# A Robot Controller Using Learning by Imitation

**Gillian Hayes** * **John Demiris** **

Department of Artificial Intelligence
University of Edinburgh
5 Forrest Hill
Edinburgh EH1 2QL
Scotland

**Abstract.** Roboticists have already invested considerable energy in building robot controllers which model the learning capacities of single animals. In this paper we present a new type of controller which draws upon insight from the field of imitative learning: one agent learns from perceiving and imitating the behaviour of another. We describe the architecture of an imitative learning controller and two implementations, a simulator and a robot. The learner robot follows a teacher robot through a maze and learns to associate its perceptions at locations where the teacher carries out a significant action with the action it subsequently undertakes as a result of its innate teacher-following behaviour. Such a controller limits the learning task to bouts of learning when there is something useful to be learnt. It allows a robot to learn in terms of its *own* perceptions, makes programming many nominally identical robots simpler, and opens the possibilities for cross-modal learning.

## 1 Introduction

Roboticists have invested considerable effort in implementing biologically inspired control mechanisms in robots. They have made progress in the design of reflex behaviour for robots and have started to make inroads into the field of robot learning, for example by programming robots to form associative maps of their environment or to learn from reinforcement received from the environment for successful behaviour [10, 8, 9, 12, 1]. In this paper we present a controller which utilises a new type of learning, imitative learning, whereby one robot learns how to act by perceiving and, in some sense, imitating the actions of another robot. Once it has learned, it is able to negotiate new situations on its own.

*Imitative learning* is a term taken from ethology, which ethologists use to describe a huge range of behaviour in animals [2, 3]. Indeed, there is substantial disagreement among ethologists as to what exactly constitutes "imitation" in animals, since the term is used to describe phenomena as disparate as vocal mimicry in birds, the tendency of chicks to eat more when in the company of conspecifics, and faster learning of tasks by rats or kittens who are able to observe a "teacher".

The mechanisms postulated by ethologists to underlie these phenomena are correspondingly disparate. Zajonc [13, 14] suggests that the mere presence of conspecifics may make it more likely that an animal carry out a particular behaviour; on the other hand, it may be necessary for the conspecifics also to be undertaking the same behaviour for facilitation to occur. Simple fear reduction in the presence of others may be taking place, so that innate behaviours are disinhibited rather than being positively activated. These mechanisms could explain such social facilitation behaviour as the tendency of chicks to peck at food more when others are around.

Thorpe [11] suggests that contagious behaviour may result if the sight of one animal performing some innate behaviour is the stimulus which triggers this behaviour in another as in, for example, yawning in humans or flock manoeuvres in birds. On the other hand, Humphrey [7] suggests that classical conditioning could be involved: if some stimulus always releases the same behaviour in a group of animals, then that behaviour in one animal is the conditioned stimulus for the same behaviour in another. Yet more mechanisms are postulated to underlie other sorts of imitative behaviour.

In general, the specifications of these "mechanisms" given in ethological descriptions are not sufficient (or not in the right form) to be directly implementable as a robot controller. However, such a controller would clearly make robot programming easier. Even robots which are nominally identical cannot be made to behave robustly by the simple expedient of copying a program from one to the next: no two robots are identical and no two sensors the same. We therefore adopt the following strategy.

* gmh@aifh.ed.ac.uk
** johnde@aifh.ed.ac.uk

First we allow a very general definition of “imitation” as being a phenomenon which leads to “some sort of similarity in behavior among two or more individuals” [2]. We then pick a mechanism put forward as a possible explanation of one type of imitative learning, matched dependent behaviour, and design and build a robot control system based on this. Our main aim is the construction of a robot controller that is robust and versatile, so that our system and the expressed behaviour of the robot may well not be completely faithful to the ethological case. Finally, our experiences with this system may allow us in the future to further our second aim, the elucidation of the ethological mechanisms which are hypothesised to underlie particular types of imitative learning.

## 2 Matched Dependent Behaviour

We take “matched dependent behaviour” to refer to the phenomenon whereby, for example, rats negotiating a maze receive reinforcement faster if they match their behaviour to that of a leader rat (the leader knows how to find the food), and then learn to associate the other stimuli in the environment at the time (presumably those which stimulated the leader to do the right thing in the first place) with the appropriate action. They are then able to carry out that action in the absence of the leader.

