1995-02

Obstacle Avoidance by Means of an Operant Conditioning Model

OBSTACLE AVOIDANCE BY MEANS OF AN OPERANT CONDITIONING MODEL

Eduardo Zalama, Paolo Gaudiano, and Juan López Coronado

February 1995

Technical Report CAS/CNS-95-005

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA 02215

OBSTACLE AVOIDANCE BY MEANS OF AN OPERANT CONDITIONING MODEL

Eduardo Zalama†, Paolo Gaudiano‡, Juan López Coronado†

†Departamento de Ingenieria de Sistemas y Automatica, Universidad de Valladolid
Paseo del Cauce s/n, Valladolid 47011, Spain
‡Department of Cognitive and Neural Systems, Boston University
111 Cummington Street, Boston, MA 02215, USA

Abstract

This paper describes the application of a model of operant conditioning to the problem of obstacle avoidance with a wheeled mobile robot. The main characteristic of the applied model is that the robot learns to avoid obstacles through a learning-by-doing cycle without external supervision. A series of ultrasonic sensors act as Conditioned Stimuli (CS), while collisions act as an Unconditioned Stimulus (UCS). By experiencing a series of movements in a cluttered environment, the robot learns to avoid sensor activation patterns that predict collisions, thereby learning to avoid obstacles. Learning generalizes to arbitrary cluttered environments. In this work we describe our initial implementation using a computer simulation.

1 Introduction

One of the aspects to consider when an animal or intelligent machine has to operate in an unknown environment is that it must learn to predict the consequences of its own actions. By learning the causality of environmental events, it becomes possible to predict future and new events.

Models of classical and operant conditioning have emerged from the field of psychology in order to try to explain how an organism can achieve autonomous behavior in a constantly changing environment (Rescorla & Wagner, 1972; Sutton & Barto, 1981; Grossberg, 1982).

In the classical conditioning paradigm, learning occurs by repeated association of a Conditioned Stimulus (CS), which normally has no particular significance for an animal, with and Unconditioned Stimulus (UCS), which has significance for an animal and always gives rise to an Unconditioned Response (UCR). For example, a dog that repeatedly hears a bell before being fed will eventually begin to salivate when the bell is heard. The response that comes to be elicited by the CS after classical conditioning is known as the Unconditioned response (CR).

Another related form of learning is known as *operant conditioning*. In this case an animal learns the consequences of its actions. More specifically, the animal learns to exhibit more frequently a behavior that has led to a reward, and to exhibit less frequently a behavior that has led to punishment. For example, a hungry cat placed in a cage from which it can see some food will learn to press a lever that allows it to escape the cage to reach the food. In this situation, the animal cannot simply wait for things to happen, but it must generate different behaviors and to learn which are effective. This kind of learning has also been referred to as *reinforcement learning* (Sutton & Barto, 1981). The main problem in modeling this form of learning is how to learn which of a large array of behaviors has produced the reward.

We have used a model of classical and operant conditioning proposed by Grossberg (1971, 1986), Grossberg and Levine (1987) to train a wheeled robot to avoid obstacles by learning the patterns of sensor activation that predicts an imminent collision.

The remainder of this paper is structured as follows. In section 2 we describe briefly the conditioning model (Grossberg & Levine, 1987), describing its functionality and its applicability to obstacle avoidance. In section 3 we describe our simplified implementation of the model. In section 4 we show some results and performance of the model, and finally section 5 is dedicated to conclusions and future work.

2 Grossberg's conditioning model

Grossberg & Levine's (Grossberg & Levine, 1987) implementation of Grossberg's conditioning circuit is in figure 1. This model was used by Grossberg and Levine to explain a number of phenomena from classical conditioning.

In this model the *sensory cues* (both CSs and UCS) are stored in Short Term Memory (STM) within the population labeled S, which includes competitive interactions to ensure that the most salient cues are contrast enhanced and stored in STM while less salient cues are suppressed. At the bottom of the figure the *drive node* D is represented, and conditioning can only occur when the drive node is active. This node and the nodes in the population labeled P,

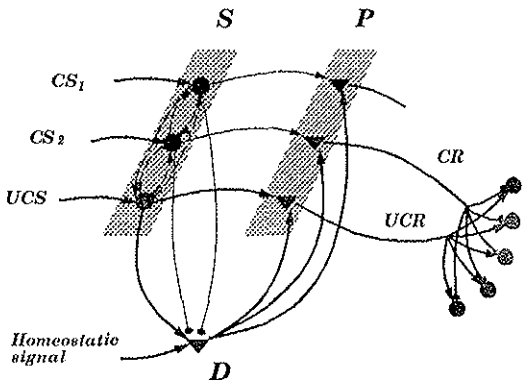

Figure 1: Conditioning circuit proposed by Grossberg & Levine. See text for details.

are represented as triangular nodes and are polyvalent cells: Polyvalent cells require the convergence of two inputs in order to become active. Finally, the neurons at the far right of the figure are represent the response (conditioned or unconditioned), and are thus connected to the motor system.

