CONNECTIONIST "FACE"-OFF: DIFFERENT ALGORITHMS FOR DIFFERENT TASKS

Dominique Valentin*
Hervé Abdi*†
Betty G. Edelman*
Annelies Nijdam*

*The University of Texas at Dallas, and †Université de Bourgogne à Dijon

ABSTRACT

We present a series of simulations contrasting the ability of a Hebbian and a Widrow-Hoff trained autoassociative memory to perform several face processing tasks in different learning and testing conditions. We show that all face processing tasks are not equally demanding, and that, in some particular circumstances, a simple learning algorithm can be more appropriate than a complex one. We also illustrate that the choice of evaluation criteria for assessing the performance of a model is crucially dependent on the task performed, especially when trying to predict human behavior. Finally, we speculate about the complexity of early learning (i.e., by infants) of faces, and suggest that this task is easier than has been generally thought.

1. Introduction

Since Kohonen (1977) first demonstrated that an autoassociative memory can be used to store and retrieve faces, this type of model has been successfully applied to diverse face processing tasks (see Abdi, 1988 for an early example and Valentin, Abdi, O'Toole, & Cottrell, 1994 for a review). Human faces are one of the most important biological and social objects in our environment. They provide us with a key not only to categorize, recognize and identify other human beings, but also to interpret their mood and emotional states.

In recent years, empirical and computational models have increased our awareness of the complexity underlying face processing tasks. This complexity stems essentially from the perceptual properties of faces as visual stimuli. Human faces are highly similar, they all have two eyes, a nose, and a mouth arranged roughly in the same configuration. Yet, inspite of only subtle differences, human observers are able to categorize, recognize and identify a large number of faces. In this paper, we further examine the intricacies inherent to these tasks by training an autoassociative memory to perform several face processing tasks using two learning algorithms: the Hebbian and Widrow-Hoff algorithms. These two learning algorithms vary in their level of complexity, and thus their relative performance provides us with some insight into the varying demands of the tasks performed. Specifically, by comparing their performance on similar tasks, we can evaluate the nature and/or type of processing required by the tasks. Our simulations show that all face processing tasks are not equally demanding, and that in some particular circumstances, a simple learning algorithm may be more appropriate than a complex one. We also illustrate the importance of the evaluation criteria selected to assess the performance of a model. Finally, we demonstrate how misleading some common performance evaluation criteria can be when trying to predict human behavior.

This paper is organized as follows: After describing the autoassociative memory model, we present a series of simulations contrasting the performance of the two learning algorithms for several tasks in different learning and testing conditions. We have tried to keep our discussion as non-technical as possible, relegating technicalities to footnotes for aficionados. These technicalities can be skipped without any problem.

2. Autoassociative memories

Autoassociative memories are a special case of associative memories (Anderson, Silverstein, Ritz, & Jones, 1977) in which the input patterns are associated with themselves. They are content-addressable memories, able

to reconstruct a whole pattern of information from part of it. For example, such memories can reconstruct face images from partial or degraded input (Kohonen, 1977).
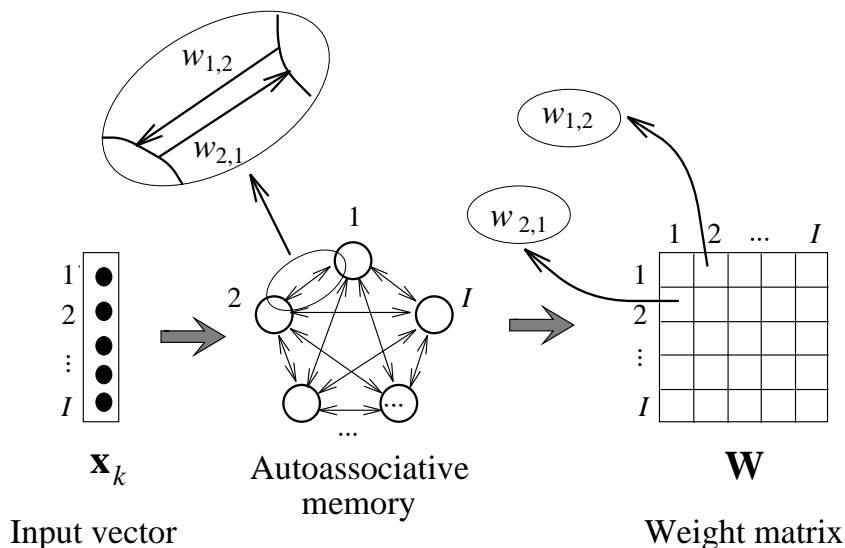

FIGURE 1. Illustration of an autoassociative memory

Formally, an autoassociative memory is a network of I units or cells fully interconnected by modifiable connections or synapses (see Figure 1). The set of connections is represented by an $I \times I$ matrix $\mathbf{W}$, in which a given element $w_{i,j}$ represents the strength of the connection between cell i and cell j. The connections are generally symmetrical (i.e., $w_{i,j}=w_{j,i}$). Each of the K objects to be stored in an autoassociative memory is represented by an I-dimensional vector denoted $\mathbf{x}_k$. Each element in this vector is input to one cell of the autoassociative memory. Learning results from the modification of the strength of the connections following the presentation of a set of patterns. These modifications can be implemented using two different learning rules. The first one, the *Hebbian learning rule*, requires minimal computation. It is generally described as "unsupervised", because no feedback concerning the correctness of a response is provided during the learning process. The second one, the *Widrow-Hoff learning rule*, requires more computation and time. It is often considered a "supervised" form of learning because its usual implementation[1] involves feedback about the similarity between the response provided by the model and a desired target.

---

[1]Other implementations of the Widrow-Hoff learning rule (e.g., explicit sphericization of the weight matrix obtained by setting the eigenvalues to unity) make the distinction between supervised and unsupervised learning less clear for autoassociative memories than

2.1. *Hebbian learning*

Hebbian learning is based on work by Hebb (1949) who proposed that learning is a purely local phenomenon expressible in terms of synaptic change. Precisely, he theorized that the synaptic change depends on both presynaptic and postsynaptic activities:

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency as one of the cells firing B, is increased (p. 62).

In essence, Hebb's learning rule states that the change in synaptic strength is a function of the "temporal correlation" between presynaptic and postsynaptic activities. Specifically, the synaptic strength between two neurons A and B increases whenever the two neurons fire simultaneously, and decreases when the two neurons fire independently.

In a linear associator, the Hebbian learning rule sets the change of the connection weights to be proportional to the product of the input and the output of the cells. For an autoassociator, the value of the weight of the connection between two cells i and j is proportional to the activity of cell i and cell j.

More formally, a set of K patterns is stored in a linear autoassociative memory by multiplying each pattern vector by its transpose and summing the resulting outer-product matrices. This learning rule can be defined[2] as:

$$(1) \qquad \mathbf{W} = \gamma \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^T$$

where T denotes the transposition operation and γ the proportionality constant. Usually, for convenience, γ is set to 1. We shall follow this tradition in this paper.

Retrieval of a pattern, such as a face image, is performed by presenting the pattern vector as input to the memory. Specifically, recall of the k-th

---

for heteroassociative memories. However, the important point is that, regardless of the implementation, Widrow-Hoff learning is always more computationally intensive than Hebbian learning.

[2] A possible problem with Equation 1 is that repeated application of the rule can lead to unbounded growth of the weights. There are other, more sophisticated, ways of defining Hebbian learning that can avoid this problem. For the set of simulations reported here, they will give the same pattern of results, but Equation 1 offers the great advantage of simplicity and mathematical tractability.

pattern is achieved as

$$(2) \qquad\qquad \widehat{\mathbf{x}}_k = \mathbf{W}\mathbf{x}_k$$

where $\widehat{\mathbf{x}}_k$ represents the response of the memory. One way to estimate the quality of this answer is to compute the squared correlation between $\widehat{\mathbf{x}}_k$ and $\mathbf{x}_k$ ($r^2_{\widehat{\mathbf{x}}_k, \mathbf{x}_k}$ or simply r^2). This index gives the proportion of common variance between the original and the reconstructed pattern and so, r^2 equals to 1 indicates perfect reconstruction of the input pattern.

