Dissociated overt and covert recognition as an emergent property of lesioned attractor networks

Martha J. Farah
Randall C. O'Reilly
Shaun P. Vecera

ONR Technical Report

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

January 1992

This research was supported by the Cognitive Science Program, Office of Naval Research, under Grant No. N00014-91-J1546. Reproduction in whole or part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1992 January	Technical

4. TITLE AND SUBTITLE
Dissociated overt and covert recognition as an emergent property of lesioned attractor networks

5. FUNDING NUMBERS
(G)
N0001491J1546

6. AUTHOR(S)
Farah, Martha J.
O'Reilly, Randall C.
Vecera, Shaun P.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Carnegie Mellon University
Department of Psychology
Pittsburgh, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER
CN - ONR - 2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Science Program (1142CS)
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
442K 004

11. SUPPLEMENTARY NOTES

12a DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited.

12b DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

SEE REVERSE SIDE

14. SUBJECT TERMS
face recognition, neural networks, modelling, prosopagnosia, covert recognition, consciousness

15. NUMBER OF PAGES
65

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Abstract

Covert recognition of faces in prosopagnosia, in which patients who cannot consciously or overtly faces nevertheless manifest recognition when tested in certain indirect ways, has been interpreted as the functioning of an intact visual recognition system deprived of access to other brain systems necessary for consciousness. We propose an alternative hypothesis: That the visual recognition system is damaged but not obliterated in these patients, and that it is an intrinsic property of damaged neural networks that they will manifest their residual knowledge in just the kinds of tasks used to measure covert recognition. In support of this, we build a simple recurrent parallel distributed processing model of face recognition and lesion the parts of the model corresponding to visual processing. At levels of damage yielding overt recognition performance comparable to patients described in the literature, the model demonstrates covert recognition in three different tasks: Savings in re-learning correct face-name associations relative to incorrect pairings, semantic priming of occupation decisions on printed names by faces having the same or different occupations, and faster perceptual analysis of previously familiar than unfamiliar faces. Implications for the nature of prosopagnosia, and for other types of dissociations between conscious and unconscious perception, are discussed.

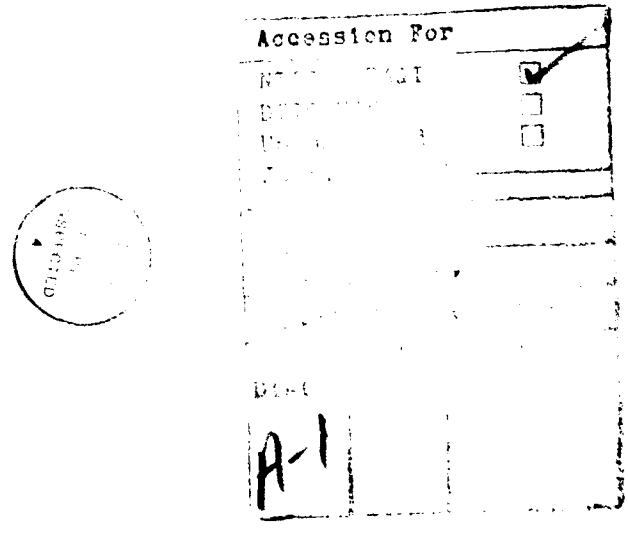
Dissociated overt and covert recognition as an emergent property of lesioned attractor networks

Martha J. Farah
Randall C. O'Reilly
Shaun P. Vecera

Carnegie Mellon University

Address for correspondence:

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213



Abstract

Covert recognition of faces in prosopagnosia, in which patients who cannot consciously or overtly faces nevertheless manifest recognition when tested in certain indirect ways, has been interpreted as the functioning of an intact visual recognition system deprived of access to other brain systems necessary for consciousness. We propose an alternative hypothesis: That the visual recognition system is damaged but not obliterated in these patients, and that it is an intrinsic property of damaged neural networks that they will manifest their residual knowledge in just the kinds of tasks used to measure covert recognition. In support of this, we build a simple recurrent parallel distributed processing model of face recognition and lesion the parts of the model corresponding to visual processing. At levels of damage yielding overt recognition performance comparable to patients described in the literature, the model demonstrates covert recognition in three different tasks: Savings in re-learning correct face-name associations relative to incorrect pairings, semantic priming of occupation decisions on printed names by faces having the same or different occupations, and faster perceptual analysis of previously familiar than unfamiliar faces. Implications for the nature of prosopagnosia, and for other types of dissociations between conscious and unconscious perception, are discussed.

Neuropsychological dissociations between visual perception and awareness of visual perception

In recent years neuropsychology has seen what Weiskrantz (1990) has called an "epidemic" of dissociations involving the loss of conscious awareness in particular perceptual or cognitive domains. Many of these dissociations involve vision. In such cases, patients may deny being able to see or recognize visual stimuli, and indeed perform poorly on certain direct tests of visual perception, but may nevertheless manifest considerable knowledge of the stimulus on certain other, generally indirect, tests of perception. In this article we will focus on prosopagnosia, the impairment of face recognition following brain damage, and the dissociation that has been observed in some cases between the loss of face recognition ability as measured by standard tests of face recognition, as well as patients' own introspections, and the apparent preservation of face recognition when tested by certain indirect tests. Our goal is to elucidate the underlying causes of this dissociation, and its implications for both the nature of prosopagnosia and for the neural correlates of conscious and unconscious perception. Before reviewing the findings to be accounted for in prosopagnosia, we will provide some broader context by briefly reviewing the other syndromes in which visual perception and awareness of visual perception have been dissociated. We will return to these other syndromes, and the possible generalizability of our conclusions regarding prosopagnosia to these other syndromes, in the General Discussion.

The phenomenon of blindsight, in which cortically blind patients who deny having any visual experience can localize and discriminate visual stimuli, was the first neuropsychological dissociation involving conscious awareness to be studied in detail. Although it was initially subject to

much skepticism, two decades of careful research have demonstrated to most people's satisfaction that the dissociation is real, and current efforts center on elucidating the specific neural systems responsible for the nonconscious components of visual perception in blindsight (see Weiskrantz, 1990, for a review).

More recently, similar phenomena have been described in other populations of brain-damaged patients. However, unlike the kinds of relatively low-level visual abilities retained by patients with blindsight, such as discrimination of stimulus location, orientation, or color, which may be mediated by subcortical visual pathways, these dissociations involve very high-level forms of visual perception and recognition. The first of this set of dissociations between high-level perception, on the one hand, and awareness of perception, on the other, was described by Volpe, LeDoux and Gazzaniga (1979) in a study of extinction.

Extinction refers to the impairment in perception of a contralesional stimulus when presented simultaneously with an ipsilesional stimulus. Volpe et al. tested the ability of right parietal-damaged patients to perceive contralesional visual stimuli in two ways. First, the patients were shown a tachistoscopic presentation of a pair of stimuli (line drawings or words), one on each side of fixation, and asked to name what they saw. In this task, the patients manifested visual "extinction" of the left stimulus by the right, which is typical of right parietal-damaged patients: the right stimulus was generally named correctly, but the left stimulus was not, and patients sometimes even denied that the left stimulus had been presented. In contrast, the patients performed well in a second kind of task with the same stimuli. When asked whether the two stimuli presented on a given trial were the same or different, the patients were highly accurate, even though this task requires perception of the left stimulus. Volpe et al. interpreted their findings

as revealing "a breakdown in the flow of information between conscious and non-conscious mental systems."

Another form of visual recognition in the absence of conscious awareness of recognition can be found in certain patients with pure alexia. Pure alexic patients are, by definition, impaired in reading but have roughly normal auditory word comprehension and writing, and their underlying deficit is therefore inferred to be one of visual word recognition. To the extent that they are able to read, they do so by a slow and laborious letter-by-letter strategy, and their reading can therefore be obliterated entirely by presenting words briefly. However, with brief presentations of words, some pure alexic patients are able to derive considerable information from the words, even though they report being unable to recognize the words and even though they cannot name the words (e.g., Shallice & Saffran, 1986; Coslett & Saffran, 1989). For example, with presentations too brief for any explicit reading, these patients are able to discriminate words from orthographically legal nonwords, and to classify words as belonging to a certain category (e.g. animals, foods) at levels far above chance.

In the past few years a fourth type of dissociation between visual recognition and awareness of recognition has been reported, and has already become the most thoroughly studied of the high-level implicit vision syndromes. This is the finding of so-called "covert recognition" of faces by prosopagnosic patients. Prosopagnosia is an impairment of face recognition, which can occur relatively independently of impairments in object recognition, and which is not caused by impairments in lower-level vision, or memory. Prosopagnosic patients are impaired in tests of face recognition such as naming faces or discriminating familiar from unfamiliar faces, and are also impaired in everyday life situations that call for face recognition. Furthermore, by their own introspective reports, prosopagnosics do not feel as though they recognize faces.

Despite the impairments that prosopagnosic patients show on a wide range of tests of face recognition, and despite their own subjective sense of being unable to recognize faces, numerous demonstrations now exist that some prosopagnosic patients do indeed recognize faces at some level. These demonstrations have made use of extremely varied methodologies, so that it is unlikely that any simple methodological artifact underlies the phenomenon. The relevant research includes psychophysiological measures such as skin conductance responses (SCRs) and event-related potentials (ERPs), as well as behavioral measures such as reaction time (RT) and learning trials to criterion.

Evidence for covert recognition of faces in prosopagnosia

In the absence of theories relating psychophysiological indices to mechanistic accounts of cognition or neural information processing, it is difficult to use the psychophysiological findings to constrain a mechanistic model of covert recognition. Therefore, we will focus primarily on the behavioral data implicating covert recognition, and provide just a brief review of some representatative psychophysiological data here.

Psychophysiological evidence. Bauer (1984) presented a prosopagnosic patient with a series of photographs of familiar faces. While viewing each face, the patient heard a list of names read aloud, one of which was the name of the person in the photograph. This test has been called the "Guilty Knowledge Test" because for normal subjects the SCR is greatest to the name belonging to the pictured person, regardless of whether the subject admits to knowing that person. Bauer found that, although the prosopagnosic patient's SCRs to names were not as strongly correlated with the names as a normal subject's would be, they were nevertheless significantly correlated. In contrast, the

patient performed at chance levels when asked to select the correct name for each face.

In a different use of the SCR measure, Tranel and Damasio (1985; 1988) showed that prosopagnosic patients had larger SCRs to familiar faces than to unfamiliar faces, even though their overt ratings of familiarity versus unfamiliarity did not reliably discriminate between the two.

Renault, Signoret, Debruille, Breton & Bolger (1989) recorded ERPs to familiar and unfamiliar faces that had been intermixed in different proportions within different blocks of trials. In general, the P300 component of the ERP is larger to stimuli from a relatively infrequent category. They found that a prosopagnosic patient showed larger P300's to whichever type of face, familiar or unfamiliar, was less frequent in a block of trials, even though the patient was poor at overtly discriminating familiar from unfamiliar faces.

Behavioral evidence. The first evidence of covert recognition in prosopagnosia was gathered by Bruyer, Laterre, Seron, Feyereisen, Strypstein, Pierrard & Rectem (1983) in the context of a paired-associate face-name relearning task, and this task has become the most widely applied measure of covert recognition in prosopagnosia. Bruyer et al.'s patient was asked to learn to associate the facial photographs of famous people with the names of famous people. When the pairing of names and faces was correct, the patient required fewer learning trials than when it was incorrect, suggesting that the patient did possess at least some knowledge of the people's facial appearance. Unfortunately, this demonstration of covert recognition is not as meaningful as it could be, because Bruyer et al.'s subject was not fully prosopagnosic; he could manifest an appreciable degree of overt recognition on conventional tests of face recognition such as forced choice face naming tests.

Recently, several more severe prosopagnosic patients have been tested in the face-name relearning task, and some have shown the same pattern of faster learning of correct than incorrect face-name associations, despite little or no success at the overt recognition of the same faces. For example, de Haan, Young & Newcombe (1987) documented consistently faster learning of face-name and face-occupation pairings in their prosopagnosic subject, even when the stimulus faces were selected from among those that the patient had been unable to identify in a pre-experiment stimulus screening test.

Greve and Bauer (1990) used a different form of learning as evidence of covert recognition in prosopagnosia. They showed a prosopagnosic patient a set of unfamiliar faces, and then showed him the same faces each paired with another face, at which time he was asked the following two questions about each pair: Which of these faces have you seen before? Which of these faces do you like better? Normal subjects tend to prefer stimuli that they have seen previously, whether or not they explicitly remember having seen these stimuli, and this has been attributed to a "perceptual fluency" advantage for previously seen stimuli (Jacoby, 1984). Perceptual fluency refers to the facilitation in processing a stimulus that has already been perceived, which leads to a subjective sense of the stimulus seeming more salient, which may in turn be attributed by the subject to the attractiveness of the stimulus. Although the prosopagnosic patient was unable to discriminate previously seen from novel faces, he did show a normal preference for the previously seen faces.

Evidence of covert recognition has also come from reaction time tasks in which the familiarity or identity of faces are found to influence processing time. In a visual identity match task (see Posner, 1978) with simultaneously presented pairs of faces, de Haan, Young & Newcombe (1987a) found that a prosopagnosic patient was faster at matching pairs of previously familiar faces than unfamiliar faces, as

is true of normal subjects. In contrast, he was unable to name any of the previously familiar faces. De Haan et al. then went on to show another similiarity between the performance of the patient in this task and that of normal subjects. If the task is administered to normal subjects with either the external features (e.g., hair and jaw-line) or the internal features (e.g., eyes, nose and mouth) blocked off, with instructions to match on the visible parts of the face, normal subjects show an effect of familiarity only for the matching of internal features. The same result was obtained with the prosopagnosic patient.

