Object Recognition and Sensitive Periods:

A Computational Analysis of Visual Imprinting

Randall C. O'Reilly
&
Mark H. Johnson

Technical Report PDP.CNS.93.1
February 1993

Parallel Distributed Processing and Cognitive Neuroscience

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA

Western Psychiatric Institute and Clinic
University of Pittsburgh
Pittsburgh, PA

Neural and Behavioral Sciences
University of Southern California
Los Angeles, CA

MRC Applied Psychology Unit
Cambridge, England

Abstract

Evidence from a variety of methods suggests that a localized portion of the domestic chick brain, the Intermediate and Medial Hyperstriatum Ventrale (IMHV), is critical for filial imprinting. Data further suggest that IMHV is performing the object recognition component of imprinting, as chicks with IMHV lesions are impaired on other tasks requiring object recognition. We present a neural network model of translation invariant object recognition developed from computational and neurobiological considerations that incorporates some features of the known local circuitry of IMHV. In particular, we propose that the recurrent excitatory and lateral inhibitory circuitry in the model, and observed in IMHV, produces hysteresis on the activation state of the units in the model and the principal excitatory neurons in IMHV. Hysteresis, when combined with a simple Hebbian covariance learning mechanism, has been shown in earlier work to produce translation invariant visual representations. To test the idea that IMHV might be implementing this type of object recognition algorithm, we have used a simple neural network model to simulate a variety of different empirical phenomena associated with the imprinting process. These phenomena include reversibility, sensitive periods, generalization, and temporal contiguity effects observed in behavioral studies of chicks. In addition to supporting the notion that these phenomena, and imprinting itself, result from the IMHV properties captured in the simplified model, the simulations also generate several predictions and clarify apparent contradictions in the behavioral data.

Introduction

The problem of computing object-based visual representations can be construed as the development of invariancies to visual dimensions irrelevant for object identity. This view, when implemented in a neural network, suggests a different set of algorithms for computing object-based visual representations than the "traditional" approach pioneered by Marr (1982). A relatively simple neural network algorithm for developing translation invariant object recognition has recently been proposed (Földiák, 1991; O'Reilly, 1992), and the version proposed by O'Reilly (1992) makes several claims about the kind of neural properties necessary to implement such an algorithm. Instead of attempting to test such claims in the complex mammalian nervous system, we have taken the approach of studying a well known and simpler vertebrate object recognition system: visual imprinting in the chick.

Filial imprinting is the process whereby young precocial[1] birds learn to recognize the first conspicuous object that they see after hatching. The original work of Lorenz (1935;1937) on imprinting has given rise to half a century of active research on this process by ethologists and psychologists from a variety of different backgrounds. Evidence for imprinting has also been reported in some mammalian species such as spiny mice and guinea pigs, but the measure that yields this evidence (approach and following behavior) cannot be used successfully with species that are relatively immobile for some time after birth, such as humans.

The imprinting phenomenon is ideal for testing neural theories of object recognition because of the convenient and robust behavioral measure of visual discrimination, and the large body of data that this has produced. Also, the area of the chick brain that subserves this imprinting process has been identified, and some of its neurobiological properties studied. Therefore, it is now possible to assess a model of imprinting in the chick both with regard to its fidelity to these properties and the behavioral effects they produce. The model of visual object recognition in the chick proposed herein incorporates several features of IMHV's neural circuitry, and is evaluated on the extent to which it accounts for the distinctive behavioral features of the imprinting phenomenon.

Lorenz argued that the learning process underlying imprinting was unique on the basis of two claims; first, that once a preference had been acquired it was irreversible and would influence behavior in later life, and second, that there was a sharply defined critical time or *sensitive period* within which the learning could take place. Both of these claims have subsequently been brought into question. There is now abundant evidence that a preference for a particular stimulus can be reversed to a preference for another object under certain conditions (e.g., Salzen & Meyer, 1968). A weaker version of Lorenz's original claim of irreversibility, namely that information about the first imprinting object is overridden but not forgotten (Jaynes, 1956), has received some support from recent studies (e.g., Cherfas & Scott, 1981). Sluckin & Salzen (1961) and Bateson (1966) have argued that the sensitive period for imprinting is not as circumscribed as Lorenz claimed. Since the sensitive period can be extended by delaying the onset of exposure to a conspicuous object these authors argue it is better viewed as a self-terminating process.

[1]That is, birds which are capable of fending for themselves at birth

Our model of imprinting clarifies the issue of a self-terminating sensitive period by demonstrating how it can result from the properties of a simple neural network model. In addition to exploring the conditions under which an initial imprinting preference can and cannot be reversed, we explore other phenomena such as generalization and blending resulting from temporal contiguity. Throughout, we emphasize how these phenomena result from certain properties of a neural system designed to perform visual object recognition.

We adopt a multi-level approach to modeling the phenomena. The basic mechanism of translation-invariant object recognition can be specified and implemented with a relatively abstract neural network model, as it is based on rather general properties of neural circuitry, and not on the detailed properties of individual neurons. Using an abstract neural network model offers the advantage of simplicity and explanatory clarity—since the model only has a few simple properties, the phenomena that result can be related more directly to these properties. However, relating such a simple model to both detailed behavioral and neural data can be difficult; in general this model makes qualitative rather than quantitative predictions. For this reason, it is necessary to also construct a more realistic neural model which is based on anatomical and physiological measurements of actual IMHV neurons. Such a model is planned for future research. Issues that arise in the present model that can only be addressed in this more detailed model will be noted as such, to provide an indication of the role of the future model and its relation to the present one.

The Neural Basis of Imprinting

With a variety of neuroanatomical, neurophysiological and biochemical techniques, Horn, Bateson and their collaborators have established over a number of years that a particular region of the chick forebrain, referred to as the Intermediate and Medial part of the Hyperstriatum Ventrale (IMHV), is essential for imprinting (see Horn, 1985; Horn & Johnson, 1989; Johnson, 1991 for reviews). This region receives input from the main visual projection areas of the chick, and may be analogous to mammalian association cortex (Horn, 1985). See figure 1 for a diagram of the visual inputs to IMHV.

IMHV is a small "sausage-shaped" region immediately around the midpoint between the anterior and posterior poles of the cerebral hemispheres. This same region or co-extensive areas have also been identified by other groups studying visual imprinting (Kohsaka et al., 1979). If IMHV is indeed a crucial site for imprinting then damage to it prior to imprinting should prevent the acquisition of preferences, and its destruction after imprinting should render a chick amnesic for existing preferences. This has been confirmed. Lesions placed prior to training reduced chicks' ability to learn, and those placed shortly after training impaired chicks' ability to subsequently recognize the "familiar" object. That is, the chicks no longer showed a preference for the object to which they were initially exposed. Importantly, these lesions had no effect on several other behaviors and learning tasks, demonstrating that IMHV is specialized for a specific task. For example, while chicks with IMHV lesions easily learn to press a particular pedal to obtain a reward, they are unable to recognize the imprinting stimulus with which they were rewarded (Johnson & Horn, 1986).

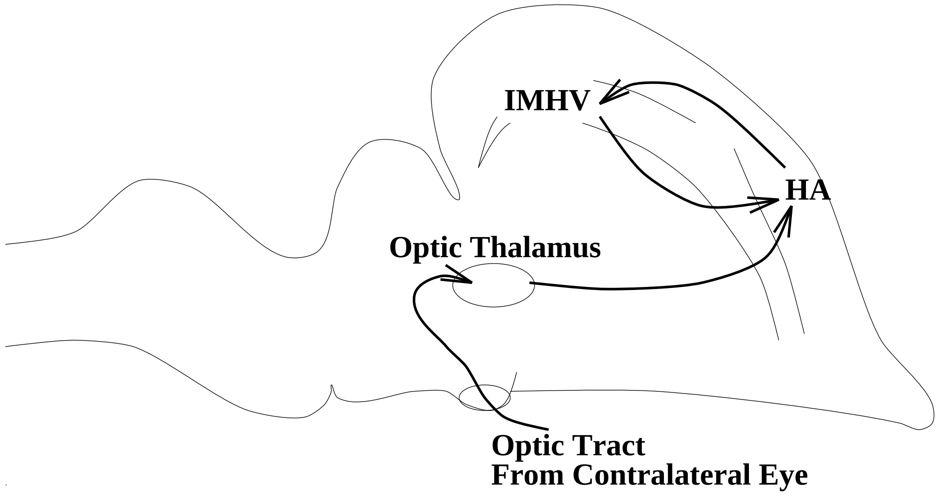

Figure 1: Visual inputs to IMHV from the thalamic pathway. Shown is a saggital view (strictly diagrammatic, as some areas are out of the plane with others) of the chick brain, with visual information coming in through the optic tract, which then synapses in the optic nucleus of the thalamus. This then projects to area HA (Hyperstriatum Accessorium), which connects reciprocally with IMHV. This pathway corresponds to the retina ⇒LGN ⇔V1,V4 ⇔IT pathway in mammals. There are other routes of visual input to IMHV, which are not shown in this figure (see Horn, 1985). The brain of a 2 day old chick is approximately 2 cm long.

