Reinforcement Learning (RL) Based Collision Avoidance Approach for Multiple Autonomous Robotic Systems (ARS)

O. AZOUAOUI*, M. OUAAZ* and A. FARAH**

*CDTA – Centre de Développement des Technologies Avancées, Laboratoire de Robotique et d'Intelligence Artificielle,
128, Chemin Mohamed Gacem, BP 245 El-Madania, 16075, Algiers, Algeria
azouaoui@hotmail.com
**ENP – Ecole Nationale Polytechnique, Laboratoire Techniques Digitales et Systèmes,
10, Avenue Hassen Badi El-Harrach, Algiers, Algeria
farah@hotmail.com

Abstract

In several complex applications, the use of multiple Autonomous Robotic Systems (ARS) becomes necessary to achieve different tasks such as foraging and transport of heavy and large objects with less cost and more efficiency. They have to achieve a high level of flexibility, adaptability and efficiency in real environments. Therefore, they must particularly have the capability to avoid collisions among them and with obstacles. In this paper, a Reinforcement Learning (RL) based collision avoidance approach for multiple ARS is suggested. Indeed, each robot must learn how to avoid the others from its interaction with the environment while reaching its target. This learning process allows ARS to benefit from the experience of the others simply by having the same score base. This approach must provide ARS with capability to acquire the collision avoidance behavior among several ARS from elementary behaviors only by trial and error search. Then, simulation results display the ability of the suggested approach to intelligently avoid collisions adaptively among ARS and with obstacles. Such an approach is capable to provide these ARS with real-time processing, more autonomy and intelligence.

Keywords: Autonomous Robotic Systems (ARS), Navigation, Collision Avoidance Behavior, Adaptive Behavior, Elementary behaviors, Reinforcement Learning (RL).

1 Introduction

With increasing demands for high precision autonomous control to achieve cooperative work by multiple Autonomous Robotic Systems (ARS), conventional control approaches are unable to adequately deal with system complexity, nonlinearities and uncertainty. Intelligent control that is experiential based rather than model based is designed as a new emerging discipline to overcome these problems [13, 3, 9]. This kind of discipline is necessary for robot control in several development of real-time robotic applications [4, 10] particularly the problem of collision avoidance among several robots. The ARS endowed with this behavior have the ability to move and be self-sufficient in multi-robot environments. These ARS must be capable of task and situation oriented behavior if they are to react usefully to their environment. An a priori modeling of all possible reactions to particular events is in most cases not possible. For this reason, the development of control systems for ARS has led to the development of adaptive systems. These systems react to changes in their environment, learn from errors in behavior and can solve some unforeseen situation classes independently [13, 5]. In fact, most of the research conducted today for learning in autonomous robots deals with the behavior-based paradigm [9]. This bottom-up approach concentrates on physical systems situated in the world and promotes simple associative learning between sensing and acting [14]. This learning in mobile robotics is aimed at avoiding the need (for the human operator) to model all of the complexities, interactions, or other influences in the real world [13, 19]. Among these adaptive approaches currently available, Reinforcement Learning (RL) is one of the most investigated approaches [13, 20, 19].

This paper deals with the *behavior-based robotics* and *intelligent control* of ARS in multi-robot environments. The aim of this work is to suggest an adaptive collision avoidance approach for multiple ARS capable to provide these robots with *real-time* processing, *more autonomy* and *intelligence*. Therefore, a reinforcement learning collision avoidance based approach for multiple ARS is suggested as one of the fundamental functions of the robots. This approach uses an adaptive method for acquisition of the elementary

behaviors to avoid collision with other robots and obstacles. To acquire the adaptive behavior, the RL is introduced. It is shown that the appropriate behaviors for collision avoidance can be successfully acquired through the suggested learning process. Because RL is concerned with the adaptive control of a robot through the use of scalar rewards (for feedback) and direct trial-and-error interaction with the environment [14, 19], it is useful for many robot problems. Indeed, the system based on RL improves its performance by receiving feedback in the form of a scalar reward (or penalty) that is commensurate with the appropriateness of its response. Thus, this behavior-based approach must provide robots with capability, to avoid collisions by interaction with the environment. In this paper, current obstacle avoidance approaches based on RL which remedy insufficiencies of classical approaches are discussed in Section 2. A RL approach essentially based on the robot interaction with the environment to acquire the obstacle avoidance behavior is suggested in Section 3. This approach uses elementary behaviors to learn how to behave in a multi-robot environment. Section 4 summarizes the simulation results.