The design of figure 1 satisfies a number of fundamental constraints. Most important are the requirements that: (1) initially, only the UCS must be able to cause a response (the UCR); (2) after conditioning, the CS must be able to elicit a response similar to the UCR; (3) in order for learning to take place, the CS and UCS must be presented nearby in time.

The system satisfies these requirements as follows: initially it is presumed that only the UCS has a strong connection to the drive node D. When the UCS is turned on (e.g., a shock), it activates the drive node, which in turn sends activation up toward the polyvalent cells P. The P that receives joint input from the UCS and D nodes becomes active, and "reads out" the UCR on the motor cells. When an CS (e.g., a light) is presented by itself prior to conditioning, it cannot activate the D node, and thus it cannot activate its P cell, and no action is generated. However, if the UCS is turned on shortly after the CS has become active, then the D node becomes active, and the CS has a brief opportunity to sample the D node's activity through an associative learning rule. At the same time, the D node will also briefly activate *all* P cells that are receiving sensory input, and thus the P cell corresponding to the CS will be active, and it will learn the pattern of motor activity being generated by the UCS. If the pairing is repeated several times, the CS will develop strong connections to the D node, while it polyvalent cell will learn to imitate the UCR. Eventually activation of the CS generates a large enough signal to activate the drive node D, which in turn helps to activate the polyvalent cell P, and reads out a response similar to the UCR, that is, the CR.

Notice that in figure 1 the drive node must also receive a *homeostatic signal* to become active. This reflects the observation that a motivated behavior, such as eating, will not be released unless the animal sees food *and* is hungry. When the CS and UCS represent a fearful cue, it is assumed that the homeostatic signal corresponds to some form of survival drive, which is presumably always active in normal animals. In this case, as we do below, on e can simply assume that the drive node only needs a strong sensory cue to become active.

One of the main characteristics of the model is its dynamical nature and temporal causality. Temporal causality refers to the fact that the association between stimulus and response can be learned even though they are presented at different times in different trials. For example when an animal presses a lever to get food, the animal will learn the effect of its action regardless of exactly how quickly the food is presented (within a window of several seconds). The present model is able to reproduce other paradigms like blocking second order conditioning (Grossberg & Levine, 1987). Second order conditioning is useful as it allows a CS previously paired with a UCS, to act as a UCS for other conditional stimuli. For instance, a bell repeatedly paired with shock eventually comes to elicit fear. If a light is now repeatedly presented shortly before onset of the bell, even though the shock is never turned on, the light will also come to elicit fear. This form of "higher order" conditioning is extremely useful as it allows animals to learn early predictors of important events.

3 Conditioning and obstacle avoidance

In this section we describe how we have applied Grossberg's conditioning model to the problem of obstacle avoidance with a mobile robot. The mobile robot we have used in the simulations is a differential-drive robot, as is shown in figure 2.

We have previously introduced a neural network controller for this type of mobile robot, which learns the robot's forward and inverse odometry through a learning-by-doing cycle (Zalama, Gaudiano, & López-Coronado, 1995). That model, which we called NETMORC, includes two neural populations that code the distance and angle to a target as registered by the robot's sensory system. The distance and angle information is used to generate the

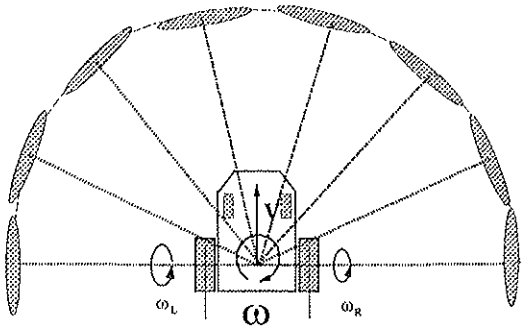

Figure 2: Mobile robot structure. The mobile robot has a set of 8 ultrasonic sensors uniform and frontal distributed through the robot.

wheel velocities required to move the robot toward the target. However, the NETMORC is not capable of avoiding obstacles.