When all the input vectors stored in the memory are mutually orthogonal, recall is perfect. If the input vectors are not orthogonal, as is the case with faces for example, the response of the memory will be degraded by the interference[3] created from similar input patterns.

2.2. *Widrow-Hoff learning*

The Widrow-Hoff learning rule (Duda & Hart, 1973), also known as the *Delta* rule (McClelland & Rumelhart, 1986) adjusts the weights of the connection matrix in order to maximize the quality of reconstruction of the input patterns[4]. The values of the weights are first computed using Hebbian learning and are then iteratively corrected using the difference between the input pattern and the pattern reconstructed by the memory. Specifically, the weight matrix is expressed at iteration $n + 1$ as:

$$
\begin{aligned}
\mathbf{W}_{[n+1]} &= \mathbf{W}_{[n]} + \eta(\mathbf{x}_k - \mathbf{W}_{[n]}\mathbf{x}_k)\mathbf{x}_k^T \\
&= \mathbf{W}_{[n]} + \eta(\mathbf{x}_k - \widehat{\mathbf{x}}_k)\mathbf{x}_k^T \\
(3) \qquad &= \mathbf{W}_{[n]} + \mathbf{\Delta}\mathbf{w}_{[n]}
\end{aligned}
$$

where n represents the iteration number, η is a small positive constant, and k is randomly chosen for each iteration. The weight matrix is updated at time $n + 1$ by computing the difference between the estimation produced by the memory (i.e., $\widehat{\mathbf{x}}_k$) and the original face vector $\mathbf{x}_k$, and by re-teaching this difference to the memory. This process is repeated across patterns until all the patterns in the learning set are reconstructed satisfactorily.

---

[3]The memory will add noise (or crosstalk) to the correct pattern (see, e.g., Kohonen, 1977 for more details).

[4]Technically, this method is equivalent to using a gradient descent technique to minimize the mean square error of reconstruction (see, e.g., Abdi, 1994a, chap. 4 and appendix for technical details).

3. Simulations

We have seen that two different types of learning can be used to train an autoassociative memory: Hebbian and Widrow-Hoff learning. In spite of its computational demands, Widrow-Hoff learning is more frequently used than Hebbian learning because of its superior performance. In this paper, we examine further the performance of Widrow-Hoff learning, as compared to the simpler Hebbian approach, to determine if the computational cost of Widrow-Hoff learning is always justified. The following simulations examine the influence of different factors on the performance of a Hebbian and a Widrow-Hoff trained autoassociative memory. We begin by looking at the relative performance of the two learning algorithms across several face processing tasks typically performed by human subjects. Next, we present simulations that manipulate factors relating to both the number and the perceptual characteristics of the stimuli. The end goals of these simulations are to illustrate the differential complexity inherent to several face processing tasks and to analyze the respective adequacy of Hebbian and Widrow-Hoff learning in relation to these tasks.

3.1. *Simulation 1*

In this simulation, we examine and contrast the ability of a Hebbian and a Widrow-Hoff trained autoassociative memory to perform several categorization tasks. Faces constitute an interesting stimulus for this type of evaluation due to both their perceptual properties and to the variety of categorization tasks available. Faces contain significant global configural information (e.g., general shape, outer contour, hairline) and detailed featural information (e.g., specific shape of the eyes, nose and mouth) that coexist. For human subjects, global and detailed information appear to be differentially useful for different face tasks (Sergent, 1986a and b). Global information seems useful to determine that a stimulus is a face ("faceness"), or that it belongs to a particular class of faces determined by a particular race or sex. Detailed information seems useful to decide whether a stimulus is a familiar ("learned" or "old") or unfamiliar ("new") face. The purpose of this simulation is to determine whether this difference in task requirements observed for human subjects can appear also in a computational model. By contrasting the performance of the two learning algorithms on different categorization tasks, we aim to determine if Widrow-Hoff learning is always essential, or if some of these tasks can be successfully accomplished by the simpler Hebbian learning.

3.1.1. *Method*

Stimuli: A total of 320 faces were digitized from slides as $225 \times 151 = 33975$ pixel images with a resolution of 16 gray levels. Half of the faces were Caucasian (80 males and 80 females). The other half were Japanese (80 males and 80 females). None of them had major distinguishing characteristics, such as facial hair, glasses or jewelry. The images were roughly aligned along the axis of the eyes, so that the eyes of all faces were approximately at the same height. For computational convenience, faces were compressed by local averaging using a 5×5 window giving $45 \times 31 = 1395$ pixel images. A control simulation carried out on the full images on a sample of faces showed that no essential information for the task simulated here is lost in the compression process (i.e., the same patterns of results emerge from both types of images). For all the following simulations, each face was coded as a 1395 pixel intensity vector obtained by concatenating the columns of its compressed image.

Experimental design: Two independent variables were manipulated. The first variable was the *type of stimulus* presented at test with four levels: 1) faces learned by the memory; 2) faces that had not been learned by the memory, but were from the same race as the learned faces; 3) faces that had not been learned by the memory and were from a different race than the learned faces; and 4) random patterns. The second variable was the *learning rule* used to train the autoassociative memory: Hebbian versus Widrow-Hoff. The dependent variable used to estimate the performance of the model was the squared coefficient of correlation computed between the images presented at test and the output of the memory.

Procedure: Two autoassociative memories were trained to reconstruct a set of Caucasian faces[5] and tested for their ability to reconstruct either the same faces, some new faces (same and different race) and random patterns. The first autoassociative memory was trained with Hebbian learning and the second one with Widrow-Hoff learning.

To test the ability of the memories to generalize to new Caucasian faces (new faces-same race condition) we used a leave-one-out jackknife technique. Training was done using 159 Caucasian faces, leaving out one Caucasian face for testing. At the end of training, the weights of the connection matrix were fixed. The remaining face was used as a memory key and the squared coefficient of correlation between the input and the output pattern was computed.

---

[5]A control simulation was carried out with Japanese faces as a training set and Caucasian faces as a testing set for the different race condition. No differences were observed between the two simulations.

This procedure was executed 160 times so that each face was used, in turn, as the testing face.

To evaluate the performance of the memories for the learned faces, as well as their ability to generalize to new faces from a different race, and to discriminate faces from random patterns, we used 160 Caucasian faces as training sets for the two autoassociative memories. The same faces, 160 Japanese faces, and 160 random patterns were then used to cue the memories and the quality of reconstruction was measured by the squared coefficient of correlation between the image presented at test and the image re-created by the memory.
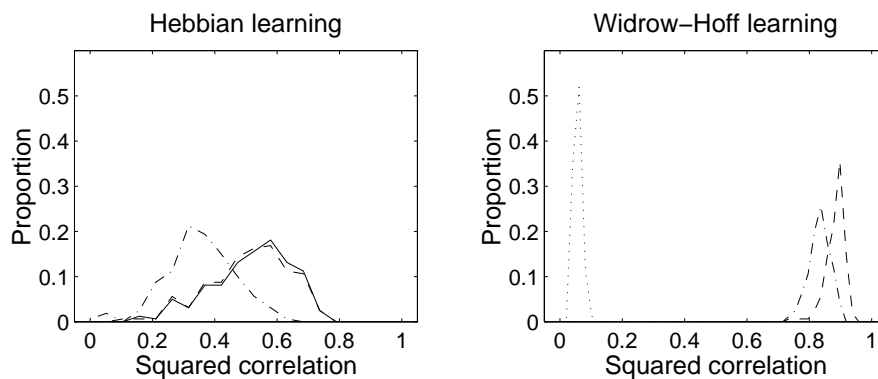
3.1.2. *Results and discussion*

FIGURE 2. Distribution of the r^2 obtained for the Hebbian memory (left panel) and the Widrow-Hoff memory (right panel) when 1) a learned face (solid line), 2) a new face from the same race as the learned faces (dashed line), 3) a new face from a different race than the learned faces (dashed-dotted line), and 4) a random pattern (dotted line) was presented at test. Note that, for readability purposes, the random pattern condition was omitted from the Hebbian learning graph and the learned condition from the Widrow-Hoff learning graph as these conditions led to constant performance (0 for the random condition with Hebbian learning and 1 for the learned condition with Widrow-Hoff learning).