In our system we concentrate particularly on the second stage of this process. A teacher robot negotiates its environment; a learner robot follows and watches for significant events in the behaviour of the teacher. The learner matches its behaviour to the teacher’s and learns to associate the stimuli in the environment at the points at which the teacher performs significant actions with the behaviour it subsequently carries out by virtue of “imitating” the teacher. Currently, no reinforcement is obtained by either the teacher or the learner. The teacher is able to negotiate its environment without reward, this behaviour being simply programmed in; and the learner has no choice but to follow the teacher. Since our initial aim is to build a useful robot controller, omission of the (generally slow, computationally expensive) reinforcement learning stage is permissible; it should, of course, be included if we intend to provide an explanation of the biological phenomenon.

In robotic imitative learning perception is clearly crucial. The learner robot must be able to perceive what the teacher robot is doing and to respond preferentially to the relevant stimuli in its environment. Perception is usually difficult; there is the conceptual problem of deciding just which stimuli are the relevant ones, and there are often the practical problems of accommodating on board the robot the large amounts of computational power needed, for example, if complete video images need to be processed to a high level of abstraction, and of getting this processing done in real time. Whilst not denying these difficulties, we contend that useful imitative learning can be carried out using simpler detectors, such as infra-red distance sensors, and simple patterns of movement, such as straight-line locomotion and deviation therefrom, and we have chosen to use these in the first stage of our investigation of the architecture of a complete controller.

We have opted initially to have the teacher robot travel through a maze, negotiating corners when it reaches them, with the learner following a certain distance behind. The learner notices when the teacher changes its direction of travel – a significant event – in order to negotiate a corner. The learner then associates what it perceives in the environment around it with the action the teacher has carried out, and imitates that action. Eventually the learner is able to make use of these associations to respond to corners correctly on its own.

Our choice of experimental testbed was governed by the need to be able to evaluate whether the robot had actually learned anything. By choosing to teach a robot to navigate a maze and then testing its learning by asking it to negotiate either the same or a new (different) maze, we have a demonstration that learning has taken place.

## 3 Architecture

In this section we describe the general architecture which governs our realisation of the matched dependent behaviour mechanism. The framework is very general, although we have described it as it is used in maze negotiation. Each module requires difficult aspects to be addressed; there will be many possible solutions, often research problems in their own right. We go on to describe two implementations of the architecture, in simulation and on autonomous mobile robots, and present our first results. The architecture comprises five modules which we describe in turn.

*Maze negotiation* The teacher’s maze negotiation behaviour is implemented in the first module. It must allow the teacher to respond appropriately to situations such as confronting a wall, right and left corners, T-junctions, and so on. Its implementation could take a very simple form, such as dead reckoning, or one could use sensors and a negotiation strategy. A reinforcement learning module could be useful: although taking many trials to learn, it would be expected to be insensitive to sensor noise and small variations in the environment. Alternatively, the robot could be led through the maze several times by a human and learn using the fast Halperin net [5, 4], which takes advantage of designer knowledge of the task and so does not need to learn the environment and task from scratch.

In the general case, this module would be responsible for producing whatever behaviour the teacher is teaching; there is no limit to the number of different instantiations and complexity of the module.

*Teacher following* In this module, which is independent of the maze negotiation module, the learner robot is programmed to follow the teacher. In the simplest implementation the learner keeps a fixed distance behind the teacher regardless of the teacher's motion; more sophisticated implementations could have the learner follow only when the teacher is moving purposively and ignore "loitering". In both cases, the learner must be able to detect the teacher and its distance from the teacher, and have strategies to deal with the situation when the teacher moves out of sight.

In the general case, this module is responsible for making the learner imitate the behaviour of the teacher. When the teacher's behaviour can be described as bulk motion the implementation of the module is quite straightforward; however, how the learner might imitate the teacher's behaviour in, say, a manipulation task involving arm movements is less clear. The learner would then need some understanding of what the teacher is doing from the teacher's point of view in order to be able to match its behaviour to the teacher's. In the case of simple motion it is much easier to insulate what the learner knows from what the teacher knows, with the sole means of communication between the two being the perception by the learner of the teacher's behaviour.

*Significant event perception* The learner robot recognises when the teacher robot is performing an action it deems significant and remembers the place at which that action is being performed. Ways of recognising that an action is significant include detecting a change in direction of movement, or detecting a change in the direction in which the teacher is pointing; the teacher could also signify it is about to undertake a significant action by emitting a sound ("watch me").