The Case for Visual Imprinting as Object Recognition

Visual imprinting in the domestic chick can be easily studied in the laboratory in the following way. Chicks are hatched and reared in darkness before being exposed to a conspicuous object such as one of those shown in figure 2. This period of exposure is called *training* and usually lasts for a period of several hours. Hours or days later, the chick is given a preference test in which it is released in the presence of two objects—the object to which it was exposed earlier, and a novel object. The extent to which the chick attempts to approach the familiar object as opposed to the novel one is measured and a *preference score* calculated. If the chick has imprinted strongly, it shows a high preference for the familiar object. We may infer from this behavior that chicks acquire information about the visual characteristics of objects to which they are exposed.

While some authors have proposed that this information is acquired in an associative manner (e.g., Hoffman & Ratner, 1973), recent evidence from neurobehavioral studies strongly suggests parallels between imprinting in birds and processes underlying object recognition in primates. Evidence from both humans and non-human primates (see Farah, 1990 and Ungerleider & Mishkin, 1982 for reviews) indicates that processes of object recognition are neurally distinct from mechanisms underlying some other forms of learning such as simple associative conditioning. McCabe et al. (1982) showed that small, localized lesions to IMHV impair the ability of chicks to acquire information about an object to which they are exposed. Similar sized lesions to some other parts of the chick forebrain did not have this

Figure 2: Objects used in the training portion of imprinting with chicks

effect. Further, while IMHV lesioned chicks were unable to imprint to an object, they were able to learn a heat-reinforced pattern discrimination task.

This preliminary evidence for a dissociation between imprinting and forms of associative learning was extended in a study by Johnson & Horn (1986), in which groups of chicks learned to press a particular pedal for a reward of exposure to an a moving imprinting object. While intact chicks, and chicks with small lesions elsewhere in the forebrain, were able to both learn the operant component of the task (to a press a particular pedal), and to imprint onto the rewarding object, chicks with IMHV lesions learned only the operant component. Thus, while IMHV lesions impaired the ability to recognize the object, they did not impair the ability to learn an operant conditioning task, enabling one to eliminate possible alternative explanations of IMHV's role in imprinting.

However, further support for IMHV's role in visual object recognition comes from imprinting and other kinds of studies. Within the context of imprinting, Johnson & Horn (1987) established that IMHV lesions impair the ability of chicks to recognize a particular individual member of their own species. Further, IMHV also seems to be crucial for the recognition of individual members of the species outside the context of imprinting, in a mate choice situation (Bolhuis et al., 1989).

Another task that has been extensively studied in chicks is so-called *one-trial passive avoidance learning* (PAL) task. Chicks will spontaneously peck at small brightly colored objects. If, however, such an object is coated with an aversive tasting substance, such as methyl anthranilate (MeA), they will withhold their pecking toward that object. Davies et al. (1988) demonstrated that chicks with IMHV lesions are unable to learn not to peck at an object coated with MeA. The authors argued that this deficit relates to problems in the recognition of the small beads used, since the intact chicks selectively withheld their pecking to the color of bead which had been coated with MeA previously.

In conclusion, evidence from the effects of IMHV lesions on a variety of learning tasks supports the contention that the region is critical for object recognition. The deficits are not attributable to defects at lower levels in the visual system since chicks with such lesions perform other tasks requiring visual information as well as intact birds (Johnson & Horn, 1986; for review see Horn & Johnson, 1989). For example, IMHV lesioned chicks can learn a simple pattern discrimination (McCabe et al., 1982). Perhaps the best description of the functional effect of IMHV lesions would that it induces object agnosia (Horn, 1985; Johnson, 1991).

The Structure of IMHV

Having established that IMHV is involved in object recognition it is interesting to examine whether any changes take place in the structure of synapses within the region following imprinting. Horn et al. (1985) under-trained one group of birds on a red box (20 minutes exposure to an object) and over-trained another group (140 minutes exposure). A third group of chicks served as dark-reared controls. Following training, samples were taken from the IMHV of each hemisphere; ultra-thin sections were cut and examined under an electron microscope. There were no differences in any measure of synapse morphology between under-trained chicks and dark reared controls. Chicks that had been trained for 140 mins differed from the other two groups in only one measure of the synapse structure: the mean length of the postsynaptic density, the thickened part of the postsynaptic membrane, which had increased by 17 percent. The change was restricted to synapses on dendritic spines (axospinous) within the left IMHV.

The postsynaptic density appears to be a site of high neurotransmitter receptor density, so that an increase in the area of this region suggests an increase in the number of receptor sites. Following imprinting, McCabe & Horn (1988) measured the presence of a particular kind of receptor (NMDA) known to be important for synaptic modification (Collingridge & Bliss, 1987). They found a significant increase in the number of NMDA sites in the left IMHV of chicks compared with dark-reared controls. Further, they found a significant positive correlation between the number of NMDA receptor sites and the degree to which a chick preferred the familiar stimulus at testing, while other factors such as locomotor activity, etc. were not significantly correlated.

It would be reasonable to expect that the changes in synaptic structure and biochemistry just described would result in changes in the spontaneous firing of neurons within left IMHV. Several multicellular and single cell studies have provided evidence that the spontaneous activity of neurons in the left IMHV are influenced by imprinting training (Payne & Horn, 1984; Davey & Horn, 1991; Brown & Horn, 1992). Some of the findings from these experiments include the following:

- In general, there is a low level of spontaneous activity in IMHV, and a similarly low level of activity even when a training stimulus is presented. This suggests a sparse representation of visual stimuli, with activity probably regulated by inhibitory interneurons.
- While effects of training are not found in other brain regions, they appear to be widely distributed within IMHV. Thus, significant effects are only obtained when a large number of multicellular sites within IMHV are combined. This suggests that the representations in this region are distributed.
- When stimuli similar to the training stimulus are presented, "generalization" effects are observed in the sense that the same patterns of responses are evoked.

While regions of the avian forebrain may be analogous to mammalian cerebral cortex (Horn, 1985), recent cytoarchitectonic studies of IMHV have revealed that it is much simpler in structure. In contrast to the 6-layered structure with many distinct cell types (12 or more according to Douglas & Martin, 1990) found in mammalian cerebral cortex, there is no clear laminar structure of IMHV, and only 4 distinctive types of cells have been identified (Tömböl et al., 1988). In figure 3 we illustrate the basic intrinsic micro-circuit described by Tömböl et al. (1988).

The four cell types that Tömböl et al. (1988) identified consist of two types of principal neurons (PNs) somewhat similar to mammalian pyramidal neurons (although they lack true apical dendrites) which are probably excitatory. The type 1 PNs are spiny, large, and possess long bifurcating axons that probably project outside the region, while type 2 PNs are medium sized with thick spiny dendrites. The presence of a high density of spines is indicative of extensive afferentation (input), as is the case with the main input cells of the mammalian neocortex, the spiny stellate cells found in layer 4 (Douglas & Martin, 1990). As shown in the figure, the two types of PN cells are interconnected such that they have a characteristic positive feedback loop.