In another RT study, de Haan, Young and Newcombe (1987b; also 1987a) found evidence that photographs of faces could evoke covert semantic knowledge of the depicted person, despite the inability of the prosopagnosic patient to report such information about the person when tested overtly. Their task was to categorize a printed name as belonging to an actor or a politician as quickly as possible. On some trials an irrelevant (i.e., to be ignored) photograph of an actor's or polician's face was simultaneously presented. Normal subjects are slower to categorize the names when the faces come from a different occupation category relative to a no-photograph baseline. Even though their prosopagnosic patient was severely impaired at categorizing the faces overtly as belonging to actors or politicians, he showed the same pattern of interference from different-category faces.

A related finding was reported by Young, Hellawell and de Haan (1988), in a task involving the categorization of names as famous or nonfamous. Both normal subjects and a prosopagnosic patient showed faster RTs to the famous names when the name was preceded by a picture of a semantically related face (e.g., the name "Diana Spencer" preceded by a picture of Prince Charles) than by an unfamiliar or an unrelated face. Furthermore, the same experiment was carried out with printed names as the priming stimulus, so that the size of the priming effect with faces and names could be

compared. The prosopagnosic patient's priming effect from faces was not significantly different from the priming effect from names. However, the patient was able to name only 2 of the 20 face prime stimuli used.

In sum, a wide variety of methods has been used to document covert recognition of faces in prosopagnosia. Although we will argue that not all viable interpretations of this phenomenon have been considered, and we will urge consideration of a new interpretation, it would seem that the correct interpretation is very unlikely to be any kind of methodological artifact. The investigators in this area have been vigorous in attempting to eliminate possible artifacts in each of the experimental paradigms they have used. Furthermore, the sheer diversity of such paradigms makes an artifactual explanation unlikely. Finally, the absence of covert recognition in some cases (e.g., Etcoff, Freeman, & Cave, 1991; Newcombe, Young & de Haan, 1989; Sergent & Villemure, 1990) suggests that it is not a result of the experimental paradigms themselves.

Interpretations of covert recognition in prosopagnosia and their implications

The foregoing results would appear to indicate that, at least in those cases of prosopagnosia who show covert recognition, the underlying impairment is not one of visual recognition per se, but of conscious access to visual recognition. Indeed, all of the interpretations so far offered of covert recognition in prosopagnosia include this assumption.

For example, Tranel and Damasio (1988) say, of their patients' SCRs, that they are "not the result of some primitive form of perceptual process, but rather an index of the rich retro-co-activation produced when representations of stimuli successfully activate previously acquired, non-damaged, and obviously accessible facial records."

Similarly, de Haan et al. (1987a) describe their subject's prosopagnosia as involving a "loss of awareness of the products of the recognition system rather than ... a breakdown in the recognition system per se." In a recent computer simulation of the semantic priming effects described above, this group modelled covert recognition as a partial disconnection separating intact visual recognition units from the rest of the system, again preserving the assumption of intact visual recognition (Burton, Young, Bruce, Johnston and Ellis, in press). Bruyer (1991) offers a similar interpretation, in terms of personal (i.e., conscious agent) and subpersonal (i.e., comprising at least the visual recognition system) levels of description: "the conscious subject does not recognize or identify familiar faces, while her/his 'information processing system' does."

A related form of explanation has been put forth by Bauer (1984), who suggests that there may be two neural systems capable of face recognition, only one of which is associated with conscious awareness. According to Bauer, the ventral cortical visual areas, which are damaged in prosopagnosic patients, are the location of normal conscious face recognition. The dorsal visual areas are hypothesized to be capable of face recognition as well, although they do not mediate conscious recognition but, instead, affective responses to faces. Covert recognition is explained as the isolated functioning of the dorsal face system. This interpretation is similar to the others in that it hypothesizes some form of intact visual recognition. It is distinctive in that the dissociation between recognition and conscious awareness is not a form of disconnection (functional or anatomical) between the visual recognition system and other brain systems that mediate conscious awareness brought about by brain damage, but is the normal state of affairs for the dorsal face recognition system. This explanation is thus analogous to most current interpretations of blindsight, according to which it reflects

the functioning of a different (in that case, subcortical) visual system from that which underlies conscious visual experience.

These interpretations of covert recognition have implications both for the nature of prosopagnosia, and for the neural bases of conscious awareness more generally. With regard to prosopagnosia, current interpretations of covert recognition imply that there are at least two kinds of prosopagnosia, with different underlying causes: one in which visual recognition is intact but unavailable to consciousness (in the case of patients with covert recognition) and one in which visual recognition is impaired (in the case of patients without covert recognition).

At a more general level, these interpretations have implications for the broad issue of the neural bases of consciousness, in that they hypothesize distinct stages of processing, and corresponding distinct neural substrates, for face recognition on the one hand and awareness of face recognition on the other. The assignment of separate brain mechanisms to information processing and awareness of information processing has roots as far back as Descartes' writings on the mind-body problem (with the pineal gland subserving awareness, in that case), and in the context of modern neuroscience has been dubbed "Cartesian materialism" by Dennett and Kinsbourne (in press). Perhaps the most general and lucid expression of this idea, applied to a variety of neuropsychological syndromes including covert recognition by prosopagnosic patients, was put forth by Schacter, McAndrews and Moscovitch (1988). They tentatively proposed that "(a) conscious or explicit experiences of perceiving, knowing and remembering all depend in some way on the functioning of a common mechanism, (b) this mechanism normally accepts input from, and interacts with, a variety of processors or modules that handle specific types of information, and (c) in various cases of neuropsychological

impairment, specific modules are disconnected from the conscious mechanism."

An alternative hypothesis: Residual functioning of an impaired visual recognition system

We will argue that the available evidence on covert face recognition in prosopagnosics is consistent with an impairment in visual recognition per se. This interpretation has implications for our understanding of prosopagnosia, in that it dispenses with the necessity of postulating different forms of prosopagnosia due to different underlying causes. Instead, cases with covert recognition are hypothesized to have more residual functioning of the visual face recognition system than cases without. It also has implications for our understanding of the neural bases of conscious awareness, in that conscious awareness of recognition is not attributed to a distinct neural system from the one subserving recognition per se. Instead, the same neural system subserves both overt and covert recognition.

The primary challenge for such an account is to explain the dissociation between overt and covert recognition, given that these two sets of phenomena are hypothesized to rely on the same neural substrates. We will argue that the difference between them lies in the robustness to brain damage of performance of the two kinds of tasks, in other words, the degree of preserved neural information processing that is required in each case. Specifically, we will argue that lower quality visual information processing is needed to support performance in tests of covert recognition (e.g., to show savings in relearning, and the various RT facilitation and interference effects) relative to the quality of information processing needed to support normal overt recognition performance (e.g., naming a face, sorting faces into those of actors and politicians).

One very general way of stating this hypothesis is to say that the covert tests of recognition are more sensitive to the residual knowledge encoded in a damaged recognition system than are the overt tests. Thus, very impaired performance on overt tests might be associated with only moderately or slightly impaired performance on the covert tests. Stating the hypothesis in this way calls attention to two questions important for evaluating the hypothesis: First, what are the precise levels of patient performance on tests of overt and covert recognition, and are they consistent with the hypothesis of a single damaged system being tapped by tests of differing sensitivity? Normal-size covert recognition effects are unlikely to be due to the functioning of a damaged system (although it would not, strictly speaking, be impossible, if the "ceiling" on covert performance were very low relative to the ceiling on overt performance). Better than chance performance by prosopagnosic patients on overt tests would also be consistent with residual functioning of the visual recognition system (although, by the same token, there is no logical reason why overt performance could not have its "floor" of chance performance above the floor of the covert tests). Second, is there any independent reason to believe that the covert tests would be more sensitive measures of residual recognition ability in a damaged recognition system?

Empirical evidence relevant to testing the alternative hypothesis. In answer to the first question, it is impossible to compare directly the covert recognition performance of prosopagnosic patients and normal subjects on the basis of the evidence currently available, so we cannot know whether their covert recognition is normal, or merely present to some degree. In some cases, data from normal subjects has either not been reported, as in the P300 study, or would be impossible to obtain, as when familiar faces and names are re-taught with either the correct or incorrect

pairings. In other cases, the problem of comparing effect sizes on different absolute measures arises. In both the SCR and RT paradigms, covert recognition is measured by differences between the dependent measures in two conditions (e.g., familiar and unfamiliar faces). Unfortunately, patients' SCRs are invariably weaker than those of normal subjects, and their RTs are longer. It is difficult to know how to assess the relative sizes of differences when the base measures are different. For example, is an effect corresponding to a 200 ms difference between RTs on the order of 2 seconds bigger than, comparable to, or smaller than an effect corresponding to a 100 ms difference between RTs of less than a second? The true scaling of RT in any given task is an empirical issue; using proportions may be a better approximation to the scale than linearity, but one cannot a priori know the true scale (see Snodgrass, Corwin & Feenan, 1990, for a discussion of these issues).

The study that comes closest to allowing a direct comparison of covert recognition in patients and normal subjects is the priming experiment of Young, Hellawell and de Haan (1988). Recall that they found equivalent effects of priming name classification for their prosopagnosic patient with either photographs or names of semantically related people. Of course, this fact alone does not imply that the face-mediated priming was normal, as face-mediated priming in this task might normally be larger than name-mediated priming. To address this problem, Young et al. cite their earlier experiment, reported in the same article, in which normal subjects were also found to show equivalent effects of face-mediated and name-mediated priming. Unfortunately, the earlier experiment differed in several ways from the latter, which could conceivably shift the relative sizes of the face-mediated and name-mediated priming effects: normal subjects in the earlier experiment performed only 30 trials each, whereas the prosopagnosic patient performed 240 trials, items were never repeated in the earlier experiment, whereas they

were in the later one, the type of prime was varied between subjects in the earlier experiment, whereas the patient received both types, different faces and names were used in the two experiments, and the primes were presented for about half as long in the earlier experiment as in the later one. Ideally, to answer the question of whether this prosopagnosic patient shows normal priming from faces, a group of normal control subjects should be run through the same experiment as the patient.

Turning now to the question of whether the prosopagnosic patients who show covert recognition also show some degree of overt recognition, consistent with a damaged but not obliterated visual recognition system, the evidence is similarly difficult to evaluate. For example, some patients' chance performance on overt tasks is consistent with the use of extreme response biases, which would mask any degree of remaining sensitivity. Among the three prosopagnosic patients studied by Tranel and Damasio (1988), two rated almost all faces as "unfamiliar," and the one who used a larger portion of the rating scale narrowly missed the .05 significance level in discriminating familiar from unfamiliar faces.

Statistical naivte concerning the concept of chance performance has also led to confusion. In some cases, the term "chance performance" has been used synonymously with poor performance. For example, de Haan et al. (1987b) present the results of an overt actor/politician face judgement task with their patient, and describe the score of 30/48 in a two alternative forced choice task as being at chance. In fact, there is only a .06 probability of achieving such a high score by guessing alone. In other cases, performance is truely not statistically different from chance (e.g. in Young & de Haan, 1988, 12/30 in a three alternative forced choice familiarity task) but the small number of trials makes this a relatively weak test for purposes of obtaining confidence in the null hypothesis.

In addition, the ability of this patient and others to occasionally identify a face by name, a task whose "chance level" is difficult to estimate but is certainly close to 0% correct, also indicates that visual recognition has not been entirely obliterated. For example, this same patient was able to identify 2 out of 20 of the faces used in the semantic priming study of Young et al. (1988).

One way in which investigators have attempted to control overt recognition performance and measure covert recognition in the absence of overt recognition is by testing patients only on faces that were not successfully identified in a screening test. For example, de Haan et al. (1987a) used only the faces that their prosopanosic patient had failed to recognize in their face-name relearning task. However, this presupposes both that there is little or no measurement error in the overt task, and that overt identification is as sensitive a test of recognition as savings in relearning. That these assumptions are problematic was demonstrated by Wallace and Farah (submitted), who followed the same screening procedure of eliminating successfully identified faces with normal subjects on faces that had been learned six months prior to the experiment, and nevertheless found savings in relearning the original face-name associations, relative to new pairings.

Computational rationale for the alternative hypothesis. The empirical data reviewed so far fail to distinguish between the original hypothesis of intact face recognition deprived on access to consciousness, and the alternative hypothesis that face recognition is impaired and that covert tasks are more sensitive than overt tasks to detecting residual functioning. Our reason for prefering the alternative hypothesis is based on a consideration of the relative computational demands of the overt and covert tests. In order to explain how these differ, we will first provide a very brief overview of computation in recurrent neural

networks. More extensive background can be found in Rumelhart and McClelland's (1986) book on parallel distributed processing models of cognition.

In parallel distributed processing models, representations consist of a pattern of activation over a set of highly interconnected neuron-like units. The extent to which the activation of one unit causes an increase or decrease in the activation of a neighboring unit depends on the "weight" of the connection between them; positive weights cause units to excite each other and negative weights cause units to inhibit each other. For the network to learn that a certain face representation goes with a certain name representation, the weights among units in the network are adjusted so that presentation of either the face pattern in the face units or the name pattern in the name units causes the corresponding other pattern to become activated. Upon presentation of the input pattern to the input units, all of the units connected with those input units will begin to change their activation in accordance with the activation value of the units to which they are connected and the weights on the connections. These units might in turn connect to others, and influence their activation levels in the same way. In recurrent, or attractor, networks, the units downstream stream will also begin to influence the activation levels of the earlier units. Eventually, these shifting activation levels across the units of the network settle into a stable pattern, or attractor state. The attractor state into which a network settles is determined jointly by the input pattern (stimulus) and the weights of the network (stored knowledge).