This module allows the learner to recognise that one of the actions it is about to do (we assume that the learner's behaviour time-lags the teacher's) is to be distinguished from the rest; from this point onwards the learner must be ready to learn. Thus the learner engages in short bouts of learning which coincide with those periods in which there is something useful to be learnt; it does not have to process a continuous stream of perceptions and associate them with a continuous sequence of actions. By directing the learner's attention the complexity of the learning process is reduced.

*Association 1: environment perception* The learner robot knows from the significant event perception module the location of and its distance from the significant event. It moves to this location and then perceives its environment. This is the stimulus with which an appropriate action is to be associated. The implementation depends very much on the sensing capabilities of the learner, which could vary from simple binary distance detectors to general purpose vision systems. In most cases one would expect to use specialised sensors, tailored to the environment. The learner must be able to discriminate sufficiently the salient features of its environment, but it must also be able to generalise well enough that when it finds itself in a situation which is similar to, but not identical with, the first, it is still able to carry out the appropriate action.

*Association 2: recognition of own action* Finally, the robot must associate an action with the stimulus it has just received from the environment: this is essentially the imitated action. In our systems we make use of the fact that the learner is performing a teacher-following behaviour all the time; thus in order to discover the appropriate action, the learner simply has to move a few steps further in its path behind the teacher and keep note of what exactly it does, and it will automatically have carried out the correct action. The sequence of learner movements just before and just after the significant event is the action that is associated with the perception. The implementation of this module could vary widely, and there is also room for variation in the design of the action recognition itself.

The form of the association itself can be simple or complex, from a set of if–then rules to a connectionist network. We are currently using rules, although further connectionist developments are envisaged once the amount of information provided by the sensors is large enough. The complexity required depends in large part on the variability in the environment, the noise in the sensors, the precision with which the learner can measure its own actions and the degree to which a robot's behaviour can be characterised as a sequence of significant, recognisable events.

## 4 Results from Simulation

We have implemented this architecture on a simulator, which was built to allow us to rapidly test the applicability of this and other architectures. The current implementation on real robots is described below.

The implementation of learning on the simulator has two phases:

**Learning phase** Robot teacher executes a series of moves which take it through the maze and the robot learner follows the steps of the teacher, associating environment configurations with actions followed when something significant happens.

**Testing phase** The robot learner, equipped with the rules it acquires from the learning phase, is requested to traverse either the same or a new (different) maze on its own.

During the first phase (Figure 1), a series of moves is executed by the robot teacher (one of MoveForward, MoveLeft, MoveRight), while the robot learner follows the teacher by maintaining a close distance and changing direction, with a suitable time-lag, whenever the teacher does so. The teacher is never allowed out of sight of the learner. The orientation of the teacher is detectable by the learner; a change in the orientation (by 90 degrees) of the teacher is the signal for the learner that something significant has happened. The learner waits until it reaches the point of change; since it knows how far behind the teacher it is, it just has to travel this distance towards the teacher. Then it invokes the association procedure. This involves sensing of the environment around it (i.e. the configuration of the walls, essentially the type of corner it is dealing with), and associating that with the appropriate action. In order to determine this action the robot waits to see what it does next, governed by the instinctive following behaviour. It calculates the action as the difference between the direction in which it was travelling before it reached the significant event location, and the direction in which it is travelling afterwards. In our particular implementation, this is either a rotation of 90 degrees clockwise, or 90 degrees anticlockwise. After the learning phase we can see that these turns are associated with the right and left corners respectively.

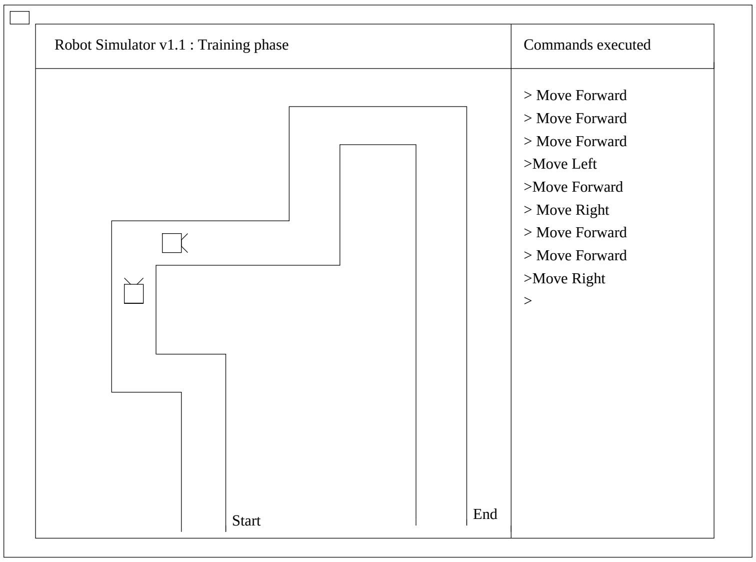


Fig. 1.: The teaching phase of the simulator. Teacher and learner robots are shown.