The other two classes of neuron identified are medium and small local circuit neurons (LCNs) which are probably inhibitory, receiving excitatory input from the PNs and projecting inhibitory output back onto them. It is not known if they also receive excitatory input from the afferents that excite the PNs. Thus, the LCNs are probably performing at least feedback inhibition, and possibly feedforward inhibition as well. Presumably, these inhibitory neurons are critical to prevent the positive feedback present in the PNs from becoming unstable, since positive feedback loops are intrinsically unstable. It should be noted that the types of PNs and LCNs and their characteristic interconnectivity found in IMHV are not commonly found in the neighboring regions that have been studied (Tömböl et al., 1988).

Two characteristics of the cytoarchitectonics of IMHV described above are important for our model: the existence of positive feedback loops between the excitatory principal neurons, and the extensive inhibitory circuitry mediated by the local circuit neurons. We will argue that these properties lead to a hysteresis of the activation state of PN's in IMHV, a feature that contributes to the development of translation invariant object-based representations.

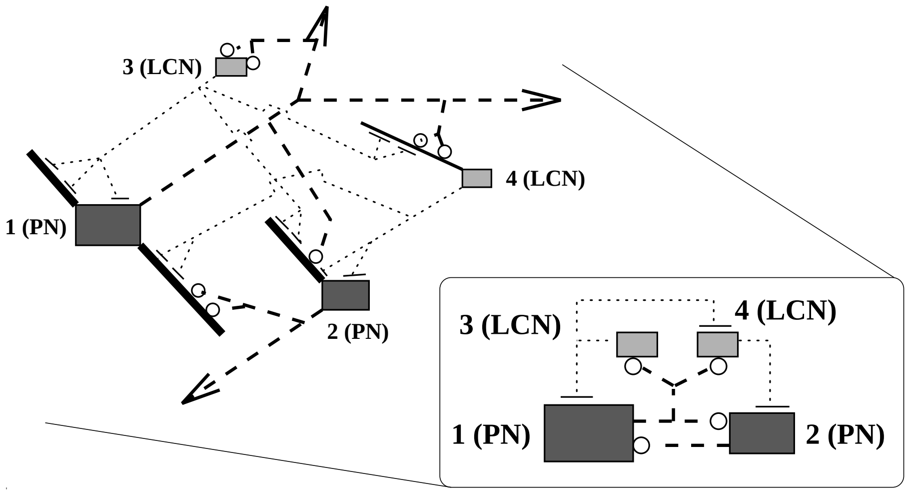

Figure 3: Schematic drawing summarizing the circuitry of IMHV at two levels of detail (simplified version in the box). Excitatory contacts are represented by open circles, and inhibitory ones by flat bars. Shown are the local circuit inhibitory neurons (LCN) and their reciprocal connectivity with the excitatory principal neurons (PN), and the recurrent excitatory connectivity between the principal neurons. In the detailed version, the thick solid lines are dendrites, while the axons are dashed or dotted lines. Both the inhibition and recurrent excitatory connectivity are used in the simplified model to produce hysteresis in the activation state of the IMHV. (After Tömböl, et al, 1988)

IMHV is extensively connected to other regions of the avian brain (Bradley et al., 1985). For example, it has prominent input from the avian primary visual projection area (the Hyperstriatum Accessorium, HA), as is shown in figure 1 and projects to regions thought to be involved in motor control, such as the archistriatum. We assume that type 2 PNs are the main *target* of projections to IMHV from area HA, while type 1 PNs are probably the main *source* of projections from IMHV, for the reasons mentioned earlier.

The Self-Organization of Invariant Representations

Despite the relative ease of its everyday execution, visual object recognition is a difficult computational problem. One of the principal reasons for this difficulty is also a clue to a potential solution: there are a practically infinite number of different images that a given object can project onto the retina. Deciphering which of the many thousands of familiar objects a given image represents is difficult because of this many-to-many correspondence. However, the ways in which a given object can produce different images on the retina are limited to a few dimensions of variability. This suggests that one could focus on eliminating the systematic variability due to these dimensions as a method of computing object-based representations.

These dimensions of variability arise principally from the projection of three-dimensional objects on to our two-dimensional retinas. Thus, the image produced by a given object can appear in a different location, orientation, and size depending on where it is located relative to our eyes. In addition, there are other dimensions of variability resulting from different lighting conditions, and from changes in the shape of the object itself. Thus, an object representation must be invariant with respect to all those dimensions on which the image can vary and still represent the object, while at the same time being selective enough to distinguish between different objects.

The algorithm proposed by Földiák (1991) and O'Reilly (1992) collapses across irrelevant dimensions by capitalizing on the idea that the visual environment naturally presents a sequence of images of the same object undergoing a transformation along one or more of these dimensions. Thus, one will see a car as a sequence of images translated in space over a short interval of time as it moves past. The information contained in this translating sequence of images is that all of them correspond to the same object in the world, by virtue of the fact that objects don't just appear and disappear instantaneously. One could refer to this information as "identity from endurance", by analogy to other visual regularity extraction, such as "shape from shading", etc. In computational terms, the world imposes a temporal smoothness constraint on the existence of objects which can be used to regularize the ill-posed problem of visual object recognition (c.f. Poggio et al., 1985; Yuille, 1990).

The temporal smoothness of the environment can be capitalized upon by a smoothness constraint (i.e., hysteresis, or the impact of previous activation states on subsequent ones) in the activation state of units in an artificial neural network, combined with an associative learning rule which causes temporally contiguous patterns of input activity to become represented by the same subset of higher-level units. These higher-level units develop representations which are invariant over the differences between the contiguous patterns. Thus, in the example of the car given above, the higher-level units would respond to the visual features of the car in any position in which they were seen.

Further, the implementation of the algorithm proposed by O'Reilly (1992) relies on the idea that neither whole objects nor individual features are subjected to the invariance transformation, but instead a range of *conjunctive features* are transformed, forming an increasingly rich and stable language in which individual objects can be represented. This language is not considered to be purely compositional, in the way that Biederman's *Geons* are (Biederman, 1987). Instead, it is contextual, so that a representation will capture both the identity of a feature and those of its neighbors, enabling spatial information to be eliminated while retaining relational or configural information.

A neural network model of visual recognition that employs a similar scheme of translation invariance is Mozer's BLIRNET model, which performs spatially invariant recognition of words (Mozer, 1991). The BLIRNET representational scheme is effective because the visual representations that are translation-invariant nevertheless encode local spatial relationships due to their conjunctive nature. After several layers of increasingly conjunctive and invariant representations, the highest layer in BLIRNET contains representations for letter-triples in any spatial location. The subset of such letter triples activated by any given word is (for a vast majority of the cases) unique, enabling one to recognize a word from this set of features alone. A similar, but vastly more complex, representational scheme could be employed for general object recognition. Indeed, the goal of the algorithm developed by O'Reilly (1992) is to enable an initially random neural network to self-organize representations which have the conjunctive qualities of the BLIRNET scheme, allowing it to be extended to novel stimulus environments.

The Algorithm

The specific biologically plausible implementation of the general ideas described above, as proposed by O'Reilly (1992), and used in the present simulations is as follows:

- Hysteresis in the activation states comes from the combined forces of lateral inhibition (which prevents other units from becoming active), and recurrent, excitatory activation loops, which cause whatever units are active initially to remain active through mutual excitation. These two forms of neural interaction have been reported in IMHV, as previously discussed (see figure 3).
- Associative learning is implemented with a simple Hebbian correlational learning rule. The specific learning rule used in the present simulations was a modification of the Competitive Learning scheme (Rumelhart & Zipser, 1986) (see Appendix A for the specific formulation), although other similar rules have been used with equal success (c.f. Földiák, 1991; O'Reilly, 1992). The general properties of a learning rule that are important for this model (and most other self-organizing models) are that it have both a positive and a negative associative character (i.e., a covariance formulation as in Sejnowski, 1977), which work together to shape the receptive fields of units both towards those inputs which excite them, and away from those that don't.
- The temporal contiguity of the visual environment is captured by "visual" stimuli that appeared in a series of different locations sequentially over time, all of which share the same features, and differ only in retinotopic location. Thus, only translational invariance is simulated, as this is the simplest form of invariance to model and understand.
- The need for recurrent excitatory activation loops and lateral inhibition requires an interactive network in the tradition of the *Interactive-Activation and Competition* (IAC) models (McClelland, 1981). The activation function used is exactly IAC with stepsize .05, decay 1, and a range of -1 to 1 with a provision that only positive activations are propagated to other units (see Appendix B for the exact equation used).