The distributions of the squared coefficients of correlation obtained in the different experimental conditions are presented in Figure 2. The left panel of this figure represents the distributions obtained with the Hebbian trained memory, and the right panel the distributions obtained with the Widrow-Hoff trained memory. To assess the discrimination power of the two learning

algorithms, we generated a *Receiver Operating Characteristic* (ROC) curve[6] from the r^2 distributions of each type of stimuli in each learning condition. This was done by using increasing values of r^2 as a criterion. Faces with an r^2 larger than the criterion were categorized as "old" and faces with an r^2 smaller than the criterion as "new". The hit and false alarm rates were obtained following standard signal detection methodology.

The results for Hebbian learning (left panel) show that, the overall r^2 distributions for learned (solid line) and new (dashed line) Caucasian faces overlap almost completely (area under ROC=0.52), although a given face is always slightly better reconstructed in the learned condition than in the new condition[7]. This overlap shows that the memory is not able to distinguish between learned and unlearned (i.e., to recognize) faces when the new faces are similar to the learned faces (i.e., other Caucasian faces). On the other hand, when the new faces are not similar to the learned faces (i.e., Japanese faces), the r^2 distribution (dashed-dotted line) differs from that of the learned Caucasian faces (area under ROC= 0.83) and new Caucasian faces (area under ROC= 0.82). Finally, the distribution for the random pattern (not shown in Figure 2) is very different from that obtained for any of the face stimuli (area under ROC=1). So, in conclusion, when Hebbian learning is used, the memory can discriminate a face from a random pattern, can "categorize" the faces according to some global characteristic such as race, but cannot recognize within this category the faces it has learned.

The distributions obtained with Widrow-Hoff learning show better face reconstruction than Hebbian learning under all conditions[8]. Moreover, because the performance for learned faces (not shown in Figure 2) is always perfect, the learned Caucasian and new Caucasian distributions do not overlap (area under ROC=1). The differences between the other distributions remain similar to those obtained with Hebbian learning. Specifically, the Widrow-Hoff trained memory discriminates only slightly better between new Japanese and new Caucasian faces than does the Hebbian trained memory (area under ROC =.88 and .82, respectively).

---

[6] The area under this curve is typically used as an unbiased estimate of the proportion of correct discrimination.

[7] The difference between a face seen as old or new is indeed real and genuine. However, the size of the overall effect (i.e., .02) would be judged meaningless by most researchers in the field.

[8] Note that because of overfitting problems, Widrow-Hoff learning will always lead to a r^2 of 1 for learned faces when the number of faces in the training set is smaller than the number of pixels (see, e.g., Duda & Hart, 1973).

This first simulation suggests that Hebbian learning is a very primitive learning algorithm, having the advantage of requiring only simple computations, but being limited in its applications with highly correlated stimuli like faces[9]. This type of learning might be useful for simulating some simple face categorization tasks based on global information (e.g., to differentiate between a face and a random pattern or between two faces of a different race), but seems less adequate to simulate tasks requiring the processing of subtle information (e.g., to distinguish between two faces of a same race. In the following simulations we investigate under what, if any, circumstances Hebbian learning is capable of making this type of distinction.

3.2. *Simulation 2*

The previous simulation showed the superiority of Widrow-Hoff learning over Hebbian learning for distinguishing between familiar and unfamiliar faces. The relatively poor performance of Hebbian learning in this task is not really surprising, because, as mentioned before, Hebbian learning of nonorthogonal stimuli suffers from interference or crosstalk (Kohonen, 1977). This interference seems to become a major problem in the case of a large number of similar patterns such as faces. The empirical question that remains is: What is the precise relationship of the amount of interference to the size of the learning set? To answer this question we analyzed the ability of both a Hebbian and a Widrow-Hoff trained memory to distinguish between learned and new Caucasian faces as the number of learned faces increased.

3.2.1. *Method*

Stimuli: The same 160 Caucasian faces as in Simulation 1 were used to train and test two autoassociative memories.
Experimental design: Three independent variables were manipulated: the *number of face* in the training sets (2, 5, 10, 20, 30, 50, and 100), the *type of face* presented at test (learned or new same race) and the *learning rule* used to train the memories. The dependent variable was again the squared coefficient of correlation between the original image presented at test and the images produced by the memories.
Procedure: Varying numbers of faces (N) were randomly selected from the original set of 160 Caucasian faces and used as training sets for two autoassociative memories. The remaining faces were used to test the ability of the memories to generalize to new faces. To estimate the effect of the number of learned faces on the performance of the memories, we used seven

---

[9]With other less correlated stimuli, the performance of Hebbian learning would be better than the one found here.

values of N: 2, 5, 10, 20, 30, 50, and 100. Note that the value $N = 1$ was not used, because, in this case, the performance is trivially perfect for learned faces, no matter which learning algorithm is used. One hundred random samples were drawn for each N condition to serve as learning sets. For each sample, in each N condition, two autoassociative memories were trained to reconstruct the sample faces. As previously, the first memory was trained using Hebbian learning and the second one using Widrow-Hoff learning. After training, the weights of the memories were fixed and all 160 Caucasian faces were presented as memory cues. For each face, the squared correlation between the original faces and the response of the memories was computed.
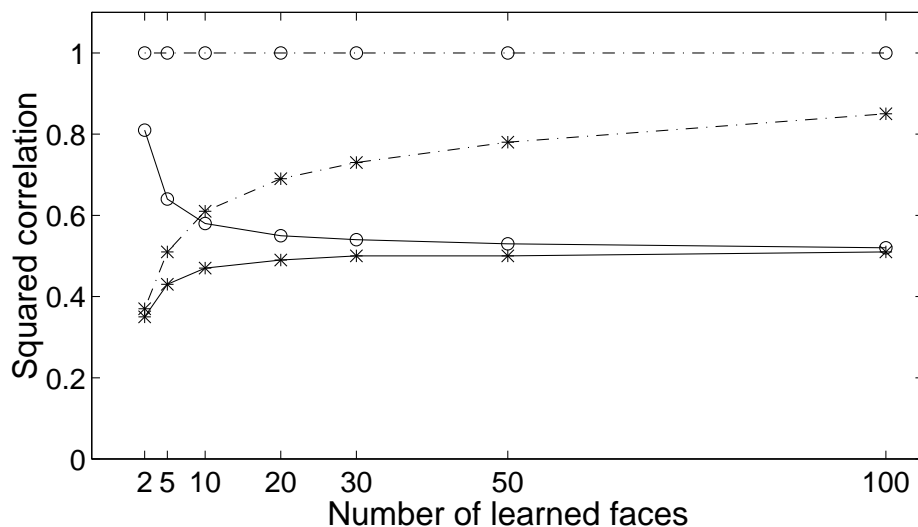
FIGURE 3. Average r^2 represented as a function of the number of faces in the training sets, the learning rule used to train the memories (the solid line represents the Hebbian rule and the dashed line [like the "-" of Widrow-Hoff] the Widrow-Hoff learning rule) and the type of face (learned faces are represented by an "o" [like old] and new faces by a star).