During the second phase (Figure 2) the robot learner is placed at the beginning of either the same or a different maze, and is requested to traverse it without the help of the teacher. It constantly senses the environment around it, attempting to match the configuration of the walls with those specified as significant configurations in the left-hand side of the rules acquired during the learning phase. When a particular configuration matches, the operator which is associated with it is applied to the robot, causing a change of direction, and then the robot continues moving forward again.

The results so far are encouraging. In the simulated implementation, the robot learner successfully learns to traverse increasingly complicated mazes. An example of the rules acquired is shown in Figure 2 where the default learner behaviour is also shown as a rule. The architecture is capable of dealing with simple learning situations. Our next step with the simulator will be to increase the complexity of the maze negotiation and teacher following modules, and particularly, the form of the association modules. It is unrealistic to expect to be able to simulate the robots' sensing capabilities properly; for this we have turned to real robots.

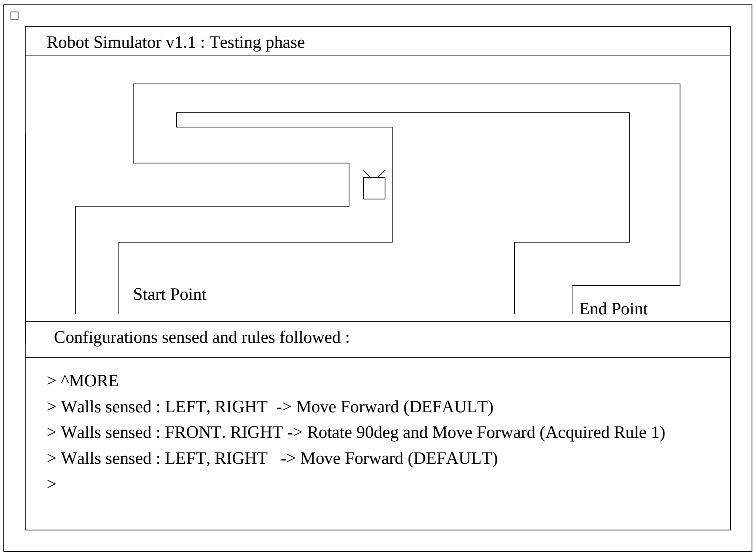


Fig. 2.: The testing phase of the simulator. The learner robot negotiates a different maze on its own.

## 5 Robot Implementation

In this section we describe how the imitative learning system is implemented to control the behaviour of two robots.

A Lego robot, of the type used for research and teaching in our Department, is used as the robot teacher. This is a small car-like vehicle, with two motors powered by batteries, augmented with a 68000-based microprocessor board (Brain Brick) and various specially built sensing bricks which process the output of infra-red, light-dependent resistor, Hall effect and bump-switch sensors. One of the Departmental research robots, Ben Hope, serves as the learner. This is an autonomous transputer-based mobile robot, cylindrical in cross-section and about 50cm high. It is equipped with a video camera, a bumper, and motion-detection, ultrasonic, and infra-red sensors; it has a radio modem for communication with a host PC if required. All processing takes place on board the robots, although data can be stored and downloaded to a workstation.

Our initial implementation has the Lego robot navigating through a maze of wooden boards about 25cm high. Currently we are most interested in experimenting with the learning modules, so we have chosen to have the teacher navigate by dead reckoning. A smarter strategy will be used when we switch the environment to a more complex corridor and office area. The learner, Ben Hope, follows, using its camera to sense a pointing device mounted on top of the Lego robot. This device is a triangular configuration consisting of five LEDs positioned in a row with another single LED situated about 10cm from the row. The LEDs are visible in a video image from Ben Hope's camera, making it a simple matter to detect both the Lego robot itself and the direction in which it is moving. The view of the teacher seen by Ben Hope is shown in Figure 3. Ben Hope's camera has quite a wide field of view, so we discard the upper part of the image before thresholding it and applying Haralick's iterated components algorithm (see [6]) to extract connected, labelled areas (Figure 4). The positions of the front and back of the teacher robot and the direction in which it is pointing can then be obtained by fitting ellipses to the two blobs. Given the height above ground of Ben Hope's camera, it is simple to calculate the distance to the Lego robot. All of the image processing is carried out on Ben Hope's framegrabber TRAM.