2 Reinforcement Learning Based Approaches

Designing robots that learn by themselves to perform complex real world tasks is still an open challenge for the field of robotics and artificial intelligence [13, 9, 20]. Indeed, several researchers have developed approaches that are based on learning from interaction with the environment and experience. Such type of learning is pertinent to intelligent control since it leads to systems that do not depend upon a priori knowledge for decision-making. A RL as a pertinent type of this learning maintains an explicit policy function that maps situations directly into actions instead of using an explicit domain model to generate a sequence of actions which is then executed open loop as in classical planning [14, 16, 20]. Indeed, approaches based on RL allow a quick response to unexpected contingencies and opportunities leading to *situated* and *reactive* systems, which are of a great interest in robotics, and particularly in multi-robot systems. *Situatedness* and *reactiveness* are in fact two important properties of systems using RL. Indeed, robots are situated since the robots themselves must control the whole interaction with the environment, i.e., the world must always be seen from the perspective of the robots [12]. In other words, the robots are situated if their control decision is based on the current situation (as determined by sensor readings and possibly a limited amount of internal state). Consequently, the robots have to be able to bring in their own experience in dealing with the current situation [12, 20]. While robots are reactive if they generate actions (behaviors) at a rate that is commensurate with the dynamics of the environment in which they are embedded. For such robots, decision making consists of evaluating a policy function, which typically requires a small constant amount of time [20]. These robots characterized also by their incremental learning adapt their policies based on experience accumulated over time. They gradually reach the correct answer through successive approximations even if their models are incomplete or inaccurate leading then to robust robots. This robustness also implies that these models can be learned during the process of RL, which allows then their use, by the robots themselves [14, 20].

Several works focused on control of a single robot. Indeed, some of them are interested in improving a collision avoidance problem based on RL through the use of neural networks [17, 18] or by providing a method to segment the sensor space [11], or by introducing several algorithms using immediate reward and delayed reinforcement [8].

In the contrary the use of RL in the collision avoidance problem for many robots is still an opening problem. For instance, an adaptive acquisition for collision avoidance among multiple autonomous mobile robots which are equipped with 'LOcally Communicable Infrared Sensory System (LOCISS)' is developed in [1]. This approach is based on RL scheme where a selected behavior is executed and is evaluated based on the displacement of three distances. The learning process should be executed to acquire the collision avoidance behaviors for a specific situation. However, considering the implementation of this approach onto the real robot constitutes a problem. The number of possible situations becomes extremely large beyond the capacity of memories which can be mounted on an actual mobile robot, because of the combinatorial explosion of the sensory patterns exchanged by the LOCISS among multiple robots. To reduce this number of combinations and to realize a feasible control mechanism which can be installed in a robot's onboard computer system, a multilayered RL scheme for acquisition of appropriate collision avoidance behaviors is proposed by Fujii *et al.* [6]. It is constituted of four layers of modular controllers corresponding to the stages of RL. Another effective collision avoidance algorithm for two robots, suggested in [7] is generated by a very simple learning process that simulates a naive human trial-and-error learning process. This approach uses only the robot's sensor outputs and a suitable reward function, where the exact

form of the reward function is learned autonomously by the robots. The authors discuss also how a robot can use its 'experience' gained in a simple environment to adjust itself to a more complex environment by automatically generating a collision avoidance algorithm for a three-robot situation using a reduced state space for the case of two robots.

In this paper, an approach based on RL is suggested to achieve a collision avoidance in multi-robot environments. This approach allows ARS to interact and adapt their behaviors to achieve the desired task.

3 Reinforcement Learning Based Collision Avoidance Approach for Multiple ARS

In this Section, the collision avoidance problem in multi-robot environments is developed using the suggested RL based approach by learning through experimentation to choose actions so as to maximize one's productivity in these dynamic environments. In such environments, the situations become complex when the number of ARS increases. To achieve a collision avoidance in such situations, it is necessary to adopt an adaptive approach to acquire behaviors to avoid collisions among ARS and with obstacles. The RL paradigm based on Q-learning algorithm is introduced in the robot learning process acquire the appropriate behavior to navigate while avoiding collisions.