Accordingly, much of the behavior of the network depends on the pattern of weights. For example, the weights determine not only which pattern becomes activated in association to an input pattern, they also determine how quickly this pattern becomes stable and how quickly a given unit or set of units reaches some pre-determined threshold of activation. Not

surprisingly, the current pattern of weights will also determine how many training cycles are needed to teach the network a new association. In ways that will be elaborated shortly, these aspects of network behavior seem closely related to the behavioral measures of covert recognition reviewed earlier: speed of perception (corresponding to settling time), speed of classifying actors and politicians (corresponding to how quickly actor or politician representations reach threshold), and, of course, paired associate learning (a direct correspondence).

When a network is damaged by eliminating units, it will be less effective at associating the patterns that it knew previously. This can be understood in terms of the idea that knowledge is stored in the weights by viewing unit damage as the permanent zeroing of all weights going into and out of the eliminated units. As more units are eliminated, the ability of the network to correctly associate previously known patterns will steadily decline until it reaches chance levels.

The impetus for our project comes from the following key idea: The set of the weights in a network that cannot correctly associate patterns because it has never been trained (or has been trained on a different set of patterns) is different in an important way from the set of weights in a network that cannot correctly associate patterns because it has been trained on those patterns and then damaged. The first set of weights is random with respect to the associations in question, whereas the second is a subset of the necessary weights. Even if it is an inadequate subset for performing the association, it is not random; it has, "embedded" in it, some degree of knowledge of the associations. Furthermore, consideration of the kinds of tests used to measure covert recognition suggest that the covert measures might be sensitive to this embedded knowledge. The most obvious example is that a damaged network would be expected to re-learn associations that it

originally knew faster than novel associations because of the nonrandom starting weights. Less obvious, but nevertheless plausible for reasons to be elaborated later, the network might settle faster when given previously learned inputs than novel inputs, even though the pattern into which it settles is not correct, because the residual weights come from a set designed to create a stable pattern from that input. Finally, to the extent that the weights continue to activate partial and subthreshold patterns over the nondamaged units in association with the input, then these resultant patterns could prime (i.e. contribute activation towards) the activation of patterns by intact routes. These mechanisms will be discussed in greater detail in the context of the individual simulations. For present purposes, the general implication of these ideas is that as a neural network is increasingly damaged, there might be a window of damage in which overt associations between patterns (e.g., faces and names) would be extremely poor while the kinds of performance measures tapped by the covert tasks might remain at high levels. Note that if this is true, it does more than just undermine the prevailing hypothesis of intact face recognition systems in those prosopagnosic patients who manifest covert recognition. It offers a specific, mechanistic hypothesis explaining the overt/covert dissociations in terms of general principles of computation in attractor networks.

In order to test this hypothesis, we developed a very simple model of face recognition, and explored the effects of damage to visual input units on network performance of three different types of tasks, corresponding to the savings in relearning paradigm, the physical matching paradigm, and the priming paradigm. Before presenting the model and simulations themselves, we will explain the concepts of activation space and weight space, which are helpful for understanding the behavior of the model.

Spatial analogies for understanding the behavior of attractor networks

Spatial analogies are useful for visualizing certain aspects of network dynamics, including the way in which the network's patterns of activation change under the influence of an input, and the way in which the ensemble of weights changes during learning. These analogies will also be useful in understanding the behavior of the present network under damage.

The activation state of the network at any point in time can be represented as a point in a high-dimensional space called activation space. The dimensions of this space represent the level of activation of each unit in the network, assuming a fixed set of weights. In addition to the dimensions representing the activation levels of the units, there is one additional dimension, representing the overall fit between the current activation pattern and the weights.

When units that are both active have a large positive weight between them, so that they reinforce each other's activation, this is an example of a good fit. If one unit is activated and another is not, and the weight connecting them is positive, or if both units are active and their is a negative (i.e., inhibitory) weight between them, the fit would be poor. This measure of fit is called "energy," with low energy representing a better fit. The energy value associated with each pattern of activation defines a surface in activation space.

When an input pattern is presented to the network, the corresponding initial position in activation space is defined by the activation levels on the input units, along with resting level values for the dimensions representing the other units in the network. The weights in the rest of the network will not fit well with uniform resting level activation values over their portion of the network (assuming they have been trained to associate a pattern with the input). Thus, the initial point in activation space will be

in a region of high energy. As activation propagates through the network, the pattern of activation changes and the point representing this pattern moves along the energy surface in activation space. The movement will be generally downwards, as the network lowers its energy, much as a ball rolls down a hill to lower its potential energy. To see why this would happen in terms of network dynamics, rather than by analogy with rolling balls, consider the examples given earlier of high and low energy activation states. For example, active units connected by negative weights (a poor fit, high energy pattern) will tend to change their activations until one is active and the other not (a good fit, low energy pattern).

The energy minima towards which the network tends are the "attractors" mentioned earlier in this article. Attractors are useful in network computation not only for associating patterns and completing partial patterns, but also for their ability to "clean up" a noisy input, by transforming a pattern similar to a known pattern into that known pattern (i.e., a pattern just uphill from an attractor will roll down into the attractor).

How quickly the network settles when presented with an input pattern depends upon how quickly it can traverse the distance between its starting point in activation space and the attractor into which it "rolls." This in turn depends on the shape of the energy "landscape" because the network's activation pattern will travel more directly (and therefore quickly) down a steep smooth incline than along more bumpy, winding terrain. The shape of the energy landscape is determined by the network's weights. In an untrained network, the landscape will be generally flat with random dips. When the network has learned a certain association, its weights will create an energy landscape in activation space in which the point corresponding to the input pattern and the attractor point corresponding to the complete associated pattern are connected by a smoothly and steeply sloping path that causes the one state to "roll" directly

down into the other. Because some patterns will have the same value on some dimensions (i.e., they will have units activated in common) the network will need barriers to prevent confusion among trajectories for different patterns. Paths bounded by these barriers can be thought of as ravines.

The weights that underlie the attractor structure of activation space can themselves be used to define a space, and this space is useful for visualizing the process of learning. In weight space, each of the weights in a network corresponds to one dimension of a space, so that we can represent the sum total of the network's knowledge as a point in this high dimensional space. If one additional dimension is now added to the space, representing the performance of the network at associating names and faces, then there will be a surface defined by each combination of weights and their associated performance. The energy of the point in activation space to which the network settles with a given set of weights is a measure of performance, with low energy (that is, good fit between the weights and the resultant activation pattern) being better performance. If, when we present the input, we also fix the activation values for the units for the associated pattern ("clamping"), then the desired weights will be those that minimize the network energy associated with this pattern. Learning consists of moving along this energy surface in weight space, changing weight values, until a sufficiently low point has been reached.

The model

The present model is intended to illustrate some very general, qualitative aspects of the behavior of damaged attractor networks in the kinds of tasks used with prosopagnosic patients. It is accordingly very simple. Figure 1 shows the architecture of the model. There are five pools of units. The face input units subserve the initial visual representation of faces, the semantics units subserve

representation of the semantic knowledge of people that can be evoked by either the person's face or name, and the name units subserve the representation of names. In a model of this kind, hidden units are helpful to learn the associations among patterns of activity in each of these three layers. These are located between the face and semantic units, (called the face hidden units) and between the name and the semantic units (the name hidden units). Thus, there are two pools of units that comprise the visual face recognition system in our model, in that they represent visual information about faces: the face input units and the face hidden units.

The connectivity among the different pools of units was based on the assumption that in order to name a face, or to visualize a named person, one must access semantic knowledge of that person (Young, Hay & Ellis, 1985). Thus, face and name units are not directly connected, but send activation to one another through hidden and semantic units. The arrows in Figure 1 show the bidirectional connectivity between layers and the within-layer connectivity. Further, each unit had a bias weight which learned the average activation level of that unit (a technique for improving the ability of the network to learn, see Rumelhart, Hinton & Williams, 1986).

Units in this model have a threshold of zero. Thus, when the activation value of a unit is positive, it will activate those units to which it is connected by positive weights and inhibit those units to which it is connected by negative weights, and when its activation value is negative it will have the opposite effects.

Faces and names are represented by random patterns of 5 active units out of the total of 16 in each pool. Semantic knowledge is represented by 6 active units out of the total of 18 in the semantic pool. The model makes no committment to any particular form of representation, beyond supposing that the representations are distributed -- that is, each face, semantic representation or name is represented by multiple

units and each unit represents multiple faces, semantic representations or names. The information encoded by a given unit will be some "microfeature" (Hinton, McClelland & Rumelhart, 1986) that may or may not correspond to an easily labelled feature (such as eye color in the case of faces). The only units for which we have assigned an interpretation are the "occupation units" within the semantic pool. One of them represents the semantic microfeature "actor" and the other represents the semantic microfeature "politician."

We created 40 distinct individuals, each consisting of a random name, face and semantic pattern (over the 16 unlabelled semantics units). Ten individuals were actors (i.e., their semantic pattern had the actor unit active in addition to the other 5 active semantics units), ten were politicians, and the remaining 20 were not assigned either of these two occupations. These 20 individuals were not tested in the simulations to be reported, but were included in training to simulate the fact that subjects know many more people than are ever tested in a given experiment. Of the 10 actors and 10 politicians, five of each were not used in training, so that we could compare the effects of familarity on network performance in Simulation 2, resulting in a training set of 30 patterns.

The network was trained to be able to associate an individual's face, semantics, and name whenever one of these was presented, using the Contrastive Hebbian Learning (CHL) algorithm (Movellan, 1990). CHL is a variation of a mean field approximation of a Boltzmann Machine (Hinton & Sejnowski, 1986; Hopfield, 1984). For each training epoch we presented one of the three representations for each individual (face, semantics, name) and trained the network to reproduce the other two. The learning rate was .01. The network was trained for 320 epochs on the complete set of 30 individuals, and for an additional 5 epochs on the set of the 10 individuals to be later tested to insure 100% accuracy for these individuals in the undamaged network.

In accordance with the CHL algorithm we used the Interactive Activation and Competition (McClelland & Rumelhart, 1989) activation function, with a step size of .01, maximum of 1, minimum of -1, rest of 0, and decay of .2.

Simulation 1
Savings in relearning face-name associations

The primary goal of this simulation was to examine the effects of different degrees of damage to the visual units (face input and face hidden units) on both overt identification of face patterns and on the difference in number of cycles needed to re-learn previously known name-face associations, relative to the number needed to learn to associate the same names and faces paired differently. Hinton and Sejnowski (1986) demonstrated savings in relearning after a variety of types of damage to a recurrent network, including unit ablation. If there is some degree of damage to the face units that can result in poor overt performance while preserving significant savings in relearning, then the savings in relearning observed in prosopagnosic patients need not imply that visual recognition per se has been spared.

Methods

The network was lesioned in two different ways: by eliminating randomly chosen units from the face input pool and from the face hidden unit pool. Seven different levels of damage were used, corresponding to removal of 2, 4, 8, 10, 12, and 14 units from the pools of 16 units, corresponding to 12.5%, 25%, 37.5%, 50%, 62.5%, 75%, and 87.5% damage.

The basic measure of overt recognition, used for comparison with covert performance in all of the simulations to be reported, was the percentage of correct name identifications of faces in a 10-alternative forced choice

among the 10 test patterns. Thus, a face was considered correctly identified if the resultant name pattern matched the correct name pattern more closely than any of the other 9 test patterns. Degree of match was quantified by the number of units having the same sign (positive or negative). This is a more lenient method of scoring overt recognition than requiring a perfect match, or even a match to within one bit.

In the first simulation, the names and faces for the ten familiar actors and politicians were paired correctly. In the second simulation, they were paired incorrectly, although never across occupation categories, because this would confound the correct-incorrect distinction with the compatability of the occupation unit pattern. In order to expedite learning, each network was required to learn only 5 name-face pairs at a time. These were presented to the network after damage for retraining in separate simulations. In order to simulate the training procedure used with patients, in which they are asked to name the face on each trial rather than select from a multiple choice set of names, we used the pattern that resulted in the name units of the network following presentation of the face as the simulation's response. This was scored as correct if it matched the target pattern to within 2 units.

In order to measure savings in relearning for correctly paired names and faces, the damaged network was retrained for 10 epochs and its performance on overt identification was assessed. This procedure was repeated 10 times with different sets of random lesions, in order to assess the reliability of the results.

Results and Discussion

Table 1 and Figure 2 show the overt identification performance of the network in the 10 alternative forced choice task after different amounts of damage to the two pools of visual units. By 50% damage to either pool of

units, the network is correct for only about 1 in 4 faces. With higher levels of damage performance drops further. At 62.% and 75% damage to face input units, only about 1 in 6 faces are corrected identified. At these same levels of damage to face hidden units, performance is not significantly different from 1 in 10, or chance performance.