3.2.2. Results and discussion

Figure 3 displays the average r^2 as a function of the number of faces in the training set, the learning rule used to train the memory, and the type of face presented at test. For the Hebbian trained memory, the correlation between original and reconstructed faces decreases with the number of faces in the

training set for the learned faces and increases for the new faces. An important point to note is that up to about 20 faces, the learned faces are clearly better reconstructed than the new faces. After 20 faces, the quality of reconstruction of both learned and new faces converges toward the same value[10] of $r^2 = .5$. Therefore, when only a few faces are learned, a Hebbian trained memory can "recognize" the faces on which it has been trained. When the number of faces increases, the memory begins to suffer from interference and to "forget" the faces it was previously able to "recognize." In contrast, the performance of the Widrow-Hoff trained memory remains perfect for the learned faces. In addition, although the ability of the memory to generalize to new faces improves dramatically as the number of faces increases, it always reconstructs learned faces better. Thus, the memory is always able to "recognize" the faces on which it has been trained.

As an illustration of the effect of the number of training faces on the quality of reconstruction of learned and new faces, Figure 4 displays the responses produced by an autoassociative memory trained under several regimens when a particular face is presented as a memory key. The top and the middle rows show the response of the Hebbian trained memory when, from left to right, 1, 2, 5, 10, 20, and 159 faces were learned. The top row corresponds to the responses produced to an old face (Hebbian old condition), and the middle row corresponds to the responses produced to a new face (Hebbian new condition). The bottom row shows the responses of the memory to a new face after Widrow-Hoff learning (Widrow-Hoff new condition).

From this figure, we can see that, in the Hebbian old condition (top row), 1) when only one face was learned, the response of the memory is the old face; 2) the similarity between the test face and the response of the memory decreases as the number of faces in the training set increases; and 3) when 159 faces have been learned, the response of the memory converges to the average of the faces in the training set.[11] Similarly, we can see that, in the Hebbian new condition (middle row): 1) when only one face was learned, the response of the memory is the old face; and 2) when the number of training faces increases, the response of the memory again converges to the average face. Comparing the images displayed in the first two rows, shows that a memory trained with Hebbian learning produces different responses to old and new faces only when a small number of faces have been learned. In contrast, in the Widrow-Hoff new condition (bottom row) when the number of faces in

---

[10] The problem of deciding if this specific value of $r^2 = .5$ has some deep meaning is still a matter of speculation.

[11] Or more technically, to the first eigenvector of the set of faces (for more details on the notion of face eigenvectors see, e.g., Abdi, 1994a and b).

FIGURE 4. Responses of an autoassociative memory trained with Hebbian learning (top and middle rows) or Widrow-Hoff learning (bottom row) to the face displayed in the left top corner when 1, 2, 5, 10, 20, and 159 faces were used as a training set (from left to right). In the top row the testing face was an old face (included in the training set), in the middle and bottom rows, the testing face was a new face (not included in the training set). The reconstruction of the Widrow-Hoff trained model, when the face was learned, was always perfect and is not shown.

the training set increases, the memory response increasingly resembles the original test face.

In conclusion, it seems that Hebbian learning tends to capture what is common to all the faces, whereas Widrow-Hoff learning captures the information that is specific to individual faces. In more technical terms, Widrow-Hoff learning increases the dimensionality of the space on which the faces are meaningfully represented, whereas Hebbian learning represents the faces on an unidimensional space corresponding to the first eigenvector (or average) of the set of faces. Note, however, that this last result is due to the specific statistical structure of faces. In particular, the fact that faces are very similar to each other causes the average to explain a large part of the variance of a set of faces. Thus the average constitutes a relatively good way of describing a face as opposed to other objects, which explains why

Hebbian learning is able to perform categorization tasks for faces, but not discrimination tasks. Different results would be obtained for objects with a different statistical structure.

The difference between Hebbian and Widrow-Hoff learning might result from the differential ability of the two learning algorithms to process subtle information useful to discriminate faces within a given category (e.g., same race). The small adjustments occurring during Widrow-Hoff learning permit subtle differences between faces to be taken into account in the construction of the weight matrix. The larger the number of faces, the more details are captured by the weight matrix, and the better the memory generalizes. In contrast, since no adjustments occur during Hebbian learning, details are not captured in the weight matrix and all faces tend to look the same. The next simulation further examines the influence of details on the performance of both learning algorithms.

3.3. *Simulation 3*

One way to test the idea that the superiority of Widrow-Hoff over Hebbian learning is indeed due to the processing of detailed information is to examine its performance when the quality of input is altered by adding high frequency random noise to the stimuli presented at test. If Hebbian learning does not depend upon small variations in the stimuli, this alteration should not severely affect its performance. Moreover, if the added noise is sufficient to mask the information that distinguishes a given face from another face, Widrow-Hoff should be affected and its superiority should be diminished, and perhaps even vanish[12].

3.3.1. *Method*

Stimuli: The stimuli were the same as those used in Simulation 2.
Experimental design: Four independent variables were manipulated: the *number of training faces* (2, 5, 10, 20, 30, 50, and 100), the *type of learning rule* used to train the memory (Hebbian vs. Widrow-Hoff), the *type of face* presented at test (learned vs. new), and the *amount of noise* added to the images during testing (noise level 0 through 5).
Procedure: The procedure was similar to that used in Simulation 2, with the exception that, during the testing phase, a Gaussian noise component was added to each pixel of the 160 face images before presenting them as memory cues. The mean of the Gaussian distributions was equal to zero and

---

[12]As a side consideration, adding some random noise brings the task closer to the old problematic of mathematical learning theory. The relationship between this theory and neural networks is explored by Gluck (1992).
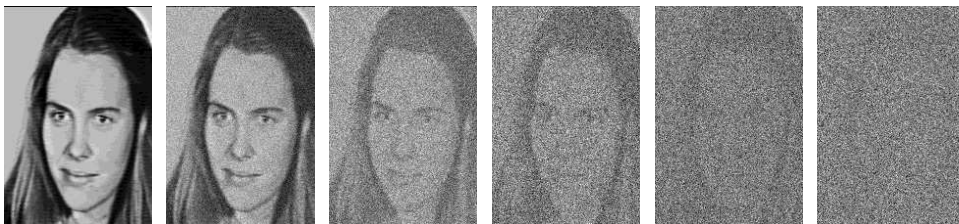
FIGURE 5. The different noise conditions used in Simulation 3 (the standard deviation of the noise is respectively: 0, .25, 1, 2, 5, and 10 times the standard deviation of the pixel distribution of the face image).

the standard deviation to 0, .25, 1, 2, 5, or 10 times the standard deviation of the pixel distribution of the face images for noise level 0, 1, 2, 3, 4, and 5 respectively. An illustration of the different noise levels is provided in Figure 5. As previously, the squared coefficient of correlation between original images (without noise) and reconstructed images was computed.