The Network Model and Methods

The detailed architecture of the model (shown in figure 4) is designed around the anatomical connectivity of IMHV and its primary input area, HA. The *input* layer of the network, layer 0, is considered to represent area HA, which contains cells with properties similar to the simple and complex retinotopic feature detectors described by Hubel and Wiesel in the cat visual cortex (Hubel & Wiesel, 1962; Horn, 1985). HA then projects to area IMHV, which we have divided conceptually into two different *layers*[2] according to the two types of PNs in IMHV described by Tömböl et al. (1988). In the model, the first "IMHV" layer represents the IMHV type 2 PNs, which are likely to receive afferent input. Thus, layer 1 of the model receives inputs from layer 0 of the model. Layer 2 of the model represents the other type of IMHV PN, type 1, which are likely to produce efferent output from IMHV due to their long, bifurcating axons. Layer 2 sends outputs to layer 1 of the model, which in turn sends outputs to layer 2, creating the recurrent feedback loop. There were 9x8 or 72 units in layer 0, and 24 units in layers 1 and 2.

Within each layer of the model, strong lateral inhibition exists in the form of relatively large negative weights between all units in the layer. While they are not explicitly simulated, this reflects the influence of a large number of GABAergic inhibitory interneurons in IMHV (Tömböl et al., 1988), and its relatively low levels of spontaneous activity. The strong inhibition was implemented in the model with weight values of 3.0. With this level of inhibition, only one unit in each layer became active at any time, which makes analysis and understanding of the system much clearer. In the real system, it is assumed that inhibition does restrict activity to a relatively low level, but clearly the *Winner Take All* extreme in the model is not realistic. The generalizability of the results to distributed patterns of activity throughout the system is an issue that will be addressed with the more detailed model. Preliminary results indicate that the same basic effects are found.

Between the two principal cell types in IMHV, excitatory connections exist. In the model, these excitatory connections were of the same strength as the excitatory connections from layer 0 to layer 1, and they were subject to the same learning rules. It is possible that these excitatory connections within IMHV would be stronger in the chick, but since no neurophysiological data is available on this issue, the effect of this variable was not manipulated. Both the inhibitory and excitatory connections provide the recurrent excitation which leads to the hysteresis effect necessary for the learning algorithm. However, only the strength of the excitatory connections was adjusted by the learning rule. This is in accordance with biochemical and neuroanatomical evidence consistent with excitatory (axospinous) connections being the site of plasticity within IMHV, and the involvement of NMDA receptor sites in imprinting (Horn et al., 1985; McCabe & Horn, 1988).

While the units in layer 1 of the model received excitatory input from both layers 0 and 2, the units in layer 2 did not receive any excitatory input from higher layers that are known to exist in the chick but are not present in the model, resulting in lower levels of activation in this layer compared to layer 1. To compensate for this, layer 2 had a lower level of decay than the other layers (.5 vs. 1). It is likely that the principal neurons (type 1 PN) in IMHV receive reciprocal excitatory connections from the areas that they project to, as this is the predominant form of connectivity in the mammalian cortex and in other parts of the avian brain (see Horn, 1985; Douglas & Martin, 1990 for reviews), which would produce an effect similar to the decay manipulation in the model. Indeed, most of the regions which receive

[2]Note that the laminar distinction in the model between these two component cells of IMHV is not intended to suggest that the cells are arranged as such in the IMHV itself, but rather serves to reflect the functional distinction between the two types of principal neurons

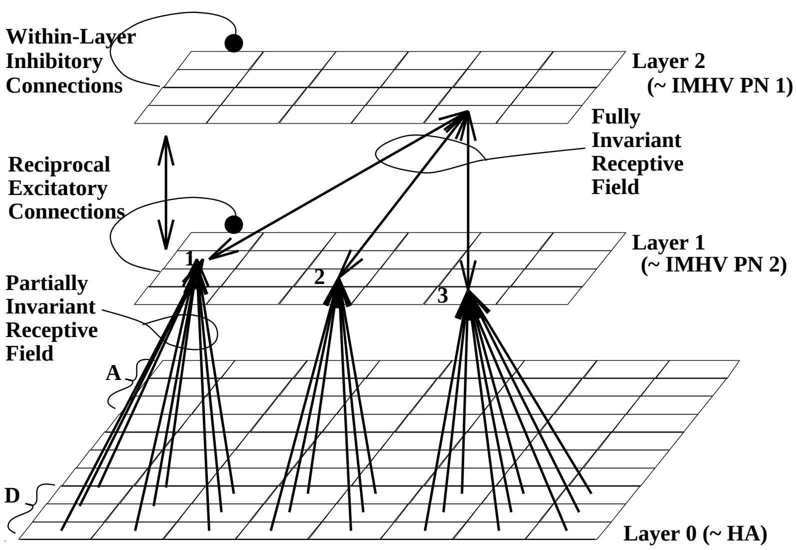

Figure 4: Network architecture used for the simplified model of IMHV, showing the three layers (Layer 0 represents HA, Layer 1 represents IMHV PN type 2, Layer 2 represents IMHV PN type 1), and their interconnectivity. The network is shown being trained on stimulus D, and the 3 different units numbered 1-3 in layer 1 have partially invariant fields which capture local invariance over portions of the different positions of stimulus D. These different units project reciprocally to the layer 2 unit which has a fully invariant representation of stimulus D by combining the partially invariant fields from layer 1.

input from IMHV also have projections back to it (Bradley et al., 1985).

In the model, the decreased decay produces a greater degree of hysteresis in layer 2 than in layer 1, which is reflected in the kinds of receptive fields that form in the two layers. While the different receptive field types are not the focus of the present studies, we have assumed that the input layer should have less invariant fields than the output layer simply because the input receives less stable inputs from the simulated "HA" (the patterns of activity on HA are moving around due to motion in the environment). See O'Reilly (1992) for more discussion of the graded invariance transformation over increasingly deeper layers.

Training in the model consisted of presenting a set of feature bits (assumed to correspond to a given object) in sequential positions across the input layer 0 (simulated HA) with either right or left motion (randomly chosen). At each position of a stimulus, the weights between all units in the system were adjusted according to the Hebbian learning rule once the activation state of the network had reached equilibrium for each position (defined here as the point at which the maximum change in activation went below a threshold of .0005). The activation state was initialized to zero between different objects, but not between positions of a single object, enabling a state resulting from an object in one position to exert the desired hysteresis effect on the next state with the object in the next position. The hysteresis caused units in the model that were active for a given position of a simulated object to remain active for subsequent positions. In combination with the Hebbian associative learning rule, this resulted in the development of units which would respond to a particular set of features

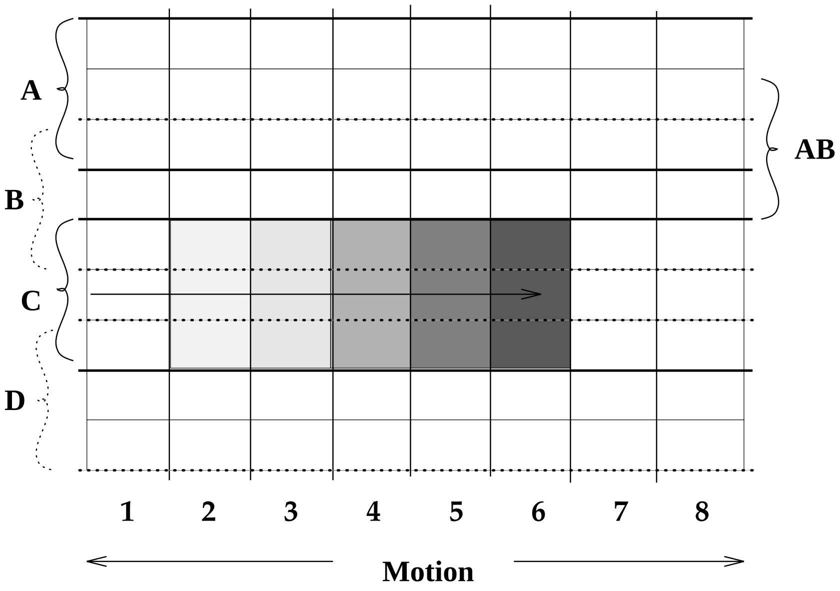

Figure 5: Stimuli used in training the network, consisting of 3 feature bits active in any of 8 different positions. For visual clarity, the positions were arranged along the horizontal axis, and the object features along the vertical, with 1 bit of overlap between each of the 4 primary stimuli. Object AB was a hybrid stimulus used in some simulations that had 2 bits in common (overlap) with both A and B, while the others had 1 bit in common with their neighbor(s).

in any of a certain range of different locations, depending on how long that unit was able to remain active.