In the present work we combine the conditioning model of figure 1 with the NETMORC in such a way that the robot can learn to avoid obstacles by modifying its angular velocity when it encounters obstacles in its path. Figure 3 illustrates the scheme we use to represent angular velocities. In this figure the leftmost node represents an angular velocity of $-w_m rad/s$, the right node represents an angular velocity of $w_m rad/s$ (where w_m is the maximum angular velocity developed by the robot), and the central node corresponds to a straight line movement. The activation pattern over the population is used to determine the wheel velocities that will move the robot. The map includes a sigmoidal transformation, whereby angular velocities close to zero are represented by a greater number of nodes. The sigmoidal function which selects the most active node in the map as a function of the angular velocity is given by:

$$n_d = \begin{cases} \frac{N}{2} + \frac{N(a_w+0.5w_m)w}{w_m(a_w+w)} & \text{if } w > 0 \\ \frac{N}{2} + \frac{N(a_w+0.5w_m)w}{w_m(a_w-w)} & \text{otherwise} \end{cases} \quad (1)$$

where n_d represents the most active node, w is the angular velocity of the robot. N is the number of nodes in the map, w_m is maximum angular velocity and, a_w controls the steepness of the sigmoid.

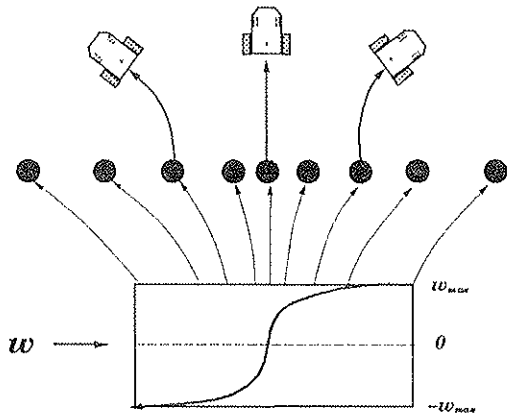

Figure 3: Angular velocity map. Each spatial position represents an angular velocity developed by the robot. The transformation has been performed by means of a sigmoidal function which permits more density of nodes for velocity values close to zero.

In our simulations, instead of only activating a single node in the population, we activate a neighborhood of nodes, with the position of maximal activation corresponding to the preferred angular velocity or direction of movement, and the activation of nearby nodes falling of as a Gaussian. We will show later that this form of distributed activity lends itself well to the problem of obstacle avoidance.

The mobile robot used in the simulations has eight ultrasonic sensors angularly distributed every 25.5°, covering the frontal plane of the robot as figure 2 shows. In the simulations we have assumed a maximum range of $5m$ for each sensors, and that a collision occurs when any of the sensors measures a distance under $1m$. An alternative solution in a practical situation could utilize information from bump sensors positioned on along the robot's perimeter.

However, it is important to point out that the conditioning model has no knowledge of the robot's geometry or of the location of the sensors, as we will show below.

In figure 4 the proposed model for obstacle avoidance is shown. In this case each sensory cue (or CS) corresponds to the signal from an ultrasonic sensors, subtracted from the maximum value of each ultrasonic sensor, so that closer obstacles are represented by larger activity. The unconditioned stimulus (US) in this case corresponds to a collision detected by the robot. For simplicity, rather than treating the collision detector as a UCS with a strong, fixed connection to the drive node, we let the collision signal activate directly the drive node. This corresponds to the situation discussed above in which an internal "survival" signal is always impinging upon the drive node. This simplification is reasonable as long as a single kind of UCS and drive are necessary.

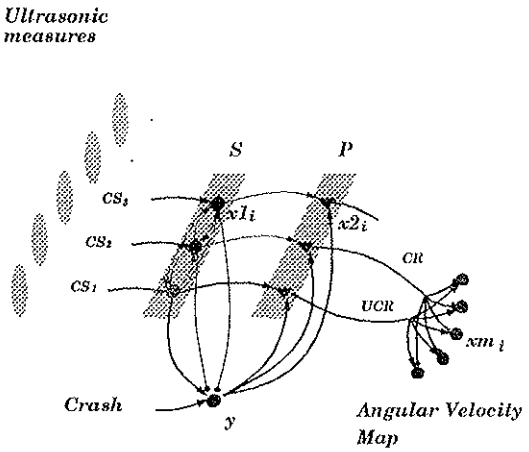

Figure 4: Conditioning model for obstacle avoidance. The ultrasonic range data represents the conditioned stimuli; the crash is the unconditioned stimulus. After conditioning, the pattern of activity across the ultrasonic sensors can predict a collision and change the angular velocity to avoid the obstacle.