3.3.2. Results and discussion

Figures 6B, C, D, E, and F display the patterns of r^2 obtained for noise levels 1 through 5. For comparison, Figure 6A redisplays the results obtained in Simulation 2 when no noise (i.e., noise level 0) was added to the test face. Three points are worth noting from these figures. First, as expected, the performance of the Hebbian trained memory is not affected at all by the addition of noise. Second, the general ability of the Widrow-Hoff trained memory to reconstruct faces is increasingly affected by the addition of noise. The larger the amount of noise is, the poorer the quality of reconstruction of both learned and new faces. Moreover, the difference in quality of reconstruction between learned and new faces decreases as the amount of noise added to the test faces increases. The memory becomes less able to "recognize" the learned faces when placed in a noisy context. Third, the effect of noise on Widrow-Hoff learning increases with the number of learned faces. In summary, when noise is added during testing, the Widrow-Hoff trained memory behaves somewhat like the Hebbian trained memory: It suffers from interference (the more faces the memory learns, the poorer the performance). In fact, Figures 6E and F, show that, when enough noise is added, Hebbian learning outperforms Widrow-Hoff learning. Again, it is worth pointing out that this result is essentially due to the specific statistical structure of faces, and might not generalize to other stimuli.
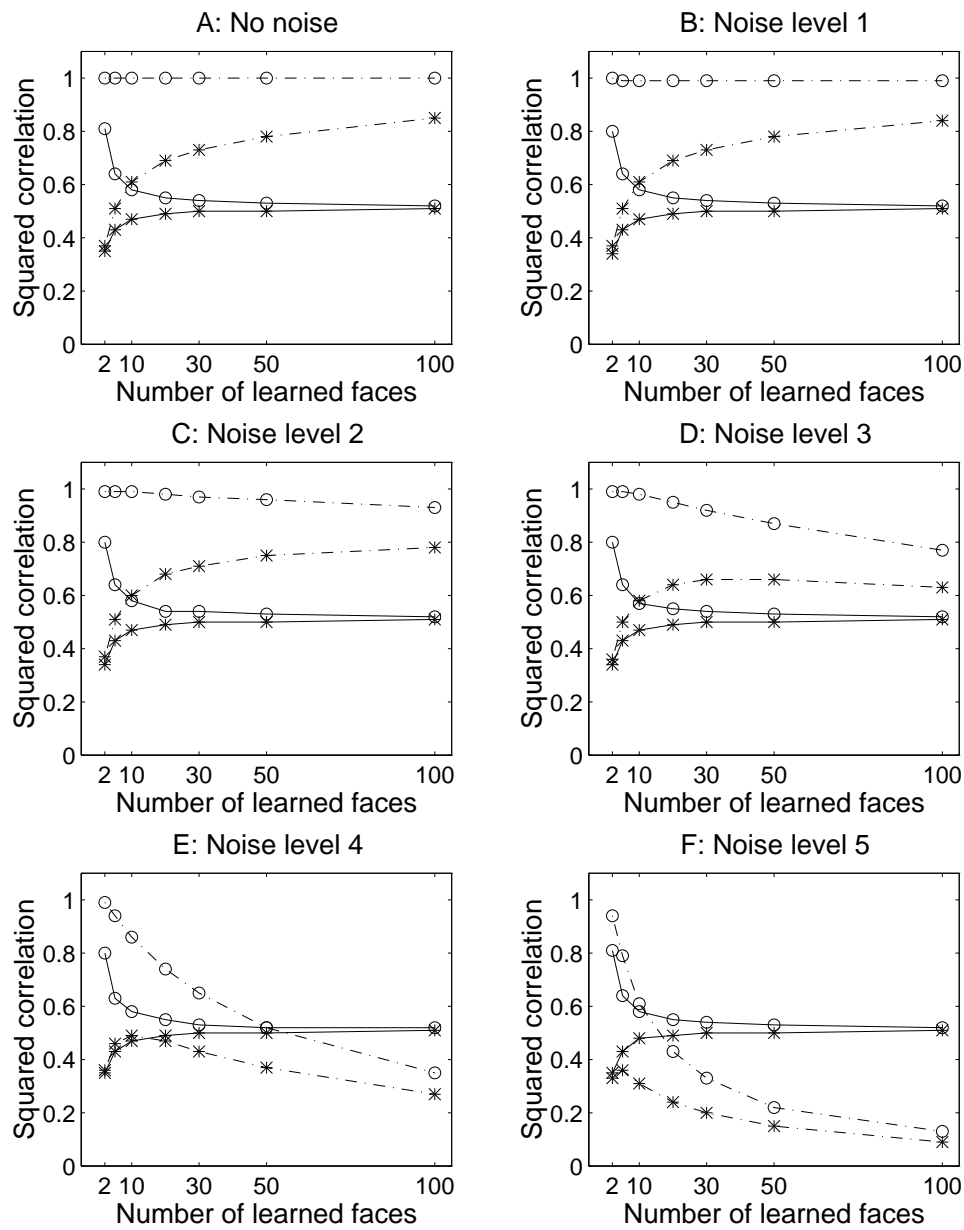
FIGURE 6. Average r^2 represented as a function of the number of faces in the training sets, the learning rule used to train the memories (the solid line represents the Hebbian rule and the dashed line the Widrow-Hoff learning rule) and the type of face (learned faces are represented by an "o" and new faces by a star). Figure 6A displays the results obtained in simulation 2 (noise level 0) and Figures 6 B, C, D, E, F display the results obtained in Simulation 3 when increasing amounts of noise were added to the face images in the testing phase. The noise levels are 0, .25, 1, 2, 5, and 10 times the standard deviation of the face pixel distributions.
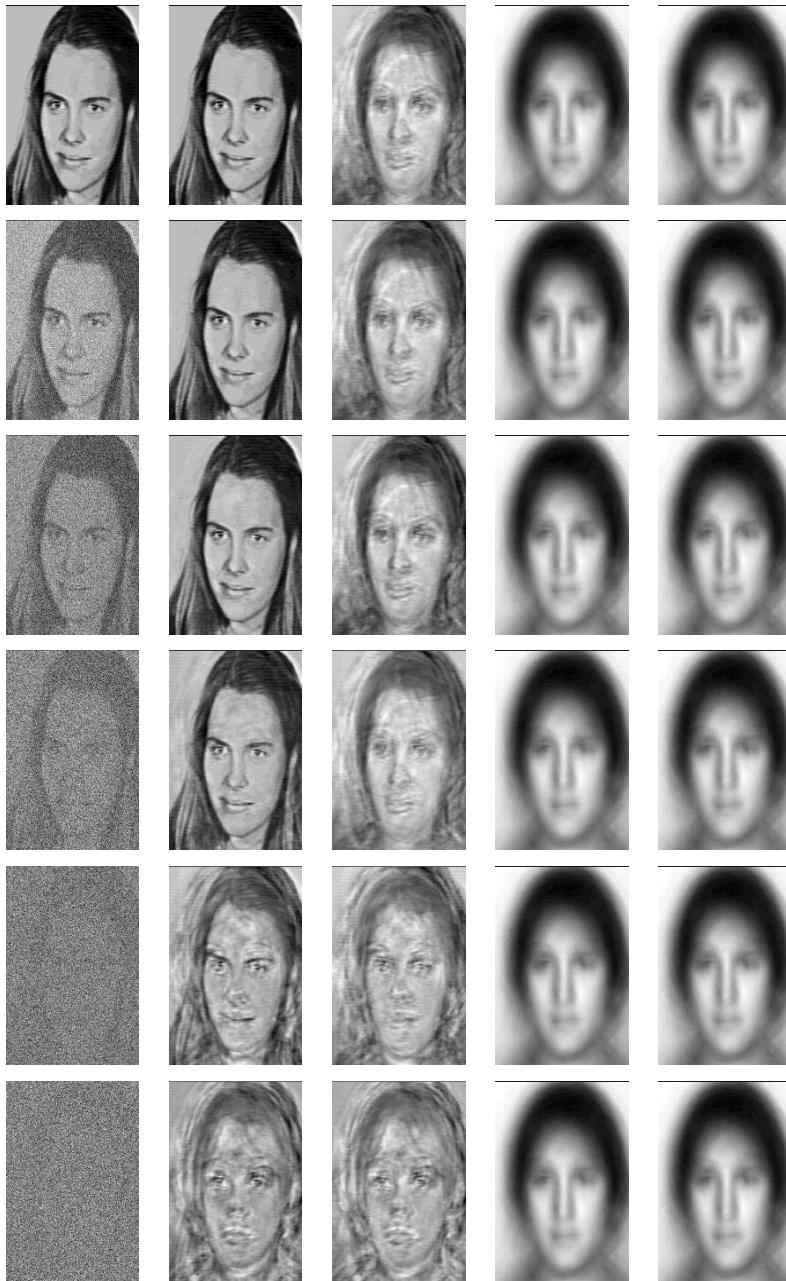
FIGURE 7. Responses of a Hebbian and a Widrow-Hoff trained memory to noisy faces. The first column shows a test face in different noise conditions, the following columns the response to this face in the 1) Widrow-Hoff old, 2) Widrow-Hoff new, 3) Hebbian old, and 4) Hebbian new conditions (see text for more details).