Despite the network's poor performance in the overt tasks under damage, it manifests covert knowledge of the faces by relearning correct name-face pairings more quickly than incorrect ones. Table 2 shows the average percent correct naming, to within a 2 unit matching criterion of the correct name, for each degree of damage to the face input and hidden units after 0 and 10 epochs of learning for correctly and incorrectly paired faces and names. Figure 3 shows the learning curves for the network after 50, 62.5 and 75% damage to the face input and face hidden units for the same pairings. Although not all levels of damage lead to equivalent performance for correct and incorrect pairings at the outset of training, the learning curve is steeper, that is learning is faster, for the correct pairings in all cases. Furthermore, this is true even with 62.5 and 75% damage to face input units, and with 50 and 75% damage to face hidden units, for which the pre-training performance of the damaged network is comparable for correct and incorrect pairings.

The phenomenon of savings in relearning correct face-name pairings in the damaged network can best be understood in terms of the way in which weight space is altered by damage. The explanation has two parts: First, we will explain that the energy associated with a particular point in weight space does not change drastically as a result of damage, and accordingly, activation patterns that were attractors before damage to the weights remain relatively low energy states (i.e., are at least close to being attractors) after damage. Second, we will explain that these relatively small changes can nevertheless result in poor performance because a particular input pattern may no longer be able to

roll into the correct attractor, even though that attractor may have been preserved. As a result of these two properties of damaged attractor networks, only a small amount of weight change (re-learning) will typically be needed to restore the performance of the network on previously learned associations.

To begin with the first part of the explanation, we will explain why the removal of units preserves the overall topography of the energy landscape of weight space, and thus the locations of attractors in activation space. When units are removed from the network, all of the weights going into and out of these units are also eliminated. This reduces the dimensionality of the weight space, creating a "projection" of the higher dimensional space onto the resulting lower dimensional space. When this happens, the values of the remaining weights may not be optimized for correct associations on their own. Therefore, the energy surface of the weight space may no longer have minima in exactly the correct locations. However, the change in shape is generally not drastic; points that were low in energy before the projection stay relatively low afterwards. A brief formal explanation of why this is so follows.

The energy of a point in the weight space is defined by the mean field algorithm to be:

$$-\Sigma_i \Sigma_j a_i w_{ij} a_j + \Sigma_i f_{stress}(a_i) \qquad (1)$$

where a_i and a_j are the activations of the units connected by weight w_{ij}, and f_{stress} is a monotonic function of the unit activation that penalizes large activations. The change in energy that results from the elimination of a single unit, a_i, is therefore a linear function of two components: the weights w_{ij} to the units to which it was connected, and the activation value of the unit. Assuming that the network settles into the same activation state for its remaining

units after damage as before, the energy of that state will have changed in direct proportion to the amount of damage. So, for example, with 75% damage the energy of the corresponding point in weight space would differ from the pre-damage point on average by 75%, as opposed to more drastic changes by orders of magnitude.

This might lead one to expect the network's performance to be highly robust in the face of rather large lesions. After all, if the attractors in the activation space associated with the remaining subset of weights have not been greatly shifted, then input patterns should still be able to roll along its old trajectory into approximately correct final states. However, this is not the case because of the second of the two properties mentioned earlier. There may be small bumps introduced into that trajectory that have the potential to deflect the network onto a different trajectory at junctures en route. (These junctures represent the crossings of paths in some, but not all, dimensions.) To the extent that input patterns are similar, that is share active units in common, there will be many such junctures as similar starting points in activation space must be channeled into different final points in activation space. Because any "wrong turn" will result in an erroneous final state, performance will be powerfully affected by these small perturbations in the energy surface.

We are now in a position to explain the phenomenon of savings in relearning. Although a small change in the energy surface can cause drastic changes in the final activation state by leading the network into a different trajectory, the amount of learning (i.e., weight change) that is required to restore the network to good performance is generally small, because it is only necessary to eliminate the critical small bumps in activation space. As already shown in the first part of the explanation, the large scale structure of the activation space will have been preserved and need not be relearned.

Simulation 2
Speed of visual perception

The goal of this simulation was to examine the effect of different degrees of damage to visual units on the speed of visual analysis of face patterns, and specifically whether speed of analysis will depend on face familiarity at levels of damage where faces are not reliably identified. This question is of interest primarily because of de Haan et al.'s (1987) demonstration that their prosopagnosic subject could perform physical same/different matching on faces more quickly when the faces were previously known to him. Presumably, the effect of familiarity on speed in this paradigm is not dependent upon same/different matching per se, but reflects a difference in the speed of deriving a visual representation that can be used to compare the appearance of the two faces. Therefore, we have not tried to implement a same/different matching paradigm here. The relevant issue is whether visual analysis of a face pattern proceeds more quickly when the face is familiar than when it is unfamiliar.

In the present model, the speed of visual perception is most directly measured by the number of cycles needed for the visual units of the network to settle into a stable pattern after presentation of a face pattern. Note that we need assume only a monotonic relation between model settling time and human RT in order to interpret the results of the present simulation.

Methods

The model was lesioned as in the previous simulation. The face portion of the 10 actor and 10 politician patterns were then presented to the network. As explained earlier, in

the description of the model, the network had been trained on half of these patterns, divided equally into 5 actors and 5 politicians. The number of cycles needed for the visual units (input and hidden) of the network to settle was recorded for each face pattern. The visual units were considered to have settled when the average change in activation of the units in a cycle was less than .001. The face input unit activations were allowed to settle by presenting the input pattern as a component of the net input to each unit, instead of simply clamping the activations (i.e., "soft" clamping). As for the previous simulation, 10 replications were performed with different random patterns of damage.

Results and Discussion

The settling times for familiar and unfamiliar face patterns are shown in Table 3 and presented graphically in Figure 4. At levels of damage causing poor or chance overt performance (see Table 1), the settling time for familiar face patterns is nevertheless faster than for unfamiliar patterns. This pattern is maintained throughout all degrees of damage to the face hidden units, and is present with as much as 50% damage to the face input units.

Why should the familiarity of the pattern affect how quickly it settles? In an intact network, a familiar input pattern will roll into an attractor representing the correct pattern of activation to which it should be associated, because the energy landscape has been tailored for this purpose. Given that much of the activation space has been shaped by learning, the trajectory of the network when presented with the input portion of an unfamiliar pattern will also tend towards attractors for the familiar patterns. However, because unfamiliar input patterns begin their trajectory in a region of activation space that has not been specifically shaped for this purpose, their trajectories into the attractor state will typically be less direct and more

circuitous. For this reason, familiar input patterns will settle faster than unfamiliar input patterns in an intact network.

When the network is lesioned, the loss of units reduces the dimensionality of the space, and the loss of weights distorts the shape of the new, lower-dimensional energy landscape. As explained previously, the large-scale topography is preserved, maintaining the settling time advantage for patterns that begin on the glide slope of attractors. In contrast, there is no reason to expect unfamiliar input patterns to find themselves any nearer to trained attractor slopes, on average, after damage than before. However, because of the susceptibility of the activation state trajectory to mischanneling by small bumps in the energy surface at potential junctures, also explained previously, overt network performance will suffer. Thus, the settling time advantage for familiar patterns is preserved in conjunction with poor overt performance.

The faster settling of familiar patterns is also relevant to Greve and Bauer's (1990) finding of greater perceptual fluency for faces seen previously by their prosopagnosic subject, as measured by attractiveness ratings. With exposure to new patterns, the damaged network will alter its weights to begin to form attractors for those patterns, although it will arrive at the best set of weights more slowly than a network that has a larger number of weights (cf. the slower learning of the increasingly damaged networks in Simulation 1). This leads to faster settling times for faces seen post-damage than for completely novel faces, even before the network has learned to accurately recognize the patterns.

Indeed, examination of the settling times for the novel patterns of Simulation 1 (i.e. the novel combinations of faces and names) shows that, at early stages of learning associated with chance overt performance in the damaged network, settling time is nevertheless reduced relative to no

learning. As shown in Table 4, at all levels of hidden unit damage and at 4 out of 7 levels of input unit damage, settling time is faster after just 5 epochs of training than before.

Simulation 3

Semantic priming of occupation decisions

The goal of this simulation was to examine the effects of different amounts of damage to the visual units on facilitation and interference caused by a face prime when judging the occupation of a named person. As a related measure of overt performance, the network was presented with the face input patterns alone to classify according to occupation.

Methods

The model was lesioned as in the previous simulations. The name portions of the 5 familiar actor and 5 familiar politician patterns were presented to the network, paired with face patterns from the same group of individuals. Each of the 10 names was presented in three conditions: alone, paired with the nonidentical same-occupation faces and paired with the different-occupation faces. The number of cycles needed for one of the occupation units, actor and politician, to attain a positive activation value was recorded. (The bias weights, learned during training, were largely inhibitory, leaving the units in a negative state in the absence of input activation.) As usual, 10 replications of the simulation with different random lesions in each of the two pools of units were carried out.

The overt ability of the network to derive occupation information from the face patterns was measured by recording which occupation unit reached threshold (i.e., became positive) after presentation of the face. For trials on which neither unit reached threshold, the network was assumed

to guess with probability .5 of being correct. The rationale for scoring performance in this way, rather than taking the larger activation of the two regardless of whether either are positively activated, is that units, like neurons, have a categorical quality to their state. In the present model, there is a categorical difference between the way in which positive and negative valued activation in a unit affects the other units to which it is connected. For example, a negatively activated unit will inhibit units to which it is connected by positive weights, but will excite them when its activation goes positive. Note that the method of scoring overt categorization was lenient in that we only require the sign of the activation to be correct.

Results and Discussion

The performance of the network on the overt occupation decision for faces is shown in Table 5. With lesions to hidden units or input units of 50% and 62.5%, the network's performance falls in the range of 59-65% correct. This is comparable to the performance of the prosopagnosic patient reported de Haan et al. (1987b), who obtained 62.5% correct on the same task.

Table 6 and Figure 5 show the number of cycles required for the correct occupation unit to become positive after presentation of a name, without an accompanying face, and with faces from the same or different occupation category. Fewer cycles are required for the occupation units to attain positive values when the face and name are from the same occupation category than when they come from different occupation categories. The effect of the face is evident at all but the most extreme levels of damage. In particular, it is evident at the levels of damage to input and hidden units whose corresponding overt performance was discussed above. The data from the no-face condition suggest that, as in de

Haan et al.'s (1987) study, the effect is primarily one of interference rather than facilitation.

The mechanism by which faces affect performance in the present model is as follows: To the extent that the presentation of a face pattern causes any activation to propagate into the rest of the network, this activation will influence the activation of the occupation units, even if it is not in itself sufficient to bring them all the way to threshold. At first glance this would seem to predict both facilitation and intereference. Facilitation would arise because the face would contribute activation towards its occupation unit, and if the name has the same occupation, less additional activation from the name pattern would be needed for that occupation unit to attain a positive value. Interference would be predicted because the negative correlation between the two occupation units' activations, over the set of known patterns, would have resulted in an inhibitory connection between them having been learned by the network, and so that the activation of either occupation unit would tend to inhibit the activation of the other. In effect, the network learns which subpatterns are consistent and inconsistent with which others, and inconsistent subpatterns (e.g., the single unit actor or the single unit politician) will tend to inhibit each other. The lack of an observed facilitation effect is attributable to mutual inhibition of the patterns for different individuals in the same occupation category counteracting the facilitation mechanism just described. That is, some of the units activated by the name, which would normally contribute activation to the occupation unit, are themselves being inhibited by the influence of the face pattern.

A similar account has already been presented by Burton, et al. (in press) to explain semantic priming by faces in prosopagnosia. They implemented an interactive activation model with local representations, and simulated the effect of faces in the occupation decision task for a prosopagnosic

subject by attenuating the connections between their equivalents of face units and semantics. However, because their simulation is local and hand-wired, it does not develop mutual inhibitory relations among subpatterns as in the present model. As a consequence, it shows as much facilitation as interference in this task.

The same mechanism proposed here will, in principle, explain Young et al.'s (1988) finding of semantic priming of names by associated faces in a name familiarity task. Indeed, Burton et al. were also able to simulate the two kinds of tasks in the same way.

Simulation 4
Covert recognition of overtly unrecognized faces

In this final section, we demonstrate that the preserved covert recognition ability in the damaged network is not the result of the network's preserved overt recognition ability for a subset of the familiar patterns. The demonstration takes the form of an existance proof. For each of the three tasks simulated, we tested the covert recognition performance of the network just on the subset of faces that it failed to recognize in the overt recognition tests.

Methods

A randomly selected 50% of the face hidden units were damaged and the resulting network was tested on the overt 10-alternative forced choice recognition test. The two faces out of 10 that were correcty identified were eliminated from the set of test faces. For the semantic priming experiment, only the 5 faces that were not correctly categorized as actors or politicians were retained in the test set. The damaged network was then tested for covert recognition in the three previously described tasks.

Results and Discussion

The network relearned the correct associations among the eight faces and names faster than the incorrect: After damage and prior to learning, it obtained a score of 0% correct for both the correct and incorrect name-face pairs. After 10 epochs of learning, more learning had taken place for the correct pairs: the network obtained a score of 50% correct for the correct pairs and 0% correct for the incorrect pairs.

As before, presentation of a face from the wrong occupation category delayed the relevant occupation unit from reaching threshold when a name was presented. The mean number of cycles to reach threshold was 70.0 when no face was presented, 33.6 when a face from the same category was presented, and 94.9 when a face from the other category was presented.

Settling time in the face units was faster for the 8 previously learned faces than for the 10 novel faces, on average 200.8 and 232.2 cycles, respectively.

In sum, the covert recognition abilities displayed by damaged attractor networks does not depend upon the presence in the test set of any overtly identified face patterns.