Indeed, unlike most learning algorithms that have been studied in the field of Machine Learning, reinforcement learning techniques allow to find optimal action sequences in temporal decision tasks where the external evaluation is sparse, and neither the effects of actions, nor the temporal delay between actions and its effects on the learner's performance is known to the learner beforehand [15]. The designated goal of learning is to find an optimal policy, which is a policy for action selection that maximizes future pay off (reward). In order to do so, most current reinforcement learning techniques estimate the value of actions, i.e., the future pay off one can expect as a result of executing an action, using recursive estimation techniques [15, 8].

3.1 ARS and Sensors

Each ARS has five infrared sensors (transmitters and receivers) as shown in Figure 1, and is capable of detecting the relative position of another robot within its sensing range. By transmitting/receiving motion information, that is, moving direction and speed, each robot can recognize other robot's motion easily [1].

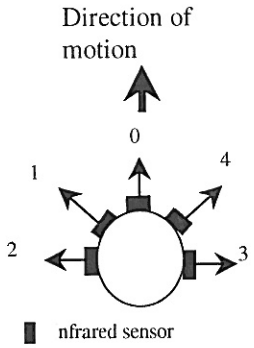


Figure 1: Robot Model.

3.2 Elementary Behaviors

The robot has no knowledge a priori of the environment where it moves. Its structure must allow to learn to behave only from interactions in the environment. The robot uses two elementary behaviors to act on the environment to change its state:

Forward : the robot moves towards its target.

Avoid : The robot turns when detecting an obstacle.

These behaviors conduct to the following actions which are used in the suggested approach:

Ignore : the robot moves towards its goal ignoring the objects around.

Follow : the robot moves towards its goal ignoring the objects around but with a reduced velocity.

TurnL : the robot performs a movement in left according to its orientation.

TurnR : the robot performs a movement in right according to its orientation.

The set of actions are then : $A = \{I, F, TL, TR\}$.

3.3 System Description

In most current real world applications, there must be the situation in which three or more robots and objects happen to aggregate in a small area. The objective of the learning process is to acquire the appropriate behaviors to get to each robot's own goal while avoiding collision to other robots and obstacles based on the information communicated by the sensors.

In this paper, a situation s in the environment is defined by the state of the sensor set. This state represents the existence or not of an ARS or an obstacle in each ranging area of each sensor. The collision avoidance problem in such multi-robot environment is solved by interacting with it. Each robot acquires the capacity to intelligently avoid collisions with other robots and with obstacles. This collision avoidance behavior is essentially based on a RL scheme acquired by interaction as shown in Figure 2. The objective of the ARS is to collect the maximum of rewards. Therefore, it

must choose the most rewarded actions and avoid the most punished. Thus, the suggested approach must make the robot capable to avoid robots and obstacles by interacting with the environment.

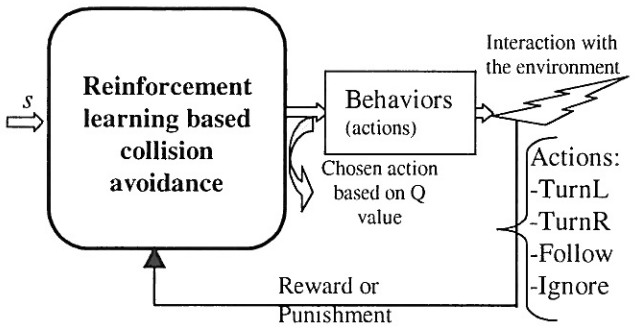


Figure 2: System Synopsis.

3.4 Learning by Interaction

In order to accomplish collision avoidance in multi-robot environment where the situation becomes very complicated, it is necessary to introduce learning schemes. This learning lets the robot acquire adaptive behaviors with little or no a priori knowledge of the environment where the robot will work [11]. In fact, the robot learns through trial and error interactions with the environment. During the navigation, each ARS must build an implicit internal map (i.e., obstacles and free spaces) from sensor data, update it and use it for intelligently controlling its collision avoidance behavior. This behavior is acquired by learning without any teaching signals from sensory information.