Four "objects" were used in the simulations, which consisted of three active features in any of the 8 possible retinotopic locations represented in the input layer. While these locations were mapped horizontally in columns in the simulation, this is really just an approximation for 8 different views of an object, which could correspond to rotation, dilation, or arbitrary translation, and not just horizontal translation. There was 1 bit out of the three features in common with the neighboring stimuli, so that object A overlapped with B by 1 bit, and B with C, C with D. A did not overlap with C or D. See figure 5 for the stimuli used.

Simulation 1: Computational and Capacity Issues

This first simulation was designed to establish the capacity and computational properties of the network as specified above. This was done by training on the entire corpus of four training stimuli (objects A-D), and showing that the model developed individuated representations for each object that were invariant across the entire range of positions of the object. Each epoch of training consisted of sweeping each stimulus in turn across the input layer, covering all 8 positions of each object. The order in which the stimuli were presented was randomized, with a "delay" (implemented by zeroing the activation states) between each sweep of the stimuli. Training continued for 100 epochs. The receptive fields for the two simulated IMHV layers developed invariant representations in a graded way. That is, the first layer developed representations specific to one stimulus in any of several (but not all) different positions of that stimulus. This layer was partially invariant with respect to position, because the receptive fields of these units did not include the entire range of positions of the stimuli. Layer 2, however, did have fully invariant representations, so that a single pattern of activation coded for a single stimulus in any position.

The development of invariance was quite robust over different parameter settings, although the specific qualities of the representations depended on having them in a certain range. In particular, larger levels of activation decay tended to reduce the hysteresis effect and caused lower levels of invariance to develop. Likewise, smaller levels of decay caused more invariance to develop on layer 1 (layer 2 was already fully invariant). These effects appeared over relatively large changes in decay (1 vs. 1.5, for example). The specific level of hysteresis (controlled by activation decay, primarily) used in the simulations was chosen because it produced a distinction between the layer 1 and 2 receptive fields, instead of making them both fully invariant. This kind of graded invariance is thought to be important for more complex kinds of invariance processing (O'Reilly, 1992).

We can conclude from this simulation that, when all objects are encountered in an interleaved manner within each epoch, the network will form fully invariant representations of each one. Thus, any sensitive period effects found in subsequent simulations are not due to a limitation of the network *per se* (e.g., that it is too small to learn multiple objects), but rather must be due to the manipulation being performed.

Simulation 2: Basic Imprinting (Learning)

Behavior

Imprinting is measured behaviorally by chicks' preferential approach to the training object. As was described above, the training consists of the presentation of a stimulus in front of the chick for a period of time. The stimulus must be moving in order to obtain strong imprinting (Sluckin, 1972). For example, Hoffman & Ratner (1973) presented naive chicks with an identical stimulus either moving or static. While chicks would readily imprint to the the stimulus when it was moving, they did not when it was static. This effect is unlikely to be simply due to the differential attraction of attention, since once chicks had been trained on the moving object, they would readily approach it even when it was static. That is, the movement was essential for training, but not for subsequent recognition. In the model, the input stimulus must be moving because this enables an invariant representation to develop, and we argue that this should also hold for IMHV and the chick.

Simulation

The simulation of the imprinting effect involved presenting a single stimulus for a long period of time, and recording the development of the preference for this stimulus over time. Since it is not possible to record preferential approach behavior from the network, some proxy for this kind of preference measure must be used. We have hypothesized that layer 2 in the model represents the output of IMHV, so we can record the activation level over this layer when different stimuli are presented to determine the preference for a presented stimulus.

The most obvious way of measuring activation level is to simply record the activation strength of the one active unit in Layer 2. However, there is a complication with the activation measure due to the lateral inhibition within a layer. If several units have become responsive to the imprinted stimulus, then they will compete with each other more strongly than if fewer units respond to the stimulus. Thus, the activation strength for the imprinted stimulus might actually be weaker than for the untrained stimulus. As discussed above, the constraint of lateral inhibition in the real system is not thought to be strong enough to eliminate all but a single active neuron. Thus, this complication would not be present, at least to the degree it is in the simplified model. To circumvent this potential measurement artifact, we instead used the total excitatory input to layer 2 (i.e., for all units in the layer) as an indication of preference. This measure eliminates the confound from the inhibitory input.

The final preference scores were computed by first summing the excitatory input to each unit over all units in layer 2 to arrive at a total excitatory input. This total was then averaged over all the different positions for each stimulus, resulting in a raw preference score for each stimulus. These raw preference scores were turned into percentage preference measures between two stimuli (e.g., A over D) by dividing each raw score by the total for both stimuli to get a percent preference score. This is analogous to the *preference score* measure used in many behavioral studies (Horn, 1985).

The basic imprinting results for the model are shown in figure 6, which indicates that a preference for the training stimulus over a novel, dissimilar, object does develop over time. Like the chick, the model shows increased preference for the training object with longer periods of training. In the model, this effect is simply due to the selective enhancement of weights to units which respond to the imprinted stimulus, and is not a surprising or novel result. In combination with the previous simulation, this result indicates that the simplified model is capable of developing translation invariant object-based representations that are sufficient to account for an imprinting-like preference for a stimulus to which the system has been exposed. We will next show that the system exhibits other characteristics of the imprinting phenomenon as well.

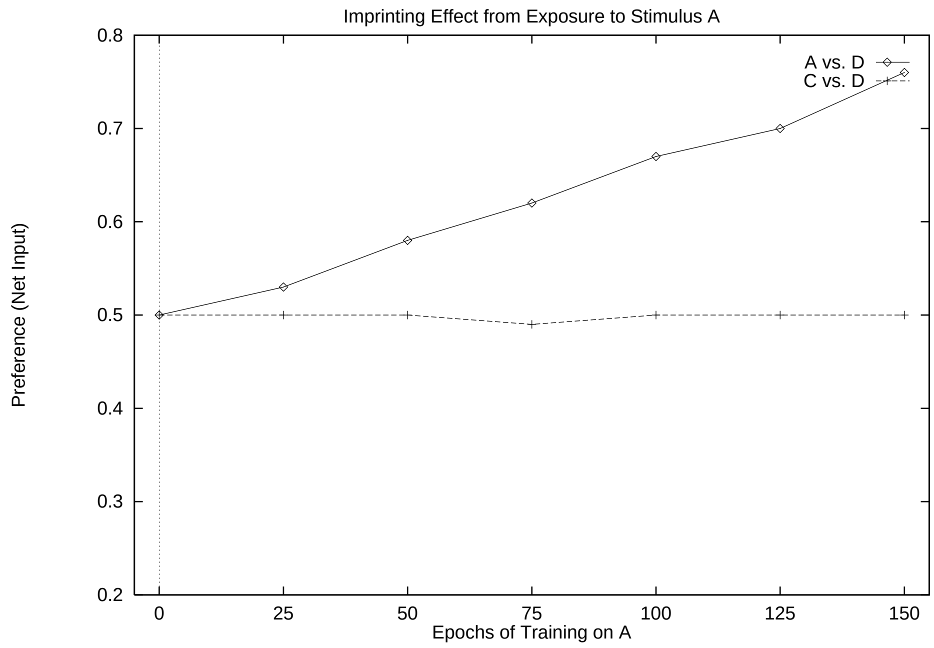

Figure 6: Simulation 2: The basic imprinting effect, showing the preference for the imprinted stimulus A as compared to a novel stimulus, D. The preference for a control stimulus C as compared to D is also shown. This preference does not deviate from chance (0.5).