General Discussion

We have shown that some very general properties of attractor networks lead to preserved performance, after network damage, for the types of tasks used to measure covert recognition in prosopagnosia. Specifically, we have simulated in varying degrees of detail three types of behavioral task used to document covert recognition. At levels of damage associated with low overt identification and categorization performance of face patterns, the network continues to manifest knowledge of the faces when tested by the covert tasks. Of additional interest is the fact that visual portions of the network were damaged in these

simulations, demonstrating that one need not conclude that visual recognition is intact in cases of prosopagnosia with covert recognition. In the remainder of this article, we will discuss the implications of these results for our understanding of covert face recognition, other covert visual abilities, prosopagnosia, and consciousness.

Covert face recognition. Previous attempts to explain covert recognition of faces in prosopagnosia have assumed that covert and overt recognition are dependent on at least partly distinct components of the cognitive architecture, somehow disconnected by brain damage, and that the visual recognition component is intact. In contrast, we have shown that the same system may subserve both overt and covert visual recognition, and that damage to this system may spare covert recognition relative to overt recognition.

Of course, the results of our simulations do not prove that our account is correct, merely that it is possible. Nevertheless, we find it plausible for three reasons: First, it follows from a set of independently motivated computational principles. These include the utility of attractor states in network computation, and the tendency of activation space to preserve its large-scale attractor structure under damage while acquiring changes in small-scale topography that impair the network's overt performance, as well as the concept of a threshold for activation flow between units. Second, it is consistent with the available data on overt and covert performance in prosopagnosic patients, specifically the occasional success in overt tasks by these patients. Third, it is a parsimonious account. It is not necessary to invoke separate brain centers for recognition and overt awareness of recognition and only one face recognition system is hypothesized (cf. Bauer, 1984). Furthermore, consideration of the lesion sites and associated perceptual deficits in cases of prosopagnosia suggest that the visual system is likely to have been damaged.

On our view, the phenomenon of covert recognition in prosopagnosia is no less interesting or important if it is explained in terms of incomplete damage to the face recognition system. The fact that recognition can be manifest in different ways, some of which are accompanied by conscious awareness and others not, and that this distinction appears to be coextensive with their vulnerability to brain damage, is of obvious high importance to the understanding of perception and the brain. We are merely pointing out the most straightforward explanation of this dissociation -- that the face recognition system is spared, and the impairment in overt recognition tasks arises elsewhere -- is not the only possibility. In addition to questioning the prevailing hypothesis, we are offering a new one, that has the advantage of being more explicit about mechanism.

Covert recognition in other syndromes. Could the same type of hypothesis account for other dissociations between perception with and without awareness? In principle it could, although there is no reason to assume that all of the syndromes reviewed earlier will have the same explanation. In some cases, there is evidence favoring the involvement of at least partially distinct systems subserving overt and covert perception. Although it has been suggested that the visual abilities in blindsight may be mediated by residual functioning of the cortical visual system (Campion, Latto & Smith, 1983), there is evidence of disproportionate involvement of the subcortical visual system in at least some of these abilities. For example, asymmetries in the processing of visual stimuli between nasal and temporal hemifields suggests that the subcortical visual system (which is receives disporortionate input from the temporal hemifield), plays a primary role in covert visual abilities in this syndrome (e.g., Rafal, Smith, Krantz, Cohen, A. & Brennan, 1990). Implicit reading in pure alexia may also be carried out by different systems from those subserving normal

explicit reading. The hypothesis of right hemisphere mediation of implicit reading (in contrast to the predominant role of the left hemisphere in normal reading) is supported by the relative absence of implicit reading for abstract words, function words, and grammatical inflections, and the lack of access to phonology, all characteristics of the right hemisphere lexicon (Coslett & Saffran, in press). Nevertheless, it is conceivable that this profile of reading abilities would also emerge from damage to the left hemisphere reading system. For example, differences between word classes such as word frequency and availability of collatoral support from semantic representations may confer different degrees of robustness to damage on them, and differences in the regularity of mapping among print, meaning and phonology could also affect the robustness of these mappings in the network after damage.

Findings of covert recognition in parietal-damaged patients may be best explained in terms of the residual functioning of a damaged visual system, rather than a dissociation between conscious and unconscious visual information processors. Farah, Monheit and Wallace (in press) showed that the dissociation observed by Volpe et al. (1979) could be obtained in normal subjects simply by placing a translucent sheet of drafting stock over the left half of the display to degrade subjects' perception of stimuli on the left. We also showed that the dissociation could be eliminated in parietal-damaged patients with extinction when the overt and covert tasks were matched for the precision of visual perception required by each. This implies that the dissociation between overt and covert perception after parietal damage is also due to differences in the quality of information needed to support performance in the two types of task, with performance in the covert task again more robust to low quality information. The nature of the information degradation appears to be different in the two cases, however. In prosopagnosia, what is degraded is the pattern

of previously learned associations within the visual recognition system, so that the effects of prior learning on perception are disrupted. In extinction, there is no structural impairment of representations, as evidenced by the ability of patients with extinction to perceive normally in the absence of a simultaneously occuring ipsilesional stimulus. Rather, the locus of degradation appears to be prior to visual recognition, affecting the input to visual recognition memory. This is consistent with our ability to simulate covert recognition in extinction by degrading the stimulus input to normal subjects.

The most general implication of the present model for the study of covert perception is that is demonstrates another mechanism by which overt and covert processing can be dissociated, beyond those previously considered. Schacter et al. (1988) list three general types of account for overt/covert dissociations: Conservative response bias in the overt tasks; disconnection from language (on the assumption that language is more involved in the overt tasks); and truely distinct and thus dissociable processing systems for overt and covert performance. To these we would add a fourth: differential susceptibility to damage of overt and covert performance. We have shown how knowledge can reside in a damaged network but be inaccessible for most purposes, for reasons quite distinct from the signal detection theory concept of bias, or a disconnection from other systems.

Prosopagnosia. The finding that some prosopagnosic patients manifest covert recognition and others do not has been taken as an indication that there are two different types of prosopagnosia, one caused by a visual perceptual impairment per se and the other by a disconnection of visual recognition and other, conscious, mental systems (e.g., Newcombe, Young & De Haan, 1989). However, our analysis suggests that these two groups of prosopagnosic patients are more likely to differ severity than in kind. In particular,

the similarity of the effects obtained when we lesioned face input units and face hidden units suggests that the presence of covert recognition may not be a precise way of discriminating different functional loci of damage. In fact, lesions further downstream in our model also showed similar effects to the ones reported here. This is a consequence of the highly interactive nature of the model. The nonlocalizability of errors resulting from damage in interactive models has been discussed in detail by Hinton and Shallice (1991) for their model of reading.

Consciousness. The dissociations between covert and overt perception in prosopagnosia and in other syndromes are of interest independent of the association between overt perception and consciousness. The fact that knowledge may be accessible in certain tasks and not in others is somewhat counterintuitive, and promises insights into how information is represented in the brain. Indeed, this has been the focus of the present paper. However, it cannot be denied that part of the fascination of these dissociations comes from the involvement of consciousness, specifically the patients' seemingly earnest denials of conscious awareness of stimulus properties of which they show knowledge in certain tasks. On the basis of our research, can we offer any insights into consciousness?

To the extent that the presence or absence of conscious awareness is coextensive with the distinction between tasks that can detect residual knowledge in a damaged system and tasks that cannot, as it so far appears to be, then on the basis of our simulations we can tentatively conclude this: Whatever precisely we mean by consciousness of perception (see Allport, 1988), its neural substrates need not be separate from the neural substrates of perception per se. The Cartesian idea by which there is some entity outside of the visual system per se which must receive the output of the visual system in order for conscious perception to occur

seems necessary according to the prevailing interpretation of covert recognition as the product of an intact visual system, However, if covert recognition reflects residual knowlege in a damaged visual system, then the Cartesian view is not necessarily true. In this case, visual recognition and awareness of recognition could both be products of the functioning of modality-specific visual cortex. Of course, this raises the question of why we can only be conscious of relatively high quality information in our visual systems. Unfortunately, this is a question for which we have no good answer.

Acknowledgements

This research was supported by ONR grant NS0014-89-J3016, NIMH grant R01 MH48274, NIH career development award K04-NS01405, and a grant from the McDonnell-Pew Program in Cognitive Neuroscience to the first author, an ONR graduate fellowship to the second author, and NSF grant BNS 88-12048 to James L. McClelland. The authors thank Clark Glymour and Jay McClelland for their insightful comments on an earlier draft of this manuscript.

Table 1
Overt Identification in 10-Alternative Forced Choice

(a)
Hidden Unit Damage

Amount of Damage (percent units eliminated)	Percent Correct Mean	Standard Error
12.5	62	4.9
25	43	5.0
37.5	43	4.8
50	24	4.3
62.5	14	3.5
75	13	3.4
87.5	8	2.7

(b)
Input Unit Damage

Amount of Damage (percent units eliminated)	Percent Correct Mean	Standard Error
12.5	64	4.8
25	56	5.0
37.5	41	4.9
50	26	4.4
62.5	17	3.8
75	17	3.8
87.5	19	3.9

Table 2
Savings in relearning correct relative to incorrect face-name pairings

(a)
Hidden Unit Lesion

Amount of Damage (% units eliminated)	Percent Correct			
	Correct Pairings		Incorrect pairings	
	0 epochs Mean SE	10 epochs Mean SE	0 epochs Mean SE	10 epochs Mean SE
12.5	58.0 5.0	98.0 1.4	6.0 2.4	10.0 3.0
25.0	26.0 4.4	82.0 3.9	8.0 2.7	14.0 3.5
37.5	34.0 4.8	62.0 4.9	8.0 2.7	18.0 3.9
50.0	18.0 3.9	50.0 5.1	10.0 3.0	12.0 3.3
62.5	20.0 4.0	36.0 4.8	4.0 2.0	6.0 2.4
75.0	12.0 3.3	36.0 4.8	18.0 3.9	14.0 3.5
87.5	6.0 2.4	24.0 4.3	16.0 3.7	12.0 3.3

(b)
Input Unit Lesion

Amount of Damage (% units eliminated)	Percent Correct			
	Correct Pairings		Incorrect pairings	
	0 epochs Mean SE	10 epochs Mean SE	0 epochs Mean SE	10 epochs Mean SE
12.5	68.0 4.7	98.0 1.4	0.0 0.0	4.0 2.0
25.0	58.0 5.0	96.0 2.0	8.0 2.7	4.0 2.0
37.5	32.0 4.7	72.0 4.5	8.0 2.7	4.0 2.0
50.0	20.0 4.0	74.0 4.4	6.0 2.4	18.0 3.9
62.5	16.0 3.7	18.0 3.9	10.0 3.0	10.0 3.0
75.0	12.0 3.3	46.0 5.0	6.0 2.4	24.0 4.3
87.5	10.0 3.0	18.0 3.9	2.0 1.4	20.0 4.0

Table 3
Settling time for familiar and unfamiliar face patterns

(a)
Hidden Unit Lesion

Amount of Damage (% units eliminated)	Number of cycles			
	Familiar face		Unfamiliar face	
	Mean	SE	Mean	SE
12.5	154.3	6.9	278.1	14.1
25	176.5	11.5	256.3	12.8
37.5	170.6	10.2	267.4	13.8
50	162.5	10.8	223.5	14.3
62.5	145.0	8.5	191.6	10.1
75	124.2	6.9	162.5	8.3
87.5	119.9	12.0	138.3	7.4

(b)
Input Unit Lesion

Amount of Damage (% units eliminated)	Number of cycles			
	Familiar face		Unfamiliar face	
	Mean	SE	Mean	SE
12.5	187.4	10.2	284.2	18.4
25	222.3	10.8	276.4	15.2
37.5	255.9	14.2	255.7	14.3
50	258.3	11.2	306.6	18.1
62.5	255.3	14.4	273.3	15.0
75	293.7	14.0	296.6	14.7
87.5	368.9	20.8	359.2	18.8

Table 4
Settling time for novel patterns before and after a small amount of learning

(a)
Hidden Unit Lesion

Amount of Damage (% units eliminated)	Number of cycles			
	Before learning		After training 5 epochs	
	Mean	SE	Mean	SE
12.5	376.0	18.8	365.0	21.1
25.0	476.2	26.9	430.9	26.3
37.5	475.3	23.9	419.1	30.0
50.0	513.9	29.2	465.8	25.7
62.5	506.6	24.9	438.3	23.6
75.0	521.6	25.7	467.2	25.3
87.5	669.4	35.0	431.6	18.2

(b)
Input Unit Lesion

Amount of Damage (% units eliminated)	Number of cycles			
	Before learning		After training 5 epochs	
	Mean	SE	Mean	SE
12.5	369.5	22.0	365.8	22.5
25.0	473.3	27.3	446.7	29.3
37.5	464.3	21.4	486.1	33.4
50.0	529.4	31.4	474.3	23.4
62.5	506.3	26.9	462.8	26.6
75.0	538.5	25.3	468.2	21.7
87.5	469.6	22.6	544.2	36.5

Table 5
Overt occupation categorization

(a)
Hidden Unit Damage

Amount of Damage (percent units eliminated)	Percent Correct Mean	Standard Error
12.5	85.5	3.1
25.0	77.0	3.3
37.5	74.0	4.4
50.0	62.5	4.5
62.5	59.5	5.1
75.0	53.0	5.0
87.5	51.5	4.5

(b)
Input Unit Damage

Amount of Damage (percent units eliminated)	Percent Correct Mean	Standard Error
12.5	88.0	1.2
25.0	86.5	2.3
37.5	73.0	3.4
50.0	64.5	4.4
62.5	59.5	4.9
75.0	57.5	4.8
87.5	57.0	5.0

Table 6

Time to categorize names according to their occupation alone and in the presence of same- and different-category faces

(a)
Hidden Unit Damage

Amount of Damage (% units eliminated)	Cycles for correct occupation unit to attain positive activation					
	No face Mean	SE	Same cat. Mean	SE	Diff cat. Mean	SE
12.5	50.5	5.5	49.0	4.4	142.8	13.6
25.0	49.8	5.9	75.3	9.1	150.5	12.2
37.5	55.3	5.8	76.1	9.3	110.1	8.9
50.0	70.8	13.3	81.6	8.9	114.0	8.8
62.5	66.6	13.9	59.6	6.5	82.1	6.9
75.0	73.0	15.1	84.0	7.6	98.0	7.8
87.5	62.8	9.1	72.9	9.2	66.9	5.0

(b)
Input Unit Damage

Amount of Damage (% units eliminated)	Cycles for correct occupation unit to attain positive activation					
	No face Mean	SE	Same cat. Mean	SE	Diff cat. Mean	SE
25.0	69.6	9.2	101.7	9.2	77.0	16.0
37.5	55.8	5.6	88.9	8.3	96.3	20.2
12.5	52.0	4.4	115.4	9.3	55.3	9.3
50.0	101.3	10.4	146.5	13.1	126.9	21.6
87.5	76.3	7.6	78.8	6.7	144.0	27.2
62.5	106.0	11.2	131.8	11.3	123.0	24.2
75.0	107.5	10.9	120.6	10.7	130.9	21.9

References

Allport, A. (1988) What concept of consciousness? In A.J. Marcel & E. Bisiach (Eds.) Consciousness in Contemporary Science. Oxford: Clarendon Press.