The state of the sensors defines a situation s. For every situation $s \in S$, the robot can take an action a from the action set A. The action $a \in A$ for the situation $s \in S$ causes the transition of the situation to $s' = e(s, a) \in S$, where e is the given transition function which defines the environment. The purpose of the reinforcement learning is finding an optimal policy to select the action a for the situation s that maximizes the discounted sum of the reinforcement signals $r(s, a)$ received over time. Watkins' Q-learning algorithm [19] gives us an efficient solution to this problem.

Learning Procedure. The adaptive acquisition process based on Q-learning is then conducted for a specific situation recognized by the sensory system according to the procedure shown in Figure 3. The score base which is the value table (i.e., Q matrix) of Watkins [19] is a series of scores allowing selection of behaviors.

An ARS learns a given behavior by being told how well or how badly it is performing as it acts in each given situation. As feedback, it receives a single information item from the environment. By successive trials and/or errors, the robot determines a mapping function which is adapted through the learning phase as shown in Figure 3. This learning procedure uses a single common value table shared and updated by all the ARS. Therefore the learning is shared within the group and the ARS learn more quickly because they take advantage of the other robot's experiences.

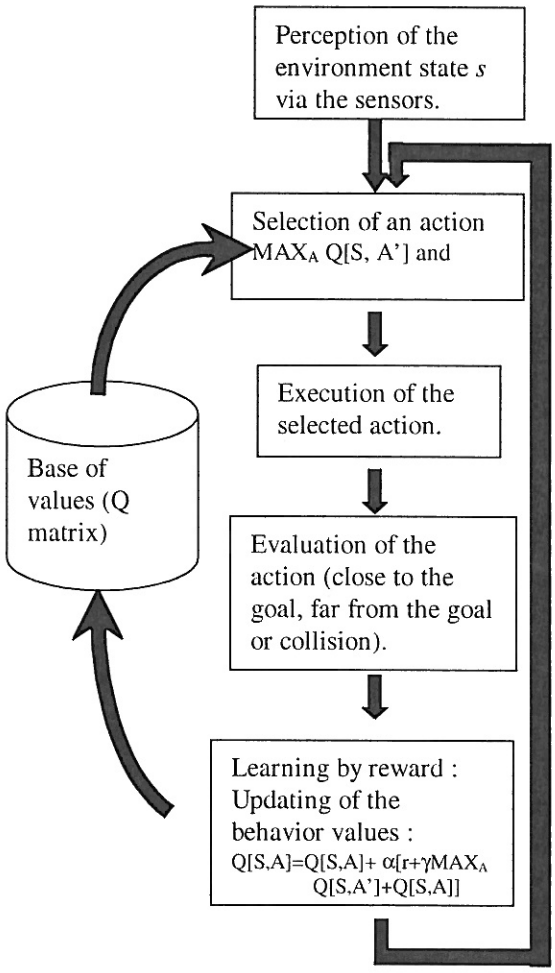


Figure 3. Learning Procedure.

Reinforcement function. The values of the reinforcement signal are usually hand-tuned and emerge after lots of experiments. These values must inform the robot if the action accomplished is good or bad. Indeed in RL, the behavior is synthesized using, as a unique source of information, a scalar, the so-called reinforcement value, which evaluates behavior actions: the robot receives either positive or negative reinforcements according to the utility (i.e., desirability) of the obtained situation as a consequence of the performed action.

In this paper, the ARS receives the following reinforcement signals during learning:

+2 if a robot reaches its target,

-0.1 If a robot moves away from the target,
-5 if a collision occurs.

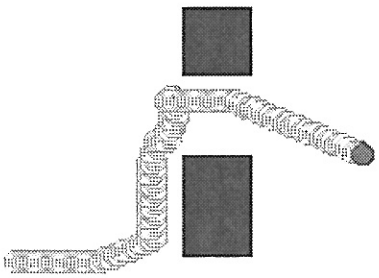
4 Simulation Results

To reflect the collision avoidance behavior of ARS based on the suggested approach, the ARS navigation is simulated in different environments and the used parameters in the learning algorithm are summarized in table 1.

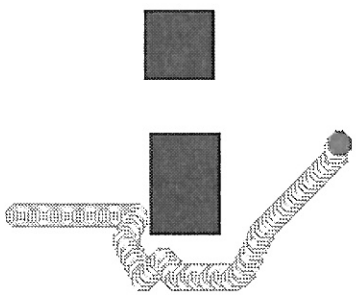
Table 1: Q-learning Parameters.