The overall idea behind the model in figure 4 is that whenever the robot collides with an obstacle, learning in the circuit will create a connection between the current pattern of ultrasonic sensor activity and the angular velocity that the robot had at the time of collision. Later activation of a similar sensor pattern will cause inhibitory activation of those angular velocities that would have caused a crash, causing the robot to change direction and steer away from an obstacle.

One of the main properties of the model is its real-time function, in the sense that it is not necessary to separate explicitly learning from normal operation. However, in order to achieve a faster learning we have performed an initial learning phase where the robot performs random exploratory movements in a cluttered, activating sequentially each node from the angular velocity map, and thus sampling various patterns of activity that lead to collision.

As the robot travels, it takes measures from the ultrasonic sensors. The complementary values are contrast-enhanced and stored within the STM field S. The population S, which was originally modeled by Grossberg as a *recurrent competitive field*, removes the inherent noise from the ultrasonic sensors and enhances the activity of those sensors receiving maximal activation, that is, those registering the closest obstacles. A more complete description of the properties of this kind of network can be found elsewhere (Grossberg, 1973, 1982). We have used a simplified discrete time version of the recurrent competitive field, which quickly and efficiently normalizes and contrast-enhances the sensor activations. Specifically, the activation x_{1i} of each neuron in population S is given by

$$x_{1i}(t) = M_x \frac{[R - I_i(t)]^+}{\sum_i x_{1i}(t)} - (1 - M_x)x_{1i}(t-1) \tag{2}$$

where I_i is the "raw" ultrasonic measures, R is the maximum range for each ultrasound, M_x is a constant that determines how much the previous activation is weighted relative to the current input, and the notation $[x]^+$ represents half-wave rectification (returns only those values of $x > 0$). The summation is taken over all ultrasonic activations, thus ensuring normalization over the entire population S.

Prior to conditioning the drive node D is activated when the robot collides against obstacles. After conditioning, sufficient activation of a pattern of sensors can also activate the drive node. This permits second order conditioning, so that after the initial learning stages the ultrasonic measures can themselves predict the collision, and lead to conditioning of other sensor patterns. The activation y of the drive node is given by:

$$y(t) = \sum_i x_{1i}(t) z_{1i}(t) - T_y + U_{CS}(t) \tag{3}$$

where U_{CS} represents the collision of the robot ($U_{CS} = 1$ if collision just occurred, and $U_{CS} = 0$ otherwise), z_{1i} the adaptive weight connecting the sensory node x_{1i} to the drive node, and T_y is a threshold that controls how easily the drive node is activated, and thus indirectly controls how easily second order conditioning will occur.

The activation x_{2i} of polyvalent cells is given by:

$$x_{2i}(t) = x_{1i}(t) f(y(t)) \tag{4}$$

where $f(y(t))$ is given by:

$$f(y(t)) = \begin{cases} 1 & \text{if } y(t) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Two different kinds of learning take place: the learning that couples sensory nodes (ultrasounds) with the drive node (the collision), and the learning of the angular velocity pattern that existed just before the collision. The first type of learning follows an associative learning law with decay:

$$z_{1i}(t) = L z_{1i}(t-1) + P x_{1i}(t) f(y(t)) \tag{6}$$

where P is the learning rate, L is the weight decay.

The equation which learns the velocity map is also of an associative form:

$$zm_{i,J}(t) = zm_{i,J}(t-1) - M x_{2i}(t)[xm_J + zm_{i,J}(t)] \tag{7}$$

where $zm_{i,j}$ represents the adaptive weights from the polyvalent cells to the nodes within the angular velocity map, M is the learning rate, and J is the winner node in the angular velocity map. However, in this case the learned weights are negative, thus when the robot collides against obstacles, the above rule learns to inhibit the actual direction of movement. Note that this learning rule is equivalent to learning a pattern of activity over a map of neurons that are coupled through mutual inhibitory connections with the angular velocity map. The use of negative connection weights is computationally more parsimonious.