Bauer, R. M. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the guilty knowledge test. Neuropsychologia, 22, 457-469.

Bruyer, R. (1991). Covert face recognition in prosopagnosia: A review. Brain and Cognition, 15, 223-235.

Bruyer, R., C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard E., & Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. Brain and Cognition, 2, 257-284.

Burton, A. M., Young, A. W., Bruce, V., Johnston, R. A. & Ellis, A. W. (in press) Understanding covert recognition. Cognition.

Coslett, H. B., & Saffran, E. M. (1989). Evidence for preserved reading in 'pure alexia'. Brain, 112, 327-359.

De Haan, E. H. F., Young, A. W. & Newcombe, F (1987a). Faces interfere with name classification in a prosopagnosic patient. Cortex, 23, 309-316

De Haan, E. H. F., Young, A. W. & Newcombe, F. (1987b). Face recognition without awareness. Cognitive Neuropsychology, 4, 385-415

Etcoff, N. L., Freeman, R., & Cave, K. R. (1991). Can we lose memories of faces? Content specificity and awareness in a prosopagnosic. Journal of Cognitive Neuroscience, 3, 25-41.

Farah, M. J., Monheit, M. A. & Wallace, M. A. (in press) Unconscious perception in the extinguished hemifield: Re-evaluating the evidence. Neuropsychologia.

Greve, K. W. & Bauer, R. M. (1990) Implicit learning of new faces in prosopagnosia: An application of the mere exposure paradigm. Neuropsychologia, 28, 135-141.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In Rumelhart, D. E., McClelland, J. L., & the PDP research group, editors, Parallel Distributed Processing: Explorations in the Microstructure of cognition. Volume 1: Foundations, chapter 3, pages 77-109. MIT Press, Cambridge, MA.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann Machines. In Rumelhart, D. E., McClellard, J. L., & the PDP research group, editors, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, chapter 7, pages 282-317. MIT Press, Cambridge, MA.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. Psychological Review, 98 (1), 74-95.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Science, U. S. A., 81, 3088-3092.

Jacoby, L. L. (1984) Incidental vs. intentional retrieval: Remembering and awareness as separate issues. In L. R. Squire and N. Butters, (Eds.) The Neuropsychology of Human memory. New York: Oxford University Press.

Marshall, J. C., & Halligan, P. W. (1988). Blindsight and insight in visuospatial neglect. Nature, 336 766-767.

McClelland, J. L., Rumelhart, D. E., & the PDP research group (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models. MIT Press, Cambridge, MA.

Movellan, J. R. (1990) Contrastive Hebbian learning in the continuous Hopfield model. In D. S. Touretzky, G. E. Hinton, & T. J. Sejnowski (Eds.) Proceedings of the 1989 Connectionist Models Summer School. San Mateo, CA: Morgan Kaufman, pp. 10-17.

Newcombe, F., Young, A. & De Haan, E. H. F. (1989). Prosopagnosia and object agnosia without covert recognition. Neuropsychologia, 27, 179-191.

Rafal, R., Smith, J., Krantz, J., Cohen, A., Brennan, C. (1990). Extrageniculate vision in hemianopic Humans: Saccade inhibition by signals in the blind field. Science, 250, 118-121

Renault, B., Signoret, J. L., DeBruille, B., Breton, F. & Bolgert, F. (1989). Brain potentials reveal covert facial recognition in prosopagnosia. Neuropsychologica, 27, 905-912.

Rumelhart, D. E., McClelland, J. L. and the PDP Research Group (1986). Parallel distributed processing: Explorations in the microstructure of cognition, Vol. I: foundations. Cambridge, Mass.: Bradford Books.

Schacter, D. L., McAndrews, M. P., & Moscovitch, M. (1988). Access to consciousness: dissociations between implicit and explicit knowledge in neuropsychological syndromes. In L. Weiskrantz (Ed.), Thought Without Language. Oxford: Oxford University Press.

Seidenberg, M. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96, 523-568.

Shallice, T., & Saffran, E. (1986). Lexical processing in the absence of explicit word identification: Evidence form a letter-by-letter reader. Cognitive Neuropsychology, 3, 429-458.

Snodgrass, J. G., Corwin, J. & Feenan, K. (1990). The pragmatics of measuring priming: Applications to normal and abnormal memory. Journal of Clinical and Experimental Neuropsychology, 12, 61.

Tranel, D. & Damasio, A. R. (1985). Knowledge without awareness: An automatic index of facial recognition by prosopagnosics. Science, 228, 1453-1454.

Tranel, D. & Damasio, A. R. (1988) Nonconscious face recognition in patients with face agnosia. Behavioral Brain Research, 30, 235-249.

Volpe, B. T., LeDoux, J. A., and Gazzaniga, M. S. (1979). Information processing of visual stimuli in an 'extinguished' field. Nature 282, 722-724.

Wallace, M. A. & Farah, M. J. (submitted) Savings in relearning as evidence for covert recognition in prosopagnosia.

Weiskrantz, L. (1990) Outlooks for blindsight: Explicit methodologies for implicit processes. (Ferrier Lecture, 1989) Proceedings of the Royal Society of London, B239, 247-278.

Young, A. W. & De Haan, E. H. F. (1988). Boundaries of covert recognition in prosopagnosia. Cognitive Neuropsychology, 5, 317-336.

Young, A. W., Hay, D. C. & Ellis, A. W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognising people. British Journal of Psychology, 76, 495-523.

Young, A. W., Hellawell, D. & De Haan, E. H. F (1988). Cross-domain semantic priming in normal subjects and a prosopagnosic patient. Quarterly Journal of Experimental Psychology, 38A, 297-318.

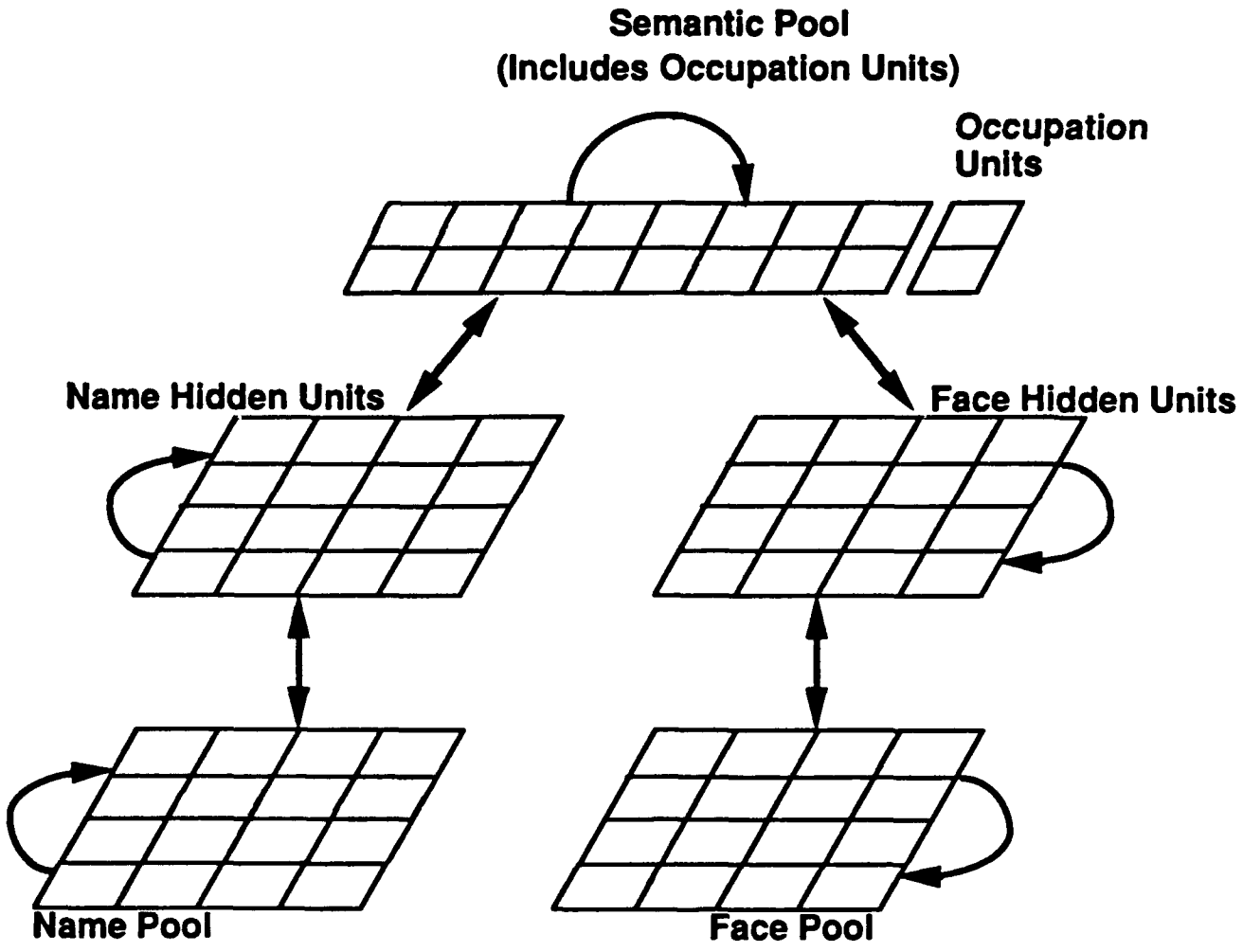
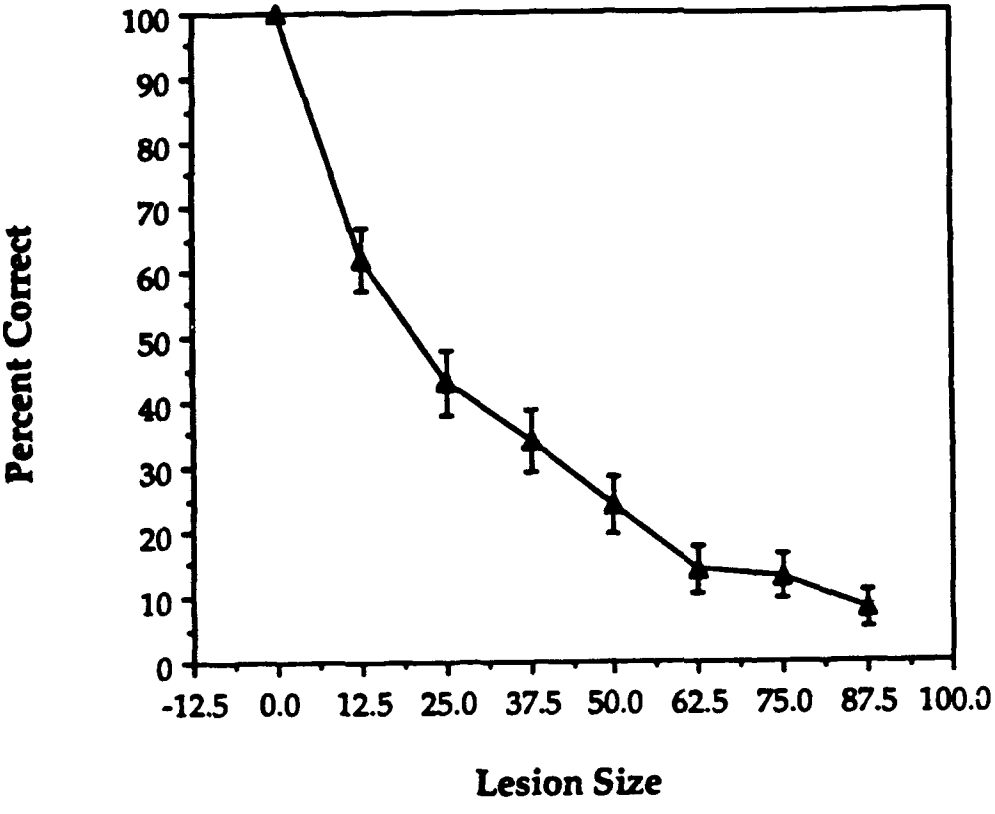
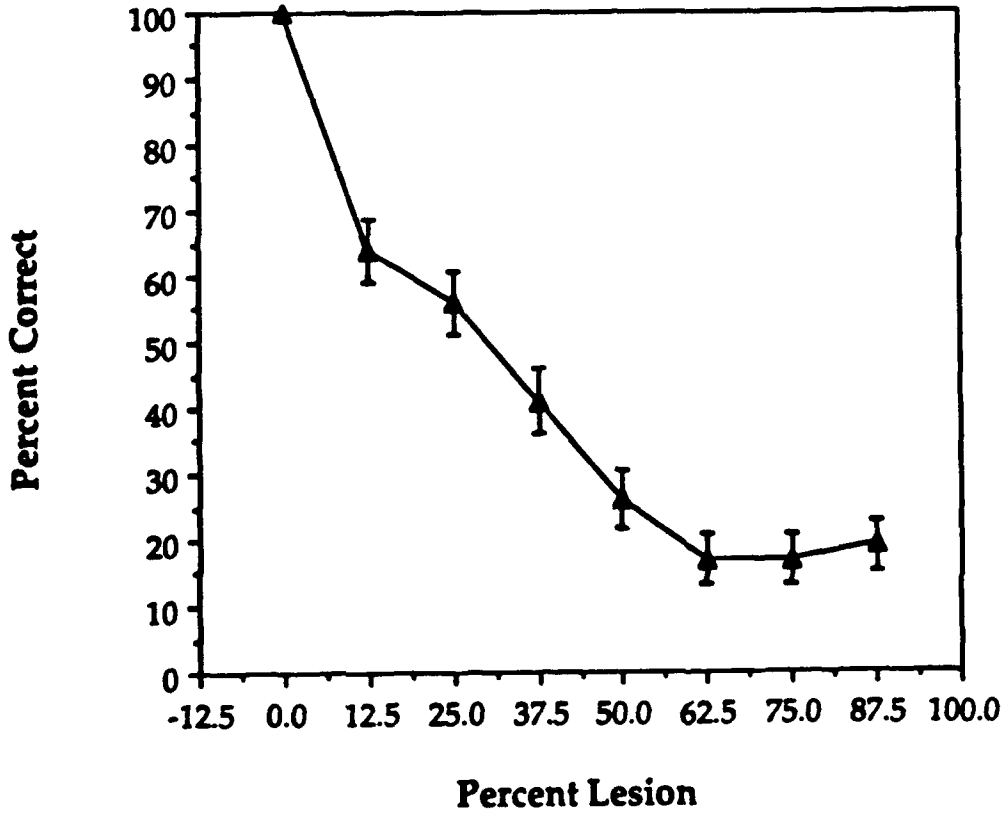

