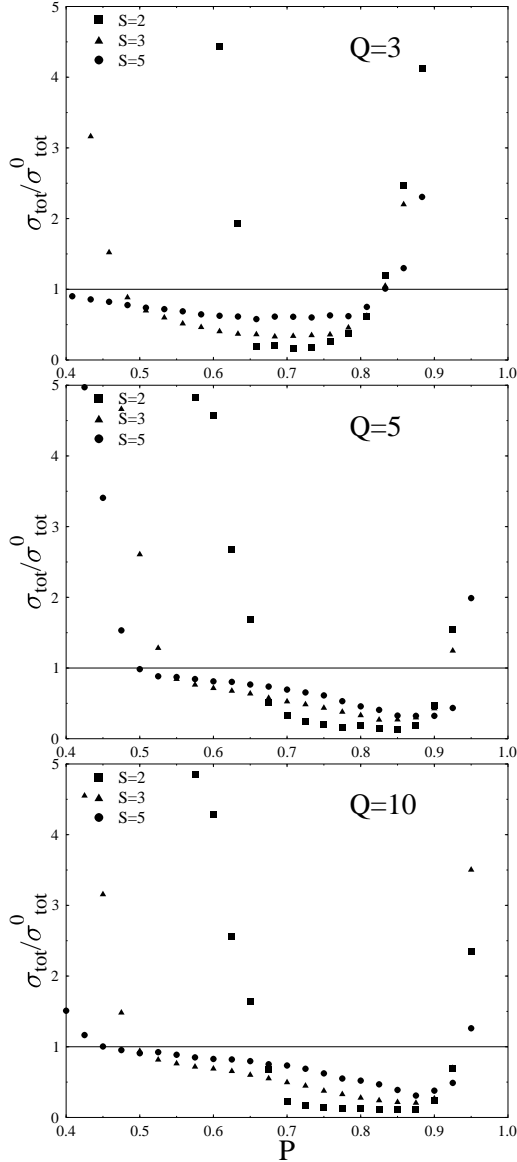


Figure 6: Normalized variance versus the strategy bias P for $Q = 3, 5, 10$, and $N = 1000$

the agents must decide how to bid taking many factors into account. In games, it is the payoff mechanism that contains the brains of the system.

Reinforcement learning through game dynamics in MAS has attracted much attention recently. In particular, learning to coordinate has been widely studied [24, 9, 13]. These approaches are generally based on the Q -learning framework, and the main object of research is to demonstrate convergence to an optimal or equilibrium solution. Convergence is no longer an issue in a dynamic setting where the functions the agents are trying to learn is changing over time [28].

Shaerf *et al.* [23] studied the problem of adaptive load balancing as an illustration of reinforcement learning in multi-agent systems. Their statement of the problem is quite similar to ours. They studied a system of N agents using M resources, each having a time dependent capacity C . If it is idle, an agent may submit a new job of some size to a resource. The resource is chosen using the resource selection rule SR. The resource selection rules are purely local, *i.e.*, have access only to the experience of a particular agent, and are the same for all agents. Schaerf *et al.*'s approach is similar to Q -learning [31] in that SRs embed within themselves an efficiency estimator that keeps track of how the resources perform over time, and a policy for selecting a resource based on this efficiency estimator. Schaerf *et al.* studied a relatively small system of 100 agents numerically for different parameter regimes and observed generally good adaptive behavior. They found that introducing communication so that agents can share information about performance of resources was detrimental to system performance. Although their work is similar to ours, it differs in important aspects. Our learning algorithm is not probabilistic, and therefore, does not allow for exploration in the reinforcement learning sense of the word. Rather, our system is constantly exploiting the best performing strategies. The second major difference is that the agents in our system are intrinsically heterogeneous, in a sense that each agent has different set of strategies. We also study a much larger system. Our metrics are also different. We find that the system is *adaptive* (able to track resource capacity changes, on average) for a wide range of parameters, it is *efficient*, meaning that fluctuations in resource utilization are small, only for a small range of parameter values. Another difference between our work and theirs is that we parameterize the degree of information sharing between agents, and find that for some values local communication improves system performance. We attribute disagreement between our findings to the different forms of the learning rules in our system and theirs.

7. DISCUSSION

We have presented a reinforcement learning model for adaptive resource allocation in a multi-agent system. Our model is applicable to large systems where agents have a choice of several resources with time-dependent capacities. The learning scheme is based on minority games on networks. Each agent uses a set of strategies to decide what resource to choose. The input to the strategy (a lookup table) is a string containing the actions of the agent's neighbors during the previous time step. However, there is no information in the string about what the winning choices were. A strategy is rewarded if it led an agent to choose a resource whose utilization is at most equal to its capacity at

that time; otherwise, the strategy is punished. The agents learn over time the best performing strategies and use them to select what resource to use.

The problem is parameterized by the number of agents, number of resources, the degree of communication in the system (specified by network connectivity), and the number and uniformity of strategies. We studied the system numerically for a wide range of parameters and system sizes, and found that some parameters lead to a very adaptive and efficient system.

There is a number of interesting questions left unanswered by our work. It would be interesting, for example, to introduce Q -learning to check whether allowing exploration in addition to exploitation will lead to a better performance. Another direction is to use information theoretic measures to replace the number of strategies and strategy bias with a single parameter that captures the complexity of an agent. We have some evidence that systems of less complex agents perform better, and that over time the agents tend to become simpler, but we have not studied this in detail yet. Still another interesting question is how information spreads through the system and how we may characterize it. Finally, we believe our approach is robust and general, and that it is worthwhile to apply it to other MAS problem domains by crafting the payoff function to the characteristics of the problem.