Parameters	Values
Learning rate α	0.4
Discounting factor γ	0.9
Exploration ratio	12.5%

The simulation results presented in Figures 4, 5, and 6 illustrate the robot learning of the collision avoidance behavior by interaction with the environment. Indeed, in Figure 4, an ARS tries to find its way to reach its target. The ARS takes two different ways as shown in Figure 4 (a) and (b). In Figure 4 (b), the passageway between the two obstacles is so confined than the robot could not pass. Therefore, it tried another way by trial and error until finding the new path conducting to its target (see Figure 4 (b)). This is made possible by the learning from interaction and particularly to the exploration capability of RL approach with its self-learning control.



(a)



(b)

Figure 4: Collision Avoidance Between an ARS and Obstacles.

In Figures 5 and 6, several ARS intersect without collisions among them and with obstacles while reaching their goals.

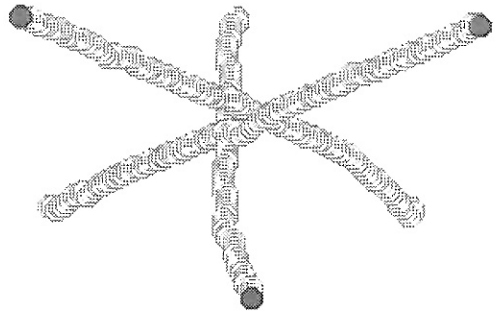


Figure 5: Collision Avoidance Among Three ARS.

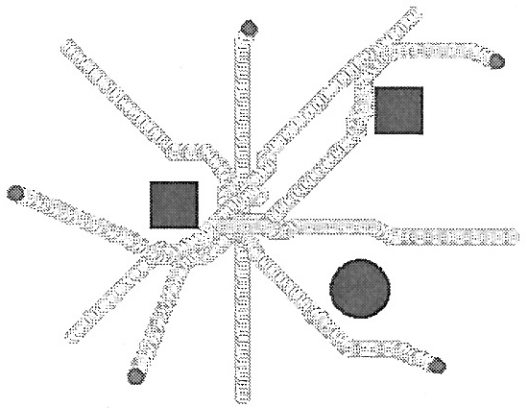


Figure 6: Collision Avoidance Among Multiple ARS in a Dynamic Environment.

These simulation results display the capability of the suggested approach to endow each ARS with an adaptive behavior acquired from interaction with the environment to achieve a desired task with restricted or no knowledge a priori.

5 Discussion and Conclusion

To solve the collision avoidance behavior problem of multiple ARS in multi-robot environments, a RL based collision avoidance approach is suggested. In such environments, it is difficult to prepare a teaching signal or to collect the sample necessary for the training. Indeed, RL allows, at least in principle, to bypass the problems of building an explicit model of the behavior to be synthesized or needing a meaningful learning base for supervised learning. With this adaptive learning approach, the robot is only guided by reinforcements fed back by the environment and is done incrementally and progressively since its parameters are updated at each

step and then are sensitive to all changes in the environment especially the value table. This table, which gives an evaluation of the selected behavior, allows learning by interaction with the environment. This learning is shared within the group by updating the same value table and the ARS learn more quickly because they take advantage of the other robot's experiences. The ARS endowed with the suggested approach succeed in their navigation without collision in an environment a priori unknown. Thus, this approach allows a *real-time* navigation based on a continual learning in a dynamic environment. This approach based on Q-learning scheme allows to learn by interaction how to avoid other ARS or obstacles even in presence of uncertainties. Indeed, in the real world applications, there must be unpredictable random noises and the answers to the complicated collision avoidance problems cannot be derived easily by human designers. It is, then, favorable that the answers to the problems can be automatically acquired through the learning process in real or simulated world (from experience).

The simulation results display the ability of the suggested approach to provide ARS with capability to intelligently avoid collisions among them and with obstacles, in unvisited environments, illustrating the *robustness* and *adaptation* capabilities of this approach.

An interesting alternative for future research is the use of this collision avoidance approach to achieve more complex tasks such as the transport of heavy and large objects with different shapes by multiple ARS navigating in formation.

References

[1] Y. Arai *et al.*, "Adaptive behavior acquisition of collision avoidance among multiple autonomous mobile robots", *Proc. of the Int. IEEE Conf. On Intelligent Robots and Systems,* Grenoble, France, pp. 1762-1767, 1997.