Semantic Pool
(Includes Occupation Units)
Occupation
Units
Name Hidden Units
Face Hidden Units
Name Pool
Face Pool

Fig 2

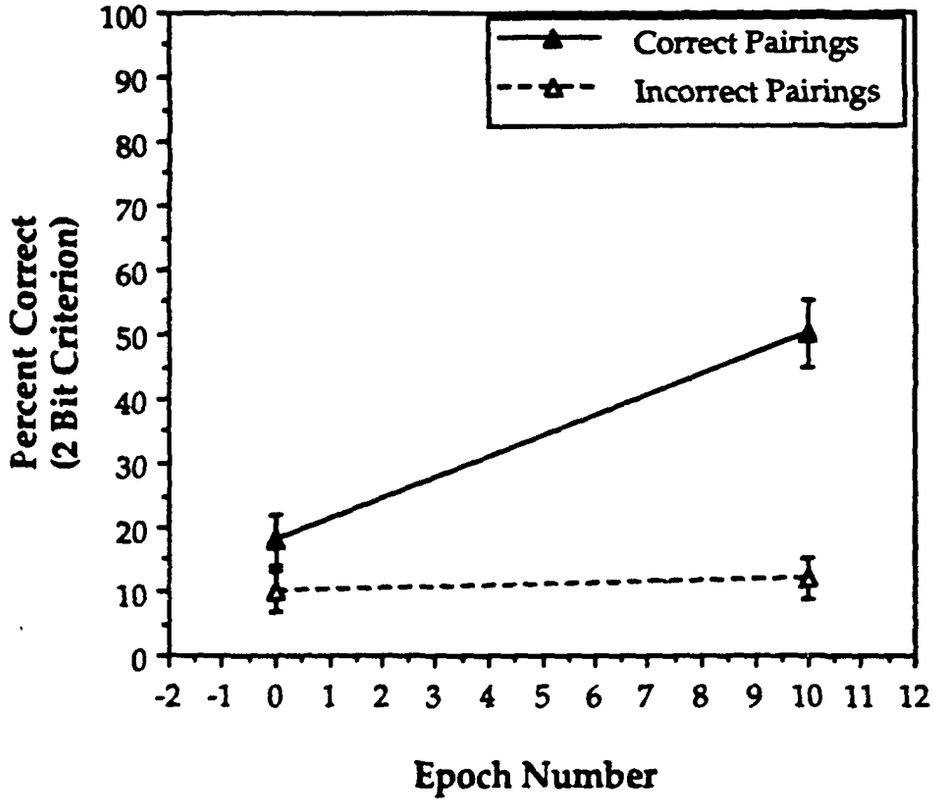
Overt Performance: Hidden Unit Lesions (Forced Choice with 10 Alternatives)



Overt Performance: Face Pool Lesions (Forced Choice with 10 Alternatives)



Savings in Relearning: 50% Hidden Unit Lesions
Percent Correct as a Function of Training



Savings in Relearning: 62.5% Hidden Unit Lesion
Percent Correct as a Function of Training

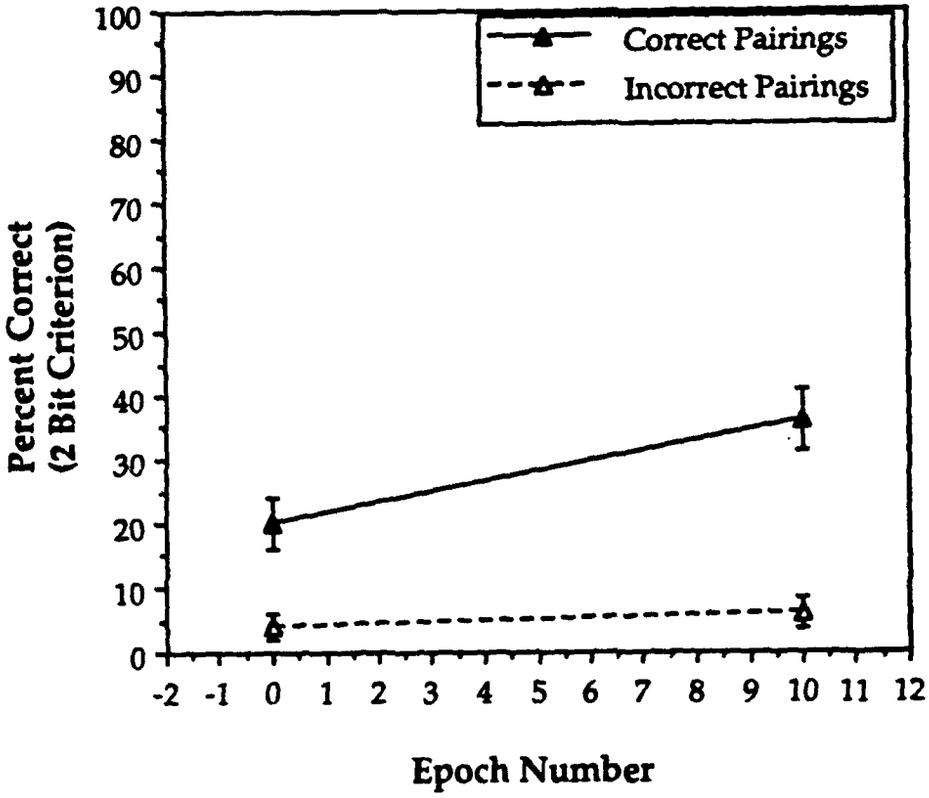
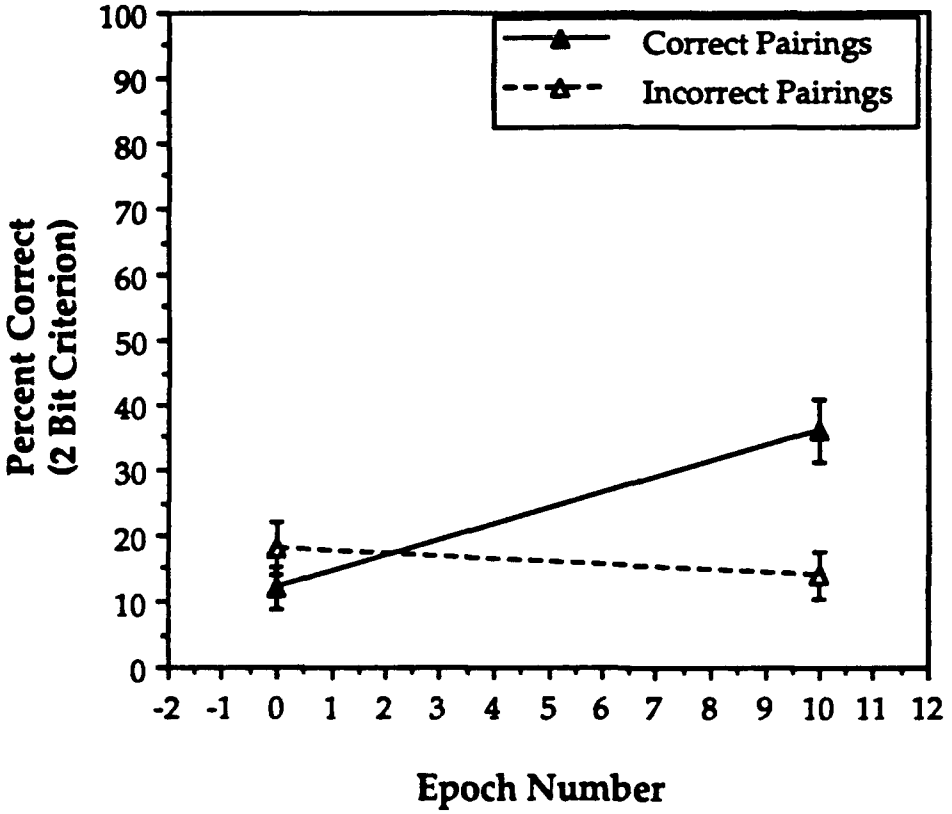
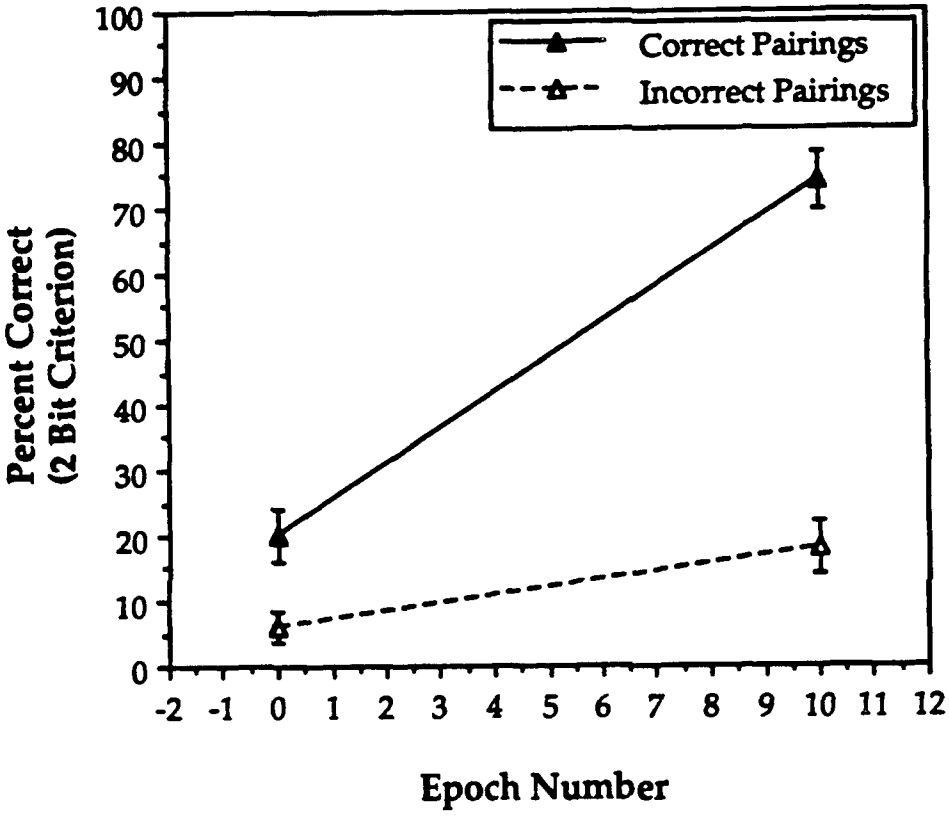