Simulation 3: Reversibility and the Sensitive Period

Behavior

Lorenz (1937) originally claimed that an imprinted preference was irreversible. Jaynes (1956) pointed that there are two senses in which imprinting could be irreversible. First, that after imprinting a bird will never again direct its filial responses to a novel object. Alternatively, that while a bird can direct its filial responses to a second object, it always retains information about the first object. There is a considerable amount of evidence that imprinting is not irreversible in the first sense (e.g., Klopfer & Hailman, 1964b; Klopfer & Hailman, 1964a; Klopfer, 1967; Salzen & Meyer, 1968; Kertzman & Demarest, 1982). Much evidence supports the second, and weaker, form of the irreversibility claim (for review see Bolhuis & Bateson, 1990; Bolhuis, 1991). Thus, while imprinted preferences can be reversed by prolonged exposure to a second object, a representation of the original object remains.

A number of factors affect whether imprinted preferences can be reversed, and if so, the extent of reversal. These factors include the length of exposure to the first object experienced by the chick, and length of subsequent exposure to a second object. For example, a prolonged exposure to the first object will prevent reversal of preference to a second object (Shapiro & Thurston, 1978). With shorter exposure to the first object, a very brief period of exposure to a second object will not result in a reversal of preference, while a longer period of exposure

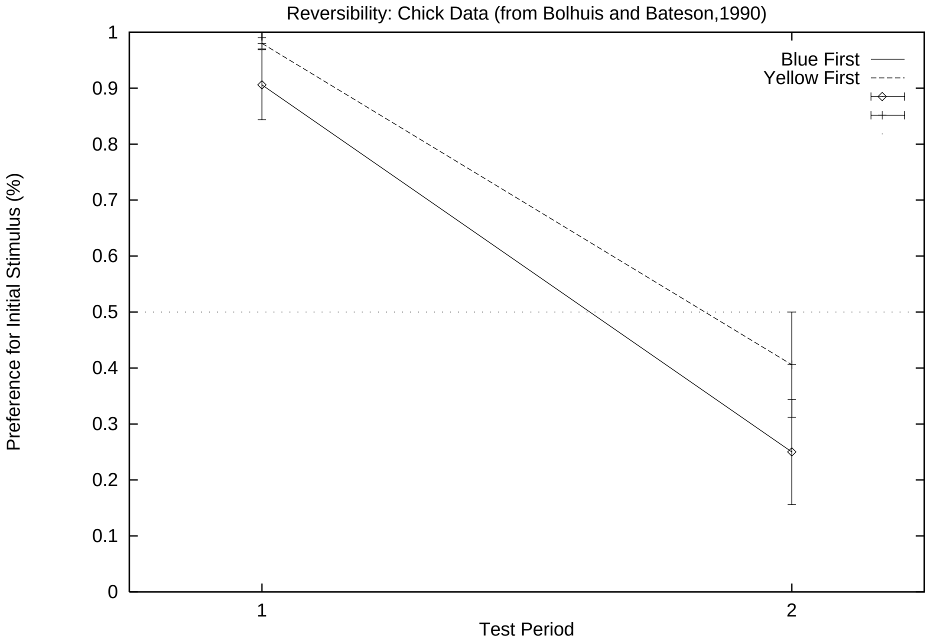

Figure 7: Data from Bolhuis and Bateson (1990) showing an initial preference (mean and standard error) for the first object (either a blue or yellow cylinder, as indicated) recorded on test 1, and a reversal of this preference after a period of exposure to the other stimulus, as recorded in test 2.

to a second object will (Salzen & Meyer, 1968).

Some of these phenomena may be illustrated by data gathered by Bolhuis & Bateson (1990). Like other experiments of this kind, these authors exposed chicks to an object for a certain period of time, in this case either 3 or 6 days. Following this, the chicks' preferences were tested, and a strong preference for the training object was found. After this, the chicks were exposed to a second object for a period, again 3 or 6 days in this experiment. When preferences are measured at this point, there is often a reversal of preference toward the second training object. Figure 7 shows the basic reversal of preference effect for a similar amount of exposure to both objects. The data also indicate a baseline preference for different colored objects (the blue object being preferred to the yellow, in this study).

Figure 8 shows the preferences obtained in various combinations of 3 and 6 day exposure durations to the same type of objects, demonstrating that the extent to which a reversal occurs depends on the length of exposure to the first stimulus (longer exposure reduces the subsequent extent of reversibility). With longer periods of exposure to the first object, the extent of training on the second object has little effect. While Bolhuis & Bateson (1990) did not systematically manipulate the length of exposure to the second stimulus, other studies have shown that this variable influences the degree of reversal (Salzen & Meyer, 1968; Einsiedel, 1975).

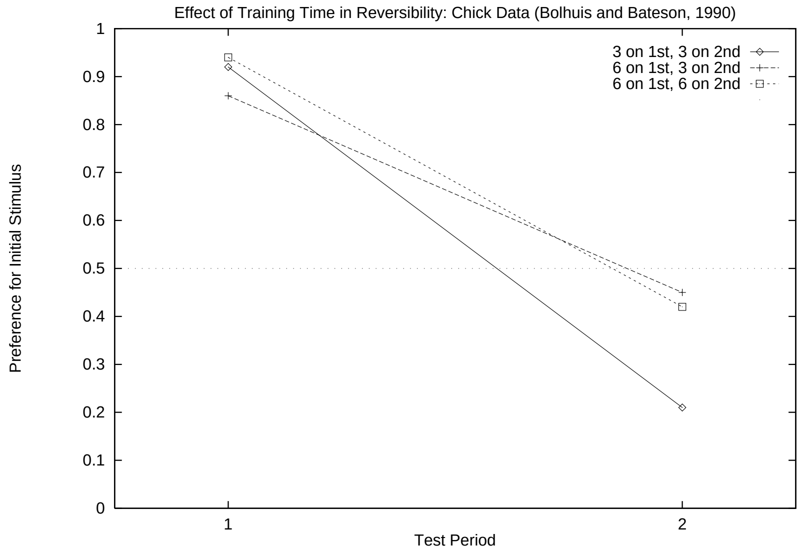

Figure 8: Data showing an initial preference for the first object after either 3 or 6 days recorded on test 1, and a reversal of this preference after a period of exposure of 3 or 6 days to the other stimulus, as recorded in test 2. This shows a decreased level of reversal with more exposure to the first object, even with an increased amount of exposure to the second object

Simulation

The implementation of reversibility in the model was straightforward: train on one stimulus for a variable amount of time, then train on a different one for a variable amount of time. For the basic effect, we trained on stimulus A for 100 epochs (100 sweeps across the input layer), and then trained again with stimulus D from various points of training on A. The results are shown in figure 9 for a network which was exposed to A for 100 Epochs, and then exposed to D for up to 300 epochs. Thus, reversibility occurs despite a relatively strong initial preference to A, with D being preferred to A after 150 epochs of exposure to D. This finding is consistent with those cited above which show that a chick can reverse its initial preference for one object after exposure to a second imprinting object. Further, the strength of preference for D increases with longer training on that object. Also shown in this figure is the continued preference for A, the initial training object, over a novel stimulus, B. This finding shows that a representation of the first object still exists, and is consistent with the second, weaker form of irreversibility observed in the chick and discussed above.

That the network displayed reversibility is not terribly surprising, as many neural networks will continue to adapt their weights as the environment changes. However, unlike many networks which display a "catastrophic" level of interference from subsequent learning (c.f. McCloskey & Cohen, 1989), this model retained a relatively intact preference for the initial stimulus over a completely novel one. The explanation for this preserved learning effect will be presented below.