Once learning has occurred, the activation of the angular velocity map is given by two components. A first excitatory component reflects the angular velocity required to reach the target without the influence of obstacles. The second, inhibitory component (because the weights are negative), moves the robot away from the obstacles as a result of the activation of ultrasound signals in the conditioning circuit. The equation that reflects this behavior is given by;

$$xm_j(t) = \exp\left[-(j - n_d(t))^2/\sigma\right] + \sum_i x_{2i}(t) zm_{i,j}(t) \tag{8}$$

where $n_d(t)$ is the index of the node that represents the desired angular velocity to reach the target without obstacles, and σ is the variance of the Gaussian.

The reason for using a Gaussian distribution of activity is depicted in figure 5. When an excitatory Gaussian is combined with an inhibitory Gaussian at a slightly shifted position, the resulting net pattern of activity exhibits a maximum peak that is shifted in a direction *away* from the peak of the inhibitory Gaussian. Hence the presence of an obstacle to the left causes the robot to shift to the right, and *vice versa*.

In an earlier paper we have described an adaptive neural controller that utilizes distance and angle information to move the robot. In this example for simplicity we use the kinematic equations directly to move the simulated robot. Specifically, for a given pattern of activity of the angular velocity nodes, we select an angular velocity according to:

$$w = \begin{cases} \frac{w_m a_w [\max_j(xm_j(t)) - N/2]}{N(a_w + 0.5 w_m ax) - w_m [\max_j(xm_j(t)) - N/2]} & \text{if } \max_j(xm_j(t)) > N/2 \\ \frac{w_m a_w [\max_j(xm_j(t)) - N/2]}{N(a_w + 0.5 w_m ax) + w_m [\max_j(xm_j(t)) - N/2]} & \text{otherwise} \end{cases} \tag{9}$$

This function is essentially the inverse of the sigmoid described by equation 1 above, where $\max_j(xm_j(t))$ represents the node number with the largest activity in the angular velocity map.

4 Experimental results

In this section we show our preliminary results on the model's performance. In a first stage, we let the model develop a set of weights by letting the robot perform movements at different angular velocities in an environment cluttered with obstacles, by activating sequentially the nodes in the angular velocity map. After this initial learning phase the robot is able to avoid obstacles in arbitrary positions.

Figure 6 shows the projections of the adaptive connections between the sensory nodes x_2 and the angular velocity map xm. In the figure you can see for example how angular velocities that make the robot turn to the right

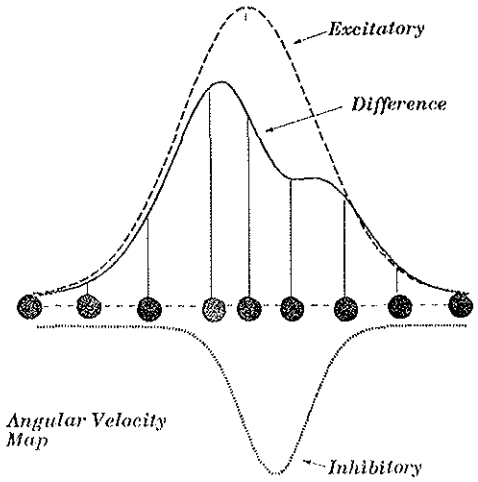

Figure 5: Positive Gaussian distribution represents the angular velocity without obstacles, and negative distribution represents activation from the conditioning circuit. The difference represents the angular velocity that will be used to drive the robot. Notice how the maximum of the excitatory Gaussian is shifted by the inhibitory Gaussian.

(nodes close to 20 in the figure) are inhibited when the robot receives ultrasonic sensor signals to the right (values close to 7).

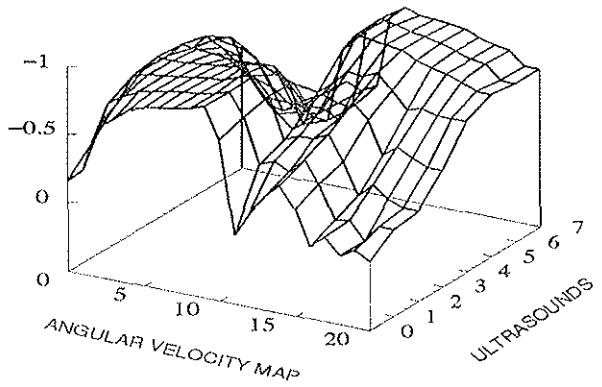

Figure 6: Representation of the adaptive connections $x_2 \rightarrow xm$. This map intrinsically codes the location of the ultrasonic sensors, in such way that when echoes are received in a given direction, that direction is inhibited in the angular velocity map.