Fig 3c

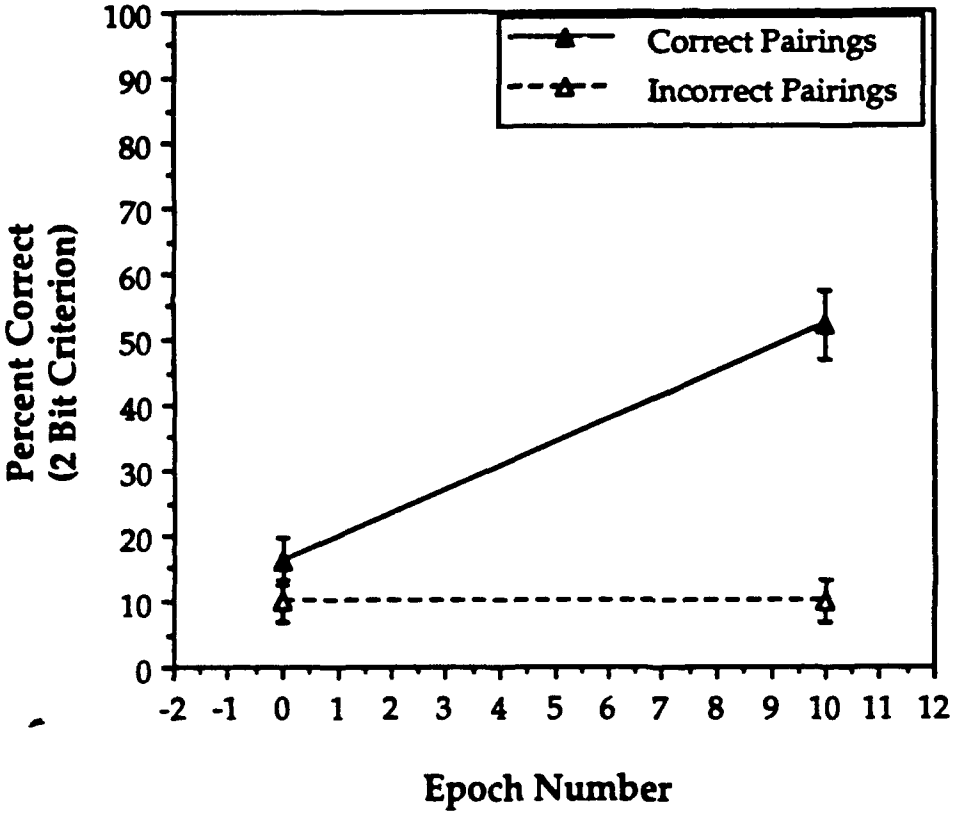
Savings in Relearning: 75% Hidden Unit Lesion
Percent Correct as a Function of Training



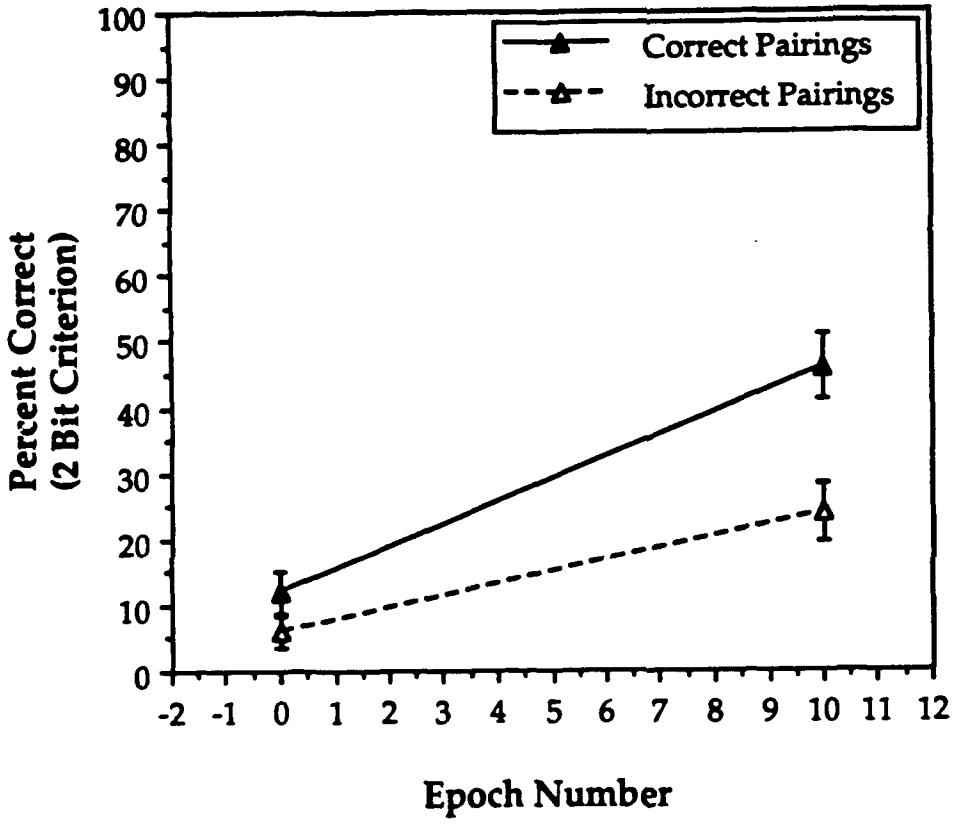
Savings in Relearning: 50% Face Pool Lesion
Percent Correct as a Function of Training



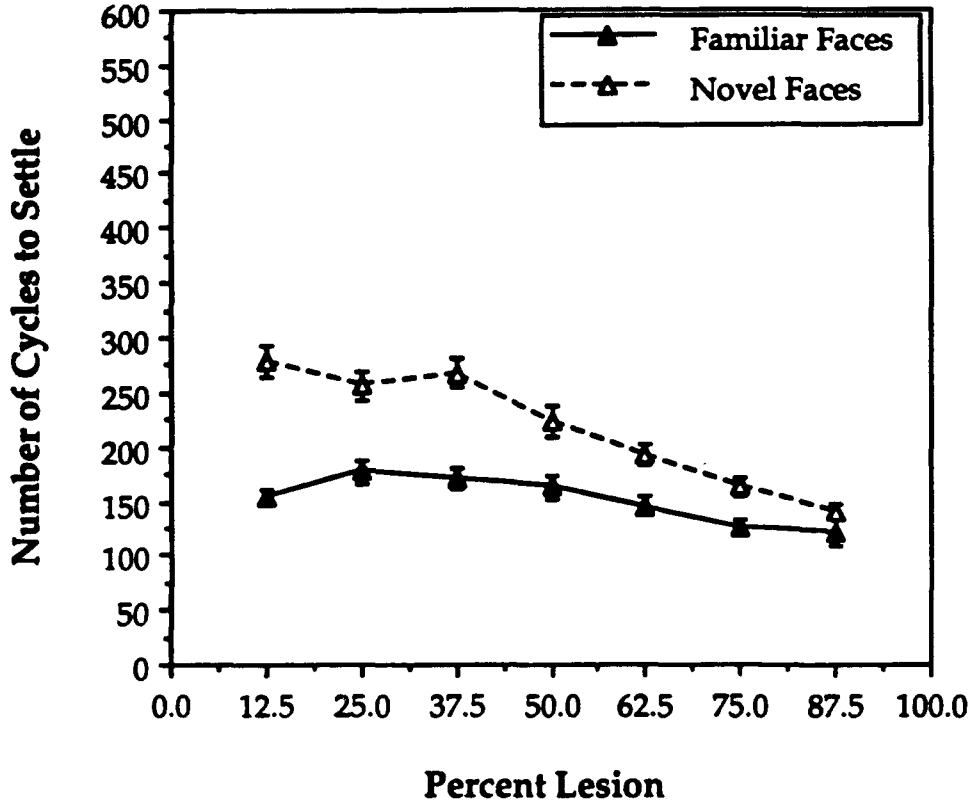
Savings in Relearing: 62.5% Face Pool Lesion
Percent Correct as a Function of Training



Savings in Relearning: 75% Face Pool Lesion
Percent Correct as a Function of Training



Same-Different Matching Task
Hidden Unit Lesions



Same-Different Matching Task
Face Pool Lesions

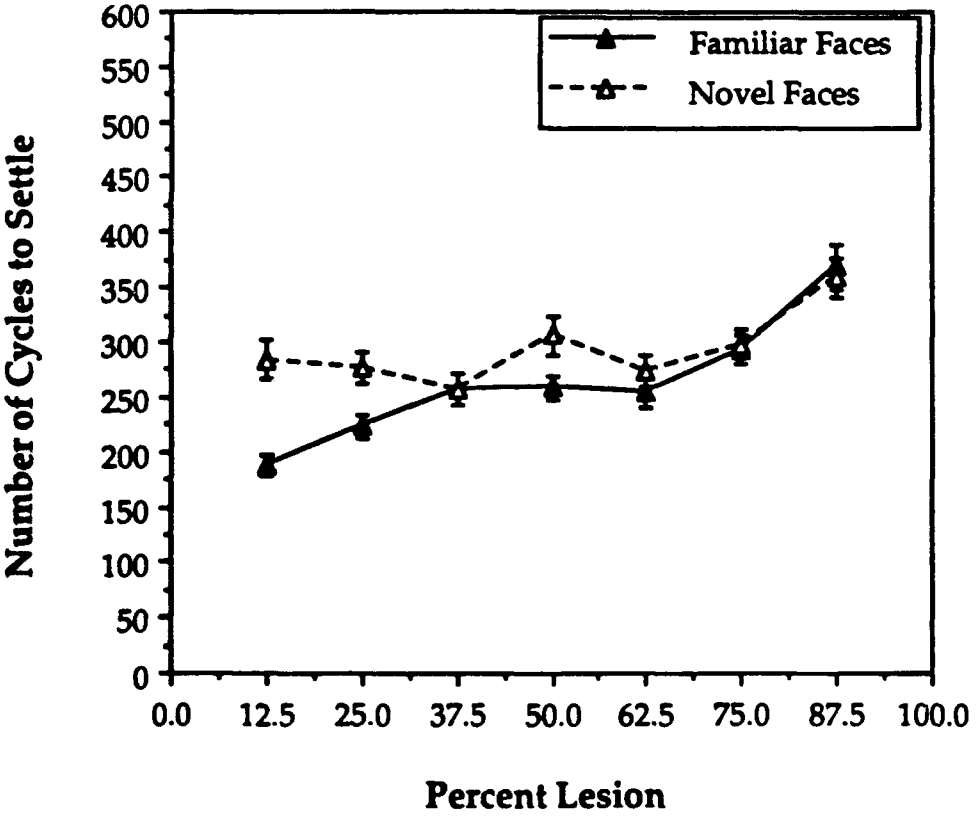
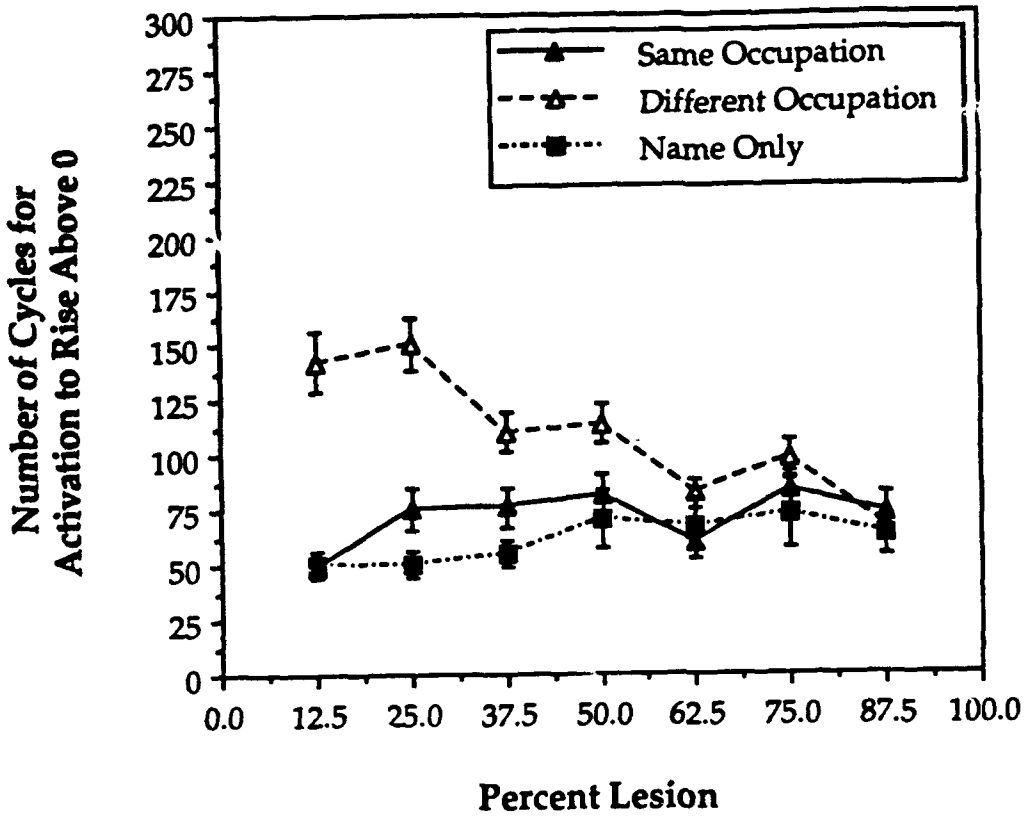


Fig 5

Semantic Priming Task
Hidden Unit Lesions



Semantic Priming Task
Face Pool Lesions