Figure 7 illustrates the effect of noise on the responses produced by a Hebbian and a Widrow-Hoff memory trained with 159 Caucasian faces. The first column displays a test face in the different noise conditions (from top to bottom, the standard deviation of the noise added to the image is respectively: 0, .25, 1, 2, 5, and 10 times the standard deviation of the pixel distribution of the face image). The second column displays the response of the Widrow-Hoff trained memory to this face when the face was learned by the memory (Widrow-Hoff old condition). The third column displays the response of the Widrow-Hoff trained memory when the face was not learned by the memory (Widrow-Hoff new condition). The fourth column displays the response of the Hebbian trained memory when the face was learned by the memory (Hebbian old condition). The fifth column displays the response of the Hebbian trained memory when the face was not learned by the memory (Hebbian new condition). Note that the Hebbian trained memory produces always the same response, regardless of the amount of noise added to the face presented as a memory key. In contrast, the response of the Widrow-Hoff trained memory differs as a function of the amount of added noise. The more noise is added, the more distorted the face image is, in both old and new conditions. Moreover, the larger the noise, the more similar to each other the responses in the old and new condition become. This result shows that, as predicted, Widrow-Hoff learning relies on detailed information, whereas this type of information is not essential for Hebbian learning. The following simulation examines further the type of perceptual information used by both learning algorithms.

3.4. *Simulation 4*

We have shown that Widrow-Hoff learning is sensitive to high frequency random noise, but Hebbian learning is not. This result is consistent with the idea that these two learning algorithms are based on two different types of information: global common information and detailed individual specific information. Previous work suggests that these different types of information are not conveyed by the same spatial frequency ranges (e.g., O'Toole, Abdi, Deffenbacher & Valentin, 1993; Sergent, 1986a and b; Valentin & Abdi, 1996). Information concerning global facial configuration common to many faces is conveyed by the low frequency ranges of a face image. In contrast, the high frequency ranges of a face image provide the highly detailed information necessary to discriminate between faces. From this dissociation between low and high frequencies, it is tempting to theorize that, whereas Hebbian learning is based essentially on the processing of low spatial frequency information, Widrow-Hoff learning can capitalize on high spatial

frequency information to perform optimally. To test this hypothesis, we
replicated Simulation 2 using, in addition to full spatial spectrum images,
low-pass[13] and high-pass[14] filtered images of the faces to train and test a
Hebbian and a Widrow-Hoff autoassociative memory (in a manner some-
what similar to O'Toole, Millward, & Anderson, 1988). This simulation
provides an indication of the differential usefulness of high and low spatial
frequency information for Hebbian and Widrow-Hoff learning, as well as the
ability of the two learning algorithms to generalize from one type of spatial
information to another.

FIGURE 8. Example of Simulation 4 stimuli. The left panel represents
an original face image (full frequency condition), the middle panel a
high-pass filtered version of the face image, and the right panel a low-
pass filtered version. The original image (left panel) is obtained by
adding the two right images.

3.4.1. Method

Stimuli: The same 160 Caucasian faces used in the previous simulations
were used as stimuli in this simulation. Three versions of each image were
created. In the first version (F), the full spatial spectrum was preserved. In
the second version (L), only the low frequency information was preserved.

---

[13] A low-pass filter eliminates high frequency components of an image while leaving the
low frequency components intact. Because high frequency components correspond to fast
changes in the characteristics (contrast and light intensity) of an image (e.g. edge, sharp
details), the effect of a low-pass filter is to blur the image.

[14] A high-pass filter eliminates low frequency components of an image while leaving
the high frequency components intact. Because low frequency components correspond to
characteristics of the image that vary slowly (e.g., overall contrast, average intensity), the
effect of a high-pass filter is to accentuate sharp edges and small details of the image.

In the third version (H), only the high frequency information was preserved. The low-pass images of the faces (L) were first obtained[15]. The high-pass images (H) were then obtained by subtracting the low-pass images from the original images. Thus, no overlapping spatial frequency information was present in the low and high frequency versions of the images. The three versions of a face image are displayed in Figure 8.

Experimental design: Five independent variables were manipulated: the *number of training faces* (2, 5, 10, 20, 30, 50, and 100), the *type of learning rule* used to train the memory (Hebbian vs. Widrow-Hoff), the *type of face* presented at test (learned vs. new), the *learning condition* (spectrum range of the face images: F, L, and H) and the *testing condition* (spectrum range of the face images: F, L, and H).

Procedure: The procedure was similar to that used in Simulation 2 with the exception that six autoassociative memories were trained to reconstruct the faces in each training sample set. The first three memories were trained using Hebbian learning with the F, L, and H versions, respectively, of the faces in the sample. The remaining memories were trained using Widrow-Hoff learning with the F, L, and H versions, respectively, of the faces. Testing of each memory was done by presenting each unlearned face, in turn, in each of the three possible versions, as memory keys. For both old and new faces, the quality of the response of the memory was evaluated by computing the squared coefficient of correlation between the test face in the learning spatial frequency condition (e.g., L version if the memory was trained with L version images) and the reconstructed images.

3.4.2. Results and discussion

Figure 9 displays the average r^2 obtained in the nine experimental conditions. The rows of this figure represent the learning conditions: F, L, and H images, respectively. Similarly, the columns represent the testing conditions: F, L, and H respectively. Thus, the panels on the diagonal represent the conditions in which no transfer occurs between learning and testing (F-F, L-L, H-H) and the off-diagonal panels represent the conditions for which a transfer occurred between learning and test (F-L, F-H, L-F, L-H, H-F, and H-L, from left to right and top to bottom). Two main points should be noted from this figure. First, Hebbian learning is affected only by the type of image presented during learning, but not by the type of image presented during test. Moreover, as previously observed, after about 20 faces, the answer of the memory for learned and new faces is identical: It is the average of the