The speed of Ben Hope is then controlled to keep it a fixed distance behind the teacher robot; effectively we are trying to keep the position of the centroid of the teacher's pointing device within a small region round the centre of the image. The direction in which Ben Hope is travelling can also be adjusted so that it follows the teacher round corners. A significant event is signalled by the orientation of the teacher robot's pointing device changing by some threshold amount. Maze wall configurations are detected by the infra-red sensors mounted around Ben Hope, and the action to be associated with this perception is the degree of turning required to bring the teacher robot back into Ben Hope's field of view. As in the simulator, the acquired association of perception and action takes the form of simple rules.

Work is currently being carried out on the integration of these modules; we hope to present the results of the learning experiments at the symposium.

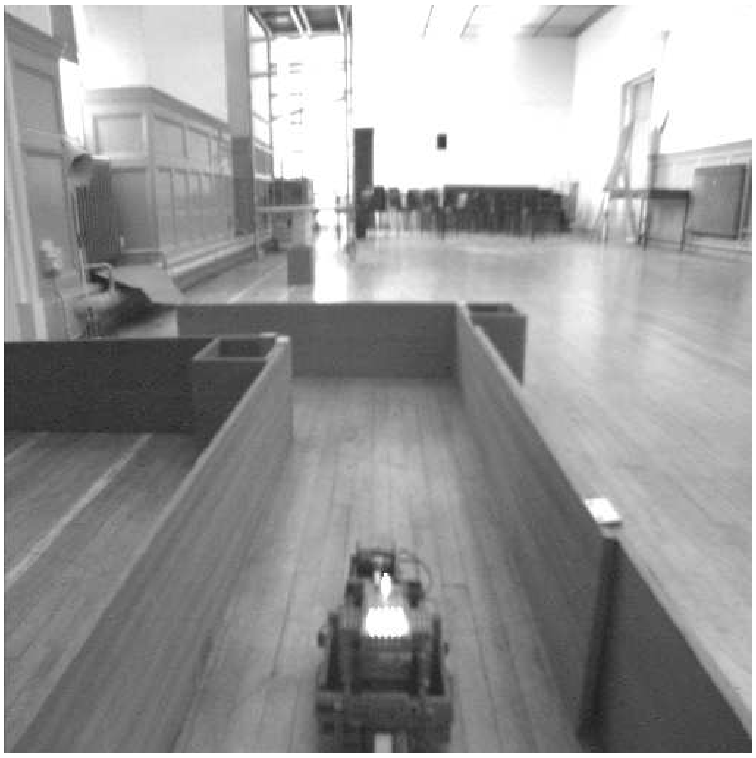

Fig. 3.: View of the teacher and part of the maze as seen by Ben Hope.

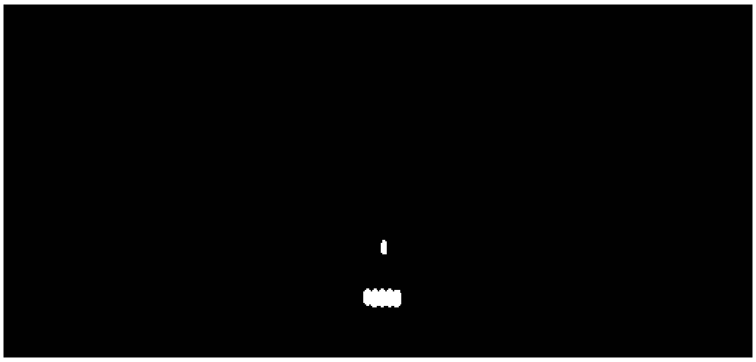

Fig. 4.: Thresholded image showing the front and back of the teacher robot (small and large blobs, respectively).

## 6 Conclusions and Summary

We have described the architecture we are using to investigate a new mechanism of robotic learning, namely imitative learning, and two implementations applied to a maze problem, one in simulation and one using robots. Results from the simulation are promising; the robot system is being integrated.