[2] Y. Arai *et al.*, Realisation of Autonomous Navigation in Multirobot Environment, *Proc. 1998 IEEE/RSJ int. Conf. On Intelligent Robots and Systems*, pp. 1999-2004, 1998.

[3] O. Azouaoui and A. Chohra, ''Evolution, behavior, and intelligence of Autonomous Robotic Systems (*ARS*),'' *Proc. 3rd Int. IFAC Conf. Intelligent Autonomous Vehicles*, Madrid, Spain, March 25-27, pp. 139-145, 1998.

[4] O. Azouaoui and A. Chohra. ''Neural group navigation approach for Autonomous Robotic Systems (*ARS*).'' *Proc. 2nd Int. ICSC Symp. on Eng. of Int. Sys.*, University of Paisley, Scotland, June 27-30, 2000.

[5] K. Berns *et al.*, "Reinforcement-Learning for the Control of an Autonomous Mobile Robot", *Proc. Of the IEEE/RSJ Int. Conf. On Intelligent Robots and Systems IROS92,* Raleigh NC, July 7-10, pp. 1808-1815, 1992.

[6] T. Fujii *et al.*, "Multilayered Reinforcement Learning for Complicated Collision Avoidance Problems", *Proc. 1998 IEEE Int. Conf. On Robotics And Automation*, pp. 2186-2191, 1998.

[7] Y. Fujita *et al.*, "Learning-Based Automatic Generation of Collision Avoidance Algorithms for Multiple Autonomous Mobile Robots", *Proc. 1998 IEEE/RSJ int. Conf. On Int. Robots and Sys.*, pp. 1553-1558, 1998.

[8] L.P. Kaelbling, "*Learning in embedded systems*", PhD Thesis, Stanford University, 1990.

[9] M.J. Mataric, "*Interaction and Intelligent Behavior*", PhD Thesis, Massachusetts Institute of Technology, 1994.

[10] M.J. Mataric, "Behavior-Based Control: Examples from Navigation, Learning, and Group Behavior", *J. of Experimental and Theoretical Artificial Intelligence, Special Issue on Software Architecture for Physical Agents, 9(2-3), H. Hexmoor, I. Horswill, and D. Kortenkamp, eds.*, pp. 323-336, 1997.

[11] H. Murao and S. Kitamura, "Q-Learning with Adaptive State Segmentation (QLASS)", *0-8186-8138-1/97 $10.00 © 1997 IEEE*, pp. 179-184, 1997.

[12] R. Pfeifer, "Building 'Fungus Eaters': Design Principles of Autonomous Agents", *Fourth Int. Conf. onSimulation of Adaptive Behavior: From Animals to Animats*, Massachusetts, USA, September 09-13, pp. 03-12, 1996.

[13] R. Pfeifer and C. Scheier, "*Understanding Intelligence*", The MIT Press, Cambridge Massachusetts, 1999.

[14] R.S. Sutton *et al.*, "*Reinforcement Learning : an introduction*". MIT Press, 1998.

[15] S. Thrun and A. Schwartz, "Issues in Using Function Approximation for Reinforcement Learning", *Proc of Fourth connectionist Models Summer School Lawrence Erlbaum Publisher, Hillsdale, NJ*, Dec., 1993.

[16] S. Thrun, A Lifelong Learning Perspective for Mobile Robot Control, *Proc of Int.Conf on Intelligent Robots and Systems IROS'94,* Vol. 1, Munich, Germany, September 12-16, pp. 23-30, 1994.

[17] C.F. Touzet, "Neural Implementation of Immediate Reinforcement Learning for an Obstacle Avoidance Behavior", *Technical Report NM 94.6, LERI-EERIE, Parc G. Besse, F-30000*, Nimes, France, 1994.

[18] C.F. Touzet, "Neural Reinforcement Learning for Behavior Synthesis", *Int. J. of Robotics and Autonomous Systems,* Vol. 22, Elsevier, pp. 251-281, 1997.

[19] C. watkins, "*Learning from Delayed Rewards*", Ph.D. Dissertation, Cambridge University, 1989.

[20] S.D. Whitehead, "Reinforcement Learning for the Adaptive Control of Perception and Action", *Technical Report 406,* University of Rochester, February 1992.