---

[15] Low pass images were obtained by convolving the original images with a 5×5 pixel window with each element set to a value of $\frac{1}{25}$.
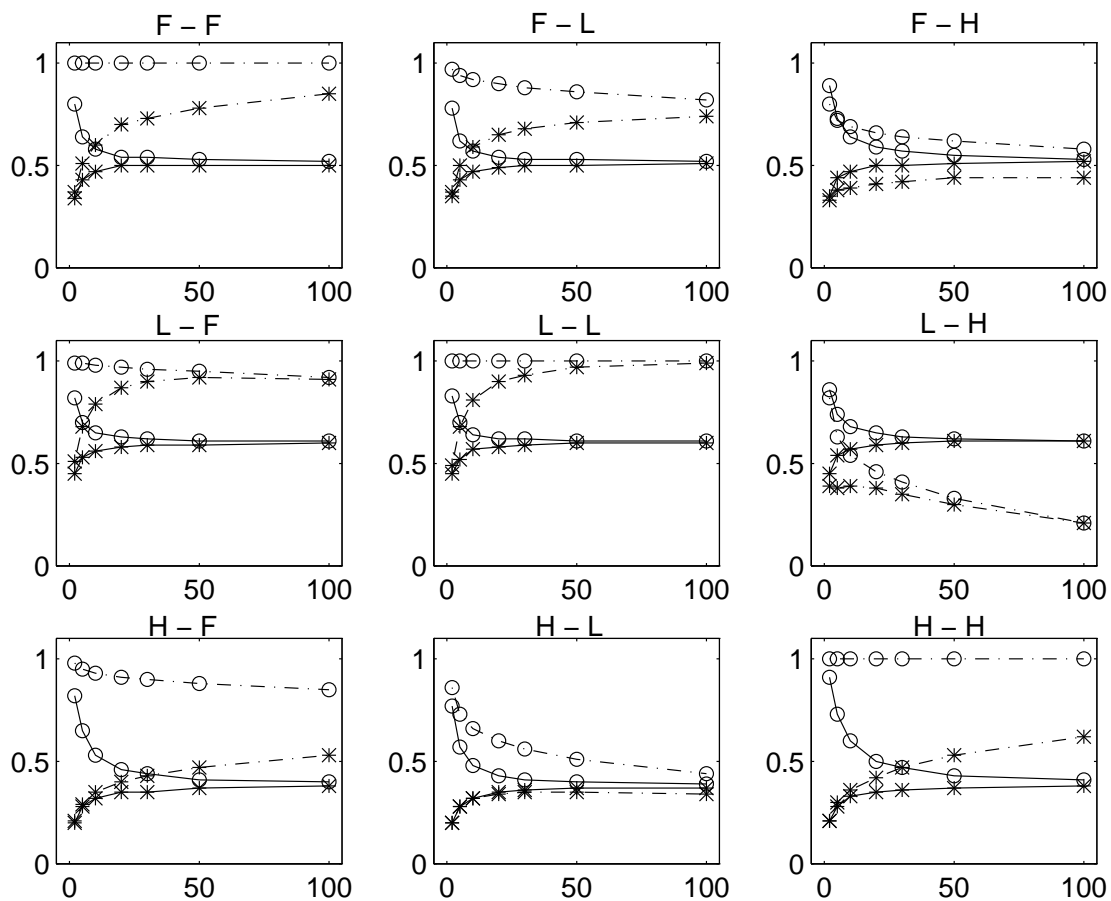
FIGURE 9. Average r^2 represented as a function of the number of faces in the training set, the learning rule used to train the memory, the type of face in the different experimental conditions of Simulation 4. The first letter of each title is the learning condition and the second letter is the testing condition (F = full spectrum, L = low-pass, and F = high-pass). As previously, Hebbian learning is represented by a solid line, Widrow-Hoff learning by a dashed line, old faces by an "o" and new faces by a star.

faces of the training set. As an illustration, the second row of Figure 10 displays the response of the Hebbian trained memory to a key face when the memory learned: 159 full spectrum face images, 159 high-pass face images, and 159 low-pass face images, respectively.

FIGURE 10. Responses of the Hebbian trained memory when full spectrum, high-pass and low-pass images were used as training sets.

Second, Figure 9 shows that Widrow-Hoff learning is affected not only by the type of image presented at learning, but also by the type of image presented during testing. The comparison of the different experimental conditions for Widrow-Hoff learning with the F-F control condition shows that filtering the images affects independently two different aspects of the memory performance: the *quality of reconstruction*, as measured by r^2, as contrasted with the difference in r^2 between learned and new faces, which assesses the *recognition power* of the memory. Both measurements vary as a function of learning and testing conditions, but in different ways. Figure 11 shows the relationship of these two measurements. Looking, first, at the memory performance when no transfer occurred between learning and testing, we see that, although faces are, generally, less well reconstructed in the H-H condition, this condition gives the best discrimination performance. This result confirms the idea that highly detailed information is necessary for discriminating between a large number of faces, and that the recognition superiority of Widrow-Hoff learning over Hebbian learning is due to the processing of high frequency information. Indeed, if a Widrow-Hoff trained memory is used to store a set of low-pass filtered images (L-L condition with a large number of faces), then its recognition performance becomes quite
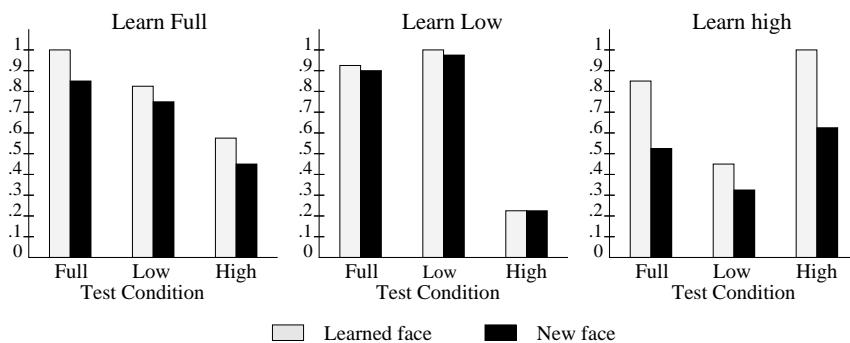
FIGURE 11. Average r^2 for the Widrow-Hoff trained memory in the different learning conditions of Simulation 4 when the training sets were composed of 100 face images. Error bars are not represented on the graphs because of the small size of the standard deviations in all conditions (from 0 to .0088).

close to that of a Hebbian trained memory, although the general quality of reconstruction remains very high.

When some transfer occurs between learning and test, the best reconstruction performance is obtained in the L-F and F-L conditions. In all the other conditions, the reconstruction performance is rather poor. In contrast, the best recognition performance is obtained when high-pass filtered images were presented during learning regardless of the image presented at test. This dissociation between quality of reconstruction and recognition performance indicates that, as was found in previous work (O'Toole et al., 1993; Sergent, 1986a and b), the optimality of a type of information depends on the task performed. Here, it is clear that, whereas low and full frequency enhance the general quality of reconstruction, high frequency information enhances the discrimination power of the model.

Finally, Figures 9 and 11, show that, when a large set of faces is learned, the performance of Widrow-Hoff learning becomes inferior to that of Hebbian learning in some conditions. This is clearly the case in the L-H condition. In this condition, both the quality of reconstruction and the recognition performance of the Widrow-Hoff trained memory become poorer than that of the Hebbian trained memory after more than five faces are learned. This pattern of results is similar to the patterns of Figures 6E and F in Simulation 3 when a large amount of noise was added to the test faces. Because adding random noise to an image is somewhat similar to selectively destroying the high frequency information in the image, this result stresses the importance of high frequency information in the identity coding of learned faces. If this

information is not present during learning, and to a lesser extent during testing, faces loose their identity and no discrimination is possible, even when the quality of reconstruction, as measured by r^2, is high (see condition L-L for example).

To illustrate the performance of the Widrow-Hoff learning algorithm, Figure 12 displays the responses of a Widrow-Hoff trained memory to a learned and a new test face in the different transfer conditions. The first column of this figure shows the learning condition and the second column the testing condition. The last two columns show the response of the memory to the test face when the face was learned by the memory (column 3) or was a new face (column 4). The rows represent the different transfer conditions, from top to bottom: 1) F-L; 2) F-H; 3) L-F; 4) L-H; 5) H-F; 6) H-L. The general quality of the memory responses can be visually estimated by comparing, for each row, the face presented in the first column, and the faces presented in the third (learned) and fourth (new) columns. Similarly, the ability of the memory to discriminate between learned and new faces can be estimated by visually comparing, for each row, the images presented in the third and fourth columns. If these two images are dissimilar, the memory is said to be able to discriminate between learned and new faces. The larger the difference between the two images, the better the discrimination power of the memory.

As expected, when full spectrum images are learned (rows 1 and 2), no visible difference appears between the learned and new reconstructions when a low-pass image is presented at test. In contrast, the learned and new reconstructions are clearly different when a high-pass image is presented at test. This observation is consistent with Figure 9 and 11, showing that, on the average, the values of r^2 are larger for learned faces than for new faces in the F-H condition, but not in the F-L condition. What is more surprising, however, is the visual appearance of the learned faces (column 3) in both testing conditions. Remember that the results summarized in Figure 11 indicated a better quality of reconstruction, as measured by r^2, in the F-L condition than in the F-H condition. Paradoxically, as can be seen in Figure 12, the reconstructed image affords better recognition of the test face for human observers in the F-H condition than in the F-L condition.