The use of this particular architecture has certain advantages. Firstly, it is not necessary to have the teacher robot learn how to follow the maze before we start; we can provide it with an algorithm tried and tested in other experiments. Secondly, and very importantly, the learner robot does not need any understanding of how the teacher perceives the world; it learns what to do based on how *it* perceives the teacher's actions, the world around it and its own actions – it is learning on its own terms. This is a very clear indication of how combining perception and action can produce useful behaviour. Moreover, the independence between teacher and learner means that different robots can be used, with different morphologies and different sensory capabilities.

The robot controller based on imitative learning which we implement is not entirely ethologically faithful. However, we expect that the insight gained from such experiments will allow us to model ethological mechanisms in the future. We do expect this type of controller to provide us with more versatile robots which learn quickly and which are relatively insensitive to small changes in the environment. Moreover, programming nominally identical robots will become easier. It is not sufficient simply to copy a program from one robot to the next: no two robots are identical, no two sensors the same. Having one robot imitate another will obviate this problem. Learning by imitation in this fashion also opens the possibilities of cross-modal learning: a robot which starts work when the sun rises could teach another to start work when it hears the birds singing.

## 7 Acknowledgments

We would like to acknowledge the support of the Department of Artificial Intelligence, University of Edinburgh, for the provision of computing, laboratory and workshop resources. Ben Hope was built under SERC grant GR/F/58523. Thanks are also due to John Hallam, Anton Dil and Andrew Fitzgibbon for comments on drafts of this paper and technical advice.

## References

1. William Chesters and Gillian Hayes, “Connectionist environment modelling in a real robot”, to appear in Proceedings of the Third International Conference on Simulation of Adaptive Behavior, Brighton, August 8–12, 1994, MIT Press.
2. J. Michael Davis, “Imitation: a review and critique”, in Perspectives in Ethology Vol. 1, P.P.G. Bateson and Peter H. Klopfer (eds.), 43–72, Plenum, New York, 1973.
3. Bennett G. Galef, Jr., “Imitation in animals: history, definition, and interpretation of data from the psychological laboratory”, in Social Learning: Psychological and Biological Perspectives, Thomas R. Zentall and Bennett G. Galef, Jr. (eds.), 3–28, Lawrence Erlbaum Associates, Hillsdale, N.J., 1988.
4. Bridget Hallam, John Hallam and Gillian Hayes, “A dynamic net for robot control”, to appear in Progress in Neural Networks, 1994.
5. Janet R.P. Halperin, A Connectionist Neural Network Model of Aggression, PhD Thesis, Department of Zoology, University of Toronto, 1990.
6. R.M. Haralick and L.G. Shapiro, Computer and Robot Vision, Vol.1, Addison-Wesley, Reading, MA, 1991.
7. G. Humphrey, “Imitation and the conditioned reflex”, Pedagogical Seminary **28**, 1–21, 1921, cited in [3].
8. Pattie Maes and Rodney A. Brooks, “Learning to coordinate behaviors”, Proceedings of AAAI 1990, Boston, MA, July 29 – August 3, 1990, 796–802.
9. Sridhar Mahadevan and Jonathan Connell, “Automatic programming of behavior-based robots using reinforcement learning”, Proceedings of AAAI 1991, July 14–19, 1991, 768–773.
10. Ulrich Nehmzow and Tim Smithers, “Mapbuilding using self-organising networks in ‘really useful robots’ ”, in From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior, Paris, September 24–28, 1990, Jean-Arcady Meyer and Stewart W. Wilson (eds.), MIT Press, 1991, 152–159.
11. W.H. Thorpe, Learning and Instinct in Animals, 2nd edition, Methuen, London, cited in [3].
12. A. Walker, J. Hallam, D. Willshaw, “Bee-havior in a mobile robot: the construction of a self-organizing cognitive map and its use in robot navigation within a complex, natural environment”, in Proceedings of the IEEE International Conference on Neural Networks, Volume 3, San Francisco, March 28 – April 1, 1993, 1451–1456.
13. R.B. Zajonc, “Social facilitation”, Science **149**, 269–274, 1965, cited in [3].
14. R.B. Zajonc, “Coactivation”, in Animal Social Psychology, R.B. Zajonc (ed.), Wiley, New York, 1969, cited in [3].

This article was processed using the TEX macro package with IRS94 style