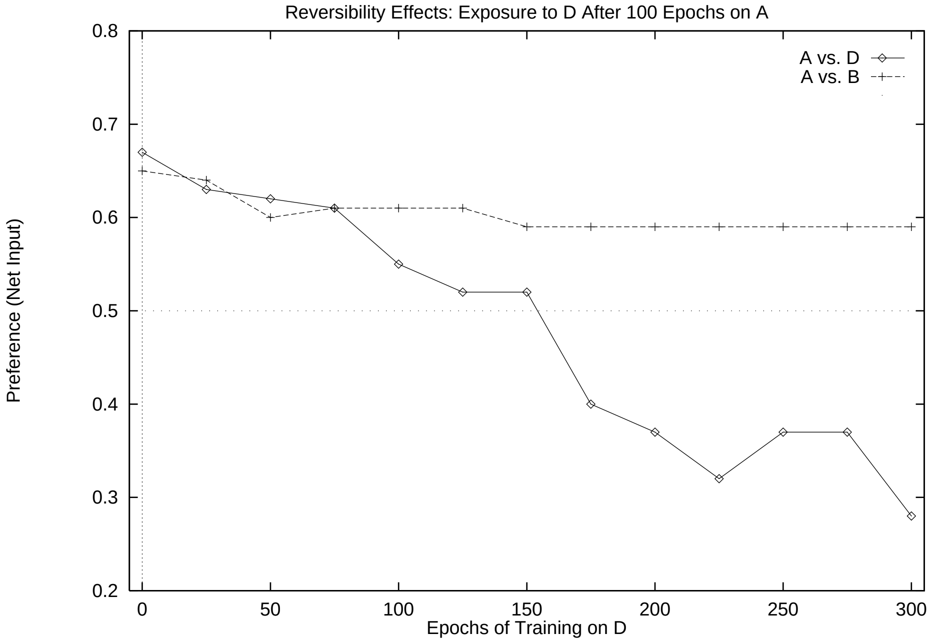

Figure 9: Simulation 3: The basic reversibility effect from training on D after initial imprinting on A. The comparison of the preference for A vs. D shows a reversal (from above 50% preference to below 50% preference). The preference for A over a second object B, remains stable despite the training on D.

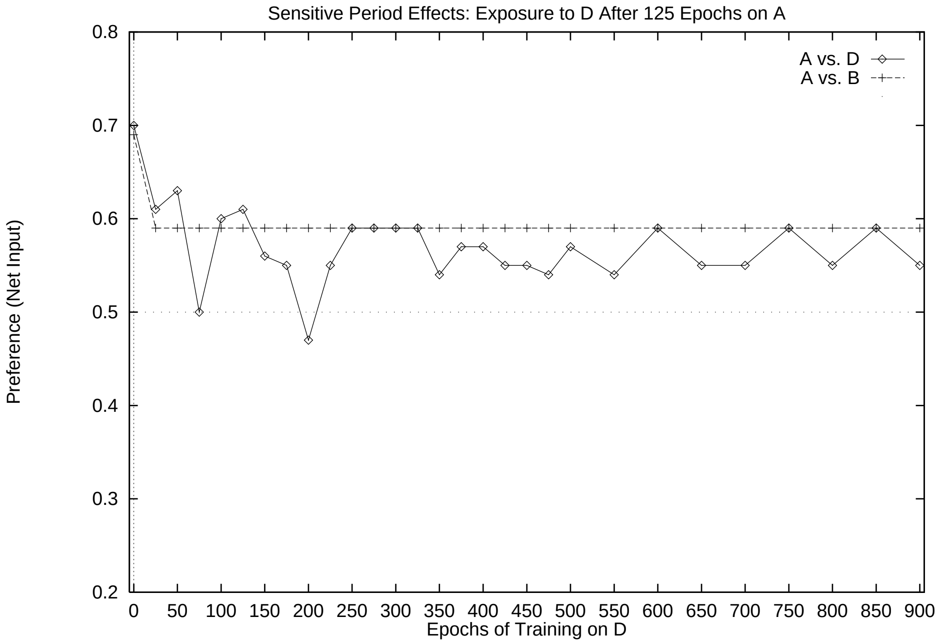

Figure 10: Simulation 3: A sensitive period effect from 125 epochs of imprinting on A before exposing to D. The A vs. D preference does not reverse, indicating a preserved preference for A despite up to 900 epochs of exposure to D

While reversibility seems to occur in the network and in chicks, there is also support for the idea of a *self-terminating* sensitive period where sufficiently long exposure to an imprinting stimulus will prevent the preference from being reversed (e.g., Shapiro & Thurston, 1978). Indeed, in the model, a sensitive period effect comes with just 25 more epochs of exposure to A before exposing the network to stimulus D. This additional exposure prevents any amount of further training on D to reverse the initial preference for A. This can be seen in figure 10. This kind of sensitive period is self-terminating because it is determined solely by exposure to the object, and not by some other maturational change in the system.

Discussion

Both the sensitive period effect and the preserved preference effect described above can be explained by an interaction between the subtractive component of the covariance learning rule and hysteresis effects. Consider a unit in layer 1 of the model network which has become active by the presentation of a stimulus in position 1 of the input layer. This unit will increment its weights to the 3 feature units that activated it, and to the unit in layer 2 that is also active. Also, according to the subtractive component of the covariance learning rule, it will also decrease its weights to all the other inactive units in the network, *including those in the other positions of the same stimulus.* When, due to hysteresis, the same layer 1 unit remains active for position 2 of the stimulus, the same subtraction will occur on weights from those input units in position 1 that were just increased on the previous time step, as they will now be inactive. Indeed, if this unit remains active for more than 2 different positions of an object, the net result will be to decrease its weights to each of these positions more than increase them.

With the weight bounding procedure used in the network, the value of a weight represents the balance of increasing and decreasing forces. Thus, the weights to a layer 1 unit representing an object over several different positions will reflect the conditional probability that the input unit is active given that the layer 1 unit is (Rumelhart & Zipser, 1986). Thus, if a layer 1 unit is active for N positions of an object, the weight from each position will equilibrate around a value of $1/N$. This is implemented by modulating weight decrease by the current value of the weight, weight increase by 1 minus the weight (see Appendix A).

Because the weights are initialized to a random value with a mean of .5, they will typically be decreasing initially whenever $N > 2$. This means that a unit that was active for a given object on one epoch will be *less* likely to be active the next epoch. As a result, some other unit will then come on, and adjust its weights for the same object. In this way, the repeated presentation of a given object will result in a *recruitment* process where multiple units will become tuned to the same object. It should be noted that this process happens gradually, with all of the recruited units taking turns becoming active, even when the system has reached equilibrium (i.e., all weights near their $1/N$ values). There are possibly other mechanisms of recruitment that could take place in neural systems based on fatigue of active neurons, graded activation of many neurons, etc. Our general claim is that recruitment of some form is important for the sensitive period phenomenon, and further that this particular form of recruitment is a plausible one.

As layer 1 units become recruited, they are always decreasing their weights to input units with which they are never contemporaneously active. This makes them not respond to the features corresponding to objects other than the one they have been exposed to, and is the reason for having a subtractive component to the weight update rule. We will refer to this as a *tuning* process. In the presence of multiple objects, the influence of recruitment is balanced by this tuning process, because units tuned to a given object will be unavailable for recruitment for another object. As was seen in Simulation 1, the network divides the representational space of layer 1 evenly among the different objects in this case.

However, when only one object is viewed for a period of time, recruitment and tuning work together to cause many units to respond very selectively to the imprinting stimulus. Thus, as training continues on the imprinting stimulus, both the tuning and recruitment effects get stronger, so that by a certain point (125 epochs in the present simulations), a majority of the units become selective to the imprinting stimulus, and are unavailable for recruitment to any new training stimulus because their weights to the other object features are near zero. Once this majority has been established, no amount of exposure to a different stimulus will cause these units to be recruited to the new stimulus, and the balance of preference will not shift. Instead, the retraining will cause the minority of initially unrecruited or less strongly recruited units to become selective to the new stimulus.