8. ACKNOWLEDGEMENTS

The research reported here was supported by the Defense Advanced Research Projects Agency (DARPA) under contracts number F30602-00-2-0573. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of any of the above organizations or any person connected with them.

9. REFERENCES

- [1] W. B. Arthur. Inductive reasoning and bounded rationality. *American Economic Review*, 84:406–411, 1994.
- [2] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211:1390–1396, 1981.
- [3] C. Boutilier, M. Goldszmidt, and B. Sabata. Sequential Auctions for the Allocation of Resources with Complementarities. In *International Joint Conference on Artificial Intelligence (IJCAI)*, Stockholm, Sweden, Aug. 1999.
- [4] A. Cavagna. Irrelevance of memory in the minority game. *Phys. Rev. E*, 59:R3783, 1999.
- [5] D. Challet and M. Marsili. Phase transition and symmetry breaking in the minority game. *Phys. Rev. E*, 60:R6271, 1999.
- [6] D. Challet and Y.-C. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A*, page 407, 1997.
- [7] D. Challet and Y.-C. Zhang. On the minority game: Analytical and numerical studies. *Physica A*, 256:514, 1998.
- [8] A. Chavez, A. Moukas, and P. Maes. *Challenger: A Multi-agent System for Distributed Resource Allocation*. In *Proc. of Autonomous Agents*, Marina del Rey, CA, Feb. 1997.
- [9] C. Claus and C. Boutilier. The dynamics of reinforcement learning in cooperative multiagent systems. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, pages 746–752, July 1998.
- [10] D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1998.
- [11] A. Galstyan and K. Lerman. Adaptive boolean networks and minority games with time-dependent capacities. *Physical Review*, E66:015103, 2002.
- [12] N. Johnson, P. Hui, D. Zheng, and C. Tai. Minority game with arbitrary cutoffs. *Physica A*, 269:493, 1999.
- [13] S. Kapetanakis and D. Kudenko. Reinforcement learning of coordination in cooperative multi-agent systems. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-02)*, Edmonton, Alberta, Canada, July 2002.
- [14] S. A. Kauffman. *The Origins of Order*. Oxford University Press, New York, 1993.
- [15] J. F. Kurose and R. Simha. A microeconomic approach to optimal resource allocation in distributed computer systems. *IEEE Transactions on Computers*, 38(5):705–717, 1989.
- [16] See <http://www.unifr.ch/econophysics/minority/> for an extensive collection of articles and references.
- [17] T. Mullen and M. Wellman. Market-Based Negotiation for Digital Library Services. In *Proc. of the 2nd USENIX Workshop on Electronic Commerce*, Oakland, CA, Nov. 1996.
- [18] R. W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.
- [19] T. Sandholm. Limitations of the Vickrey Auction in Computational Multiagent Systems. In *International Conference on Multi-Agent Systems (ICMAS)*, pages 299–306, Kyoto, Japan, Dec. 1996.
- [20] T. Sandholm. An Algorithm for Optimal Winner Determination in Combinatorial Auctions. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 542–547, Stockholm, Sweden, Aug. 1999.
- [21] T. Sandholm and S. Suri. Market Clearability. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1145–1151, Seattle, WA, 2001.
- [22] R. Savit, R. Manuca, and R. Riolo. Adaptive competition, market efficiency, phase transition. *Phys. Rev. Lett.*, 82(10):2203, 1999.
- [23] A. Schaerf, Y. Shoham, and M. Tennenholtz. Adaptive load balancing: A study in multi-agent learning. *Journal of Artificial Intelligence Research*, 2:475–500, 1995.
- [24] S. Sen, M. Sekaran, and J. Hale. Learning to coordinate without sharing information. In *(American) National Conference on Artificial Intelligence*, pages 426–431, Menlo Park, CA, 1994. AAAI Press.
- [25] R. G. Smith. The Contract Net Protocol. *IEEE Transactions on Computers*, 29(12):1104–1113, Dec. 1980.

- [26] R. V. Sole, B. Luque, and S. A. Kauffman. Phase Transitions in Random Networks with Multiple States. *SFI Working Papers*, 00-02-011, 2000.
- [27] M. Tan. Multi-Agents Reinforcement Learning: Independent vs Cooperative agents. In *Proceeding of the 10th International Conference on Machine Learning (ICML-93)*, 1993.
- [28] J. M. Vidal and E. H. Durfee. The moving target function problem in multi-agent learning. In *Proceedings of the 3rd International Conference on Multi-Agent Systems (ICMAS-98)*, 1998.
- [29] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1944.
- [30] W. E. Walsh and M. P. Wellman. A Market Protocol for Decentralized Task Allocation. In *International Conference on Multi-Agent Systems (ICMAS)*, Paris, France, July 1998.
- [31] C. J. C. H. Watkins. *Learning from Delayed Rewards*. PhD thesis, Cambridge University, Cambridge, England, 1989.
- [32] M. P. Wellman. Market-Oriented Programming: Some Early Lessons. In S. H. Clearwater, editor, *Market-Based Control: A Paradigm for Distributed Resource Allocation*, pages 74–95. World Scientific, Jan. 1996.
- [33] F. Ygge and H. Akkermans. Decentralized Markets versus Central Control: A Comparative Study. *Journal of Artificial Intelligence Research*, 11:301–333, 1999.