The dissociation between the autoassociative model measure (r^2) and human perception raises an important question about the psychological validity of this measure. The adequacy of r^2 as a predictor of human behavior has already been questioned in earlier work (Valentin, Abdi, & O'Toole, 1994; Valentin, Abdi, & O'Toole, in press) which discussed a dissociation between

FIGURE 12. Responses of a Widrow-Hoff trained memory in the six transfer conditions. The first column shows the learning condition, the second column the testing condition, the third column the response of the memory to the face presented in column 2 when the face was learned by the memory, and the fourth column the response of the memory to the face presented in column 2 when the face was not learned by the memory (see text for more details).

categorical and identity information in terms of face eigenvectors and eigen-values. This work showed that, although the reconstruction of a learned face in the subspace determined by the eigenvectors with the larger eigenvalues (i.e., the eigenvectors that capture what is common to many faces) explains most of the variance in the face set (90% on the average), this reconstruction is not identifiable by human observers. In contrast, a reconstruction in the subspace determined by the eigenvectors with small eigenvalues (i.e., the eigenvectors that capture highly detailed, individual specific information), can be easily identified by human observers, even though it explains only a small amount of variance in the face set (10% on the average).

The remaining panels of Figure 12 illustrate that the discrimination performance of the memory is better when high-pass images are learned than when low-pass images are learned. When low-pass images are learned (rows 3 and 4), the learned reconstructed image is quite similar to the new reconstructed image for both testing conditions. When high-pass images are learned (rows 5 and 6), the learned and the new reconstructed images are clearly different in the full-spectrum testing condition (row 5) and slightly different in the low-pass testing condition (row 6).

The superiority of F-H and H-F transfers over other transfer conditions is consistent with previous work (e.g., O'Toole et al., 1993) indicating that high frequency information is essential to code facial identity for both human observers and the computational model.

4. Conclusion

The simulations presented here explored the differential effects of several factors on the performance of a face autoassociative memory trained with either Hebbian or Widrow-Hoff learning. The factors examined included: the task simulated, the number of faces learned by the memory, the presence of random noise, and the spatial filtering imposed upon the face images. The main conclusions that can be drawn from these simulations are the following:

- Different face processing tasks require different levels or types of processing. Simple perceptual tasks, such as categorization by "faceness" or race, which ask only for the processing of low-spatial frequency can be performed by the simpler Hebbian algorithm. More complex perceptual tasks, such as recognition and identification of a large number of faces are based on high spatial frequency information and require the more complex Widrow-Hoff learning.
- The complexity of a face processing task depends on the number of faces to be processed. For example, while Hebbian learning was not

able to discriminate between a large number of faces, this algorithm successfully discriminates between a small number of faces.

- The appropriateness of a learning algorithm also depends on the nature and quality of the stimuli presented. For example, while Widrow-Hoff learning is, in general, more appropriate than Hebbian learning to discriminate between a large number of faces, Hebbian learning becomes more adequate, or even more efficient, when the faces are presented in a very noisy context or when only low spatial frequency information is present.
- The choice of the criterion used to evaluate the performance of a model depends crucially on the task performed. If, for example, the problem is to evaluate the quality of reconstruction of a face image in the least-square sense, or to decide if the model is able to discriminate between old and new faces, then r^2 is a good measure. If on the other hand, the problem is to predict the ability of human subjects to recognize or identify a face, then the use of r^2 becomes problematic. The question of finding a better measurement to predict human subject behavior, however, remains open for future investigation.

From a psychological point of view, our results show that when evaluating the complexity of a face processing task, it is very important to consider, not only the nature of the task, but also the number of stimuli involved, and the perceptual quality of these stimuli. For example, the fact that babies are able to recognize their mother's face among unknown faces (Busnell, Sai, & Mullin, 1989; Pascalis & de Schönen, 1995; Walton, Bower, & Bower, 1992) has often been questioned when considering their limited vision and the high similarity of facial patterns. However, the fact that a very primitive learning algorithm, such as Hebbian learning, is able to discriminate between a small number of faces, even in noisy circumstances, suggests that, in fact, discriminating between a small number of faces is not as difficult a task as one might think. Actually, the complexity of this task is a function of the number of faces to be learned. In particular, the robustness of Hebbian learning, relative to spatial filtering, and its good performance with a small number of faces, make less paradoxical the ability of infants, having limited cognitive and perceptual capabilities, to learn to recognize the faces of their caretakers. This example illustrates the usefulness of a simple computational model in helping to understand the constraints inherent to a given task in a particular context.

References

[1] Abdi, H. (1988). Generalized approaches for connectionist auto-associative memories: Interpretation, implication, and illustration for face processing. In Demongeot, J. (ed.), *Artificial Intelligence and Cognitive Sciences*, pages 151–164. Manchester University Press, Manchester.

[2] Abdi, H. (1994a). *Les réseaux de neurones*. Presses Universitaire de Grenoble, Grenoble.

[3] Abdi, H. (1994b). A neural network primer. *Journal of Biological Systems*, 2:247–281.

[4] Anderson, J., Silverstein, J., Ritz, S., and Jones, R. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 84:413–451.

[5] Bushnell, I., Sai, F., and Mullin, J. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, 7:3–15.

[6] Duda, R. and Hart, P. (1973). *Pattern classification and scenes analysis*. Wiley, New York.

[7] Gluck, M. (1992). Stimulus sampling and distributed representations in adaptive network theories of learning. In Healy, A., Kosslyn, S., and Shiffrin, R. (eds.), *From learning theory to connectionist theory*, pages 160–199. Erlbaum.

[8] Hebb, D. (1949). *The organization of behavior*. Wiley, New York.

[9] Kohonen, T. (1977). *Associative memory: A system theoretic approach*. Springer, Berlin.

[10] McClelland, J. and Rumelhart, D. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press, Cambridge.

[11] O'Toole, A., Abdi, H., Deffenbacher, K., and Valentin, D. (1993). A low dimensional representation of faces in the higher dimensions of the space. *Journal of the Optical Society of America A*, 10:405–411.

[12] O'Toole, A., Millward, R., and Anderson, J. (1988). A physical system approach to recognition memory for spatially transformed faces. *Neural Networks*, 1:179–199.

[13] Pascalis, O. and de Schönen, S. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, 18:79–85.

[14] Sergent, J. (1986a). Methodological constraints on neuropsychological studies of face perception in normals. In Bruyer, R. (ed.), *The neuropsychology of face perception and facial expression*. Lawrence Erlbaum, Hillsdale.

[15] Sergent, J. (1986b). Microgenesis of face perception. In Ellis, H., Jeeves, M., Newcombe, F., and Young, A. (eds), *Aspects of face processing*. Martinus Nijhoff, Dordrecht.

[16] Valentin, D. and Abdi, H. (1996). Can a linear autoassociator recognize faces from new orientations? *Journal of the Optical Society of America A*, 13:717–724.

[17] Valentin, D., Abdi, H., and O'Toole, A. (1994a). Categorization and identification of human face images by neural networks: A review of the linear autoassociative and principal component approaches. *Journal of Biological Systems*, 2:413–429.

[18] Valentin, D., Abdi, H., and O'Toole, A. (in press). Principal component and neural network analyses of face images: Explorations into the nature of information available for classifying faces by sex. In Dowling, C., Roberts, F., and Theuns, P. (eds.), *Progress in mathematical psychology*. Erlbaum, Hillsdale.

[19] Valentin, D., Abdi, H., O'Toole, A., and Cottrell, G. (1994b). Connectionist model of face processing: A survey. *Pattern Recognition*, 27:1209–1230.

[20] Walton, G., Bower, N., and Bower, T. (1992). Recognition of familiar faces by newborns. *Infant Behavior and Development*, 15:265–269.