Figure 7 illustrates the the robot's performance in the presence of several obstacles. The robot starts from the initial position labeled 1 and reaches a sequence of points 2-7 along the path shown by the dashed line. During the movements, whenever the robot is approaching an obstacle, the inhibitory profile from the conditioning circuit changes the selected angular velocity and makes the robot turn away from the obstacle. The presence of multiple obstacles at different positions in the robot's sensory field cause a complex pattern of activation that steers the robot between obstacles.

5 Conclusions

In this article we have described preliminary results with a model that learns obstacle avoidance for a wheeled mobile robot by means of ultrasonic information learned in a conditioning paradigm. The robot progressively learns to avoid the obstacles without the necessity of external supervision, but by negative reinforcement signals produced by the collision of the robot. One of the main properties of the model is that it is not necessary to know the robot's geometry nor the configuration of ultrasonic sensors on the robot's surface, because the robot learns from past experiences to avoid directions of movement that make the robot collide against the obstacles.

We are extending this models of conditioning to develop more complex behaviors. In particular, we are investigating conditioning circuits that permit the robot to choose among different behaviors (avoid, escape, wall following, etc.) depending on the moment-by-moment combination of sensorial information and its internal necessities. For example, a more complex system of sensory and drive nodes could be used to modulate how much

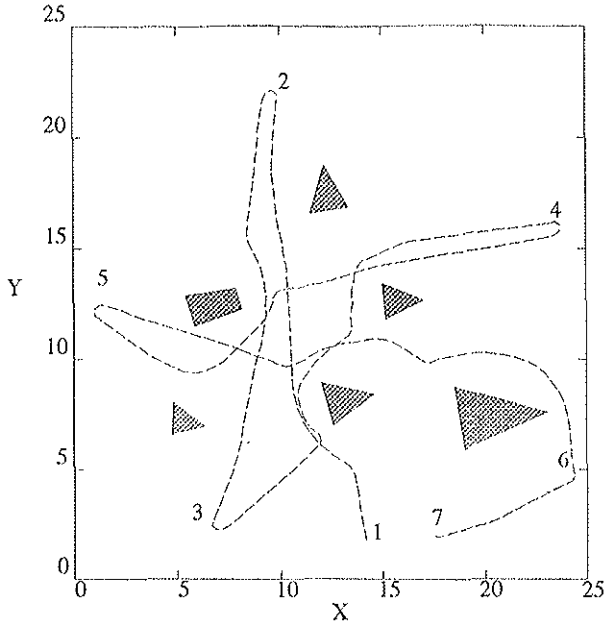

Figure 7: Trajectory followed by the robot in presence of six obstacles (shaded polygons). The robot travels from point 1 to point 7. Distances are expressed in meters. Parameters employed in simulations: $N = 21, w_m = 2.0, a_w = 0.3, M_x = 0.6, R = 5.0, T = 0.02, T_y = 0.2L = 0.9, P = 0.1, M = 0.05\sigma = 80$

the robot will try to avoid obstacles depending on its necessity to recharge its batteries.

References

Grossberg, S. (1971). On the dynamics of operant conditioning. *Journal of Theoretical Biology, 33*, 225–255.

Grossberg, S. (1982). A psychophysiological theory of reinforcement, drive, motivation and attention. *Journal of Theoretical Neurobiology, 1*, 286–369.

Grossberg, S., & Levine, D. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: blocking, interstimulus interval, and secondary reinforcement. *Applied Optics, 26*, 5015–5030.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52*, 217–257.

Grossberg, S. (Ed.). (1982). *Studies of Mind and Brain: neural principles of learning, perception, development, cognition and motor control*. Reidel, Boston.

Grossberg, S. (Ed.). (1986). *The Adaptive Brain I: Cognition, Learning, Reinforcement, and Rhythm*. Elsevier/North-Holland, Amsterdam.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H., & Prokasy, W. F. (Eds.), *Classical Conditioning II*, chap. 3, pp. 64–99. Appleton, New York.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review, 88*, 135–170.

Zalama, E., Gaudiano, P., & López-Coronado, J. (1995). A real-time, unsupervised neural network for the low-level control of a mobile robot in a nonstationary environment. *Neural Networks, 8*, 103–123.