The tuning and recruitment processes also explain the retention of preference for a trained stimulus over a novel one even after re-training. The recruited units, being tuned to a particular object, do not become recruited by the other object. Thus, the original preference for the object is preserved.

It is important to determine how general the explanation of sensitive period and preference retention effects in the model is. The two critical features of our account are the existence of hysteresis in the activation state of the units, and the existence of a subtractive component to the learning rule. The hysteresis is a crucial component of the object-recognition algorithm, and is supported by presence of recurrent excitatory loops and lateral inhibition present in the circuitry of IMHV. The second of these features, however, is a matter of some debate in the field of neural synaptic modification. While an extensive overview of this subject is beyond the scope of the present paper, it suffices to say that there is considerable empirical evidence for a subtractive synaptic modification phenomenon known as *Long Term Depression* (LTD) in several different types of neurons and species (e.g., Stanton & Sejnowski, 1989; Artola et al., 1990; Bradler & Barrionuevo, 1990; Frégnac et al., 1988).

While the two components of an LTD effect and hysteresis account for the existence of both a recruitment process and a tuning process, the specific properties of both of these components will affect the quantitative nature of the reversibility phenomenon in the chick. In particular, the level of recruitment that takes place may interact with the ability to prevent reversibility, since prevention of reversibility in the model requires the initial stimulus to have recruited a majority of the available units. To demonstrate the influence of the recruitment process, simulations were run which were identical to the previous ones in all respects except for the size of layers 1 and 2, which were doubled. In these simulations, the initial preference for stimulus A was reversible after 150 epochs of exposure to A, but not after 300 epochs of exposure to A. Thus, though it took longer to develop the bigger network to the point that it could not reverse its preference for the initial stimulus, this point did come. Examination of the development of the receptive fields for units in layer 1 of this network showed that it was not the number of units which were tuned to stimulus A that changed over time, but rather the contrast between the weights from the imprinted object and the weights from the other objects. This observation in concert with the longer training time required to prevent reversibility in the bigger network supports the gradualistic model of recruitment described above.

One variable that could be influential in the behavior of the network is the magnitude of the initial random weights. However, control simulations have shown that the initial weight magnitude is relatively unimportant unless it is lower than .125, which corresponds to a $1/N$ where N is 8, which is the level a weight will tend towards for a unit which is active for all 8 possible positions of a given object. In this case, the weights to an initially active unit will increase, and this is the only unit that will respond to the object—no recruitment will take place. The reversibility results for a simulation run with initial weights centered around .2 ($\pm$.1) show that the initial weight value does not alter the sensitive period effect. In this simulation, reversibility was tested after 150 epochs of training on A, and it was not possible to reverse the initial preference for A even after 900 epochs of exposure to D.

Both of the simulations described above indicate that the recruitment process is robust, and should be thought of as a gradual effect. The robust nature of this effect means that it could be found behaviorally with many different levels of recruitment shown in the neural firing patterns of IMHV cells. Thus, just knowing the behavior and having the simplified simulation will not enable us to make definite predictions about the absolute number of recruited cells in IMHV, but it would predict that more cells sensitive to the initially imprinted stimulus than to the second stimulus should be found in post-sensitive-period chicks, and will explain why this might be so.

So, while parameter exploration in the simplified model is helpful in establishing the validity of the explanation provided, and in understanding the basis of the findings from this model, a more satisfying quantitative picture of IMHV awaits a detailed point-neuron model. Such a model can incorporate quantitative constraints from the anatomy and physiology of IMHV (e.g., the relative strength of different synapses, and the impact of learning on these strengths), and hopefully provide more quantitative predictions about chick imprinting behavior based on these constraints than can the present model. However, the simplified model does provide insight into the kinds of mechanisms that are at work in the dynamics of the imprinting process, and the probable behavioral manifestations of these factors, and also yields a number of qualitative predictions.

Predictions

- At the behavioral level, we would predict that the length of time training on the first stimulus will affect the amount of re-training time necessary to reverse the preference for this first stimulus in favor of the second. As we have seen, support for this claim exists (e.g., Bolhuis & Bateson, 1990), but a more systematic study of the relative amount of time needed to reverse would be necessary to further constrain modeling efforts.

- We would predict that a long period of training on the initial stimulus will prevent any reversal effect no matter how long the second stimulus is presented. Given the uncertainty about what constitutes an *epoch* of training in the real system, and the effects of scaling up our model on the amount of training necessary to prevent reversibility, we are unable to make any quantitative predictions about the actual length of exposure required to prevent reversal of preference in the chick.

- At the level of firing properties of neurons in IMHV, we would predict that a cell that shows a preference for a given stimulus will retain this preference even after re-training, and further that this propensity for retention of initial preference will be correlated with the selectivity or strength of the initial preference.

- More neurons in IMHV will show a strong preference for the first training stimulus after the sensitive period for reversibility than before it.

Simulation 4: Generalization

Generalization Behavior

A number of studies have indicated that generalization from the training stimulus occurs following imprinting on an object. For example, chicks which have been exposed to an object still followed objects differing in color or shape to some extent (Jaynes 1956; 1958), and Cofoid & Honig (1961) found some evidence for generalization of the following response to different colors. More recently, Bolhuis & Horn (1992) trained groups of chicks on red-colored stimuli, before testing their preferences between two blue-colored stimuli, one of which was of the same shape as the training stimulus, and the other of which was not. Chicks showed a strong preference for the object of the same shape. Further, the original red training object was preferred over the blue object of the same shape. Thus, these authors established evidence for a gradient of generalization from the training object.

Simulations

Generalization is thought to occur in the neural system because of the tendency of overlapping input patterns (i.e., similar stimuli) to activate the same units in higher layers according to the degree of overlap present in the input layer. This is a straightforward property of most neural network models which results from the use of graded activations based on the strength of the input to a unit; a unit having 2/3 of the same input weights active in one pattern as another will have a correspondingly similar activation level, depending on the non-linearities present in the activation function.

To confirm that generalization effects similar to those found in the behavioral literature could be accounted for by our model, we ran several generalization tests on the simulations described above. Generalization in the model was tested by presenting patterns having varying degrees of overlap (similarity) with the imprinting stimuli, and recording the preference measure for these test patterns. Given that each stimulus consisted of 3 active features, we were able to vary overlap from 2/3, 1/3 and no overlap.

In the imprinting simulation (simulation 2), generalization was tested by conducting preference tests for stimulus AB (a hybrid of A and B, having 2/3 overlap with each), and stimulus B, which had 1/3 overlap with A. Each of these stimuli was compared to a novel stimulus, D. As can be seen in figure 11, imprinting on A generalized to AB, but not to B. Thus, there is a non-linearity in the generalization effect, which is due to the lateral inhibition within the network layers. Inhibition causes weakly activated units to become inhibited by more active units. In this case, the 1/3 overlap from stimulus B was not sufficient to enable the imprinted units to become active in the face of competition from other units. Thus, the imprinting training had no effect on the response to B because the units which became active in the presence of B had not been trained. However, the 2/3 overlap in AB was sufficient to enable the previously imprinted units to become active, thus causing the enhanced preference. The preference for AB was less than that for A because, although the imprinted units were activated, they received less input from inputs with enhanced weights than with the A stimulus. Thus, as was found in the chick by Bolhuis & Horn (1992), there is a gradient of generalization based on feature similarity.

Similar results were found when generalization was tested in the reversibility simulation, as can be seen in figure 12. In this case, the initial exposure to A generalized to AB, and not B, as the preferences for A and AB over D were similar initially, but B over D was near chance. Further, the subsequent exposure to D caused the preference to A and AB to reverse. This reversal was less than the preference for D over B, indicating that the non-linear generalization from the initial exposure is preserved following reversal.

The reversibility simulations were also run with AB as the retraining stimulus. In this case, the retraining stimulus is highly similar (2/3 overlap) to the original stimulus, and we know from the above simulations that the network will already generalize from training on A to AB. As can be seen in figure 13, the preference for A relative to a novel stimulus (D) increases, rather than decreases after exposure to AB. Thus, the network treats these two as roughly the same stimulus, and imprinting for one is essentially the same as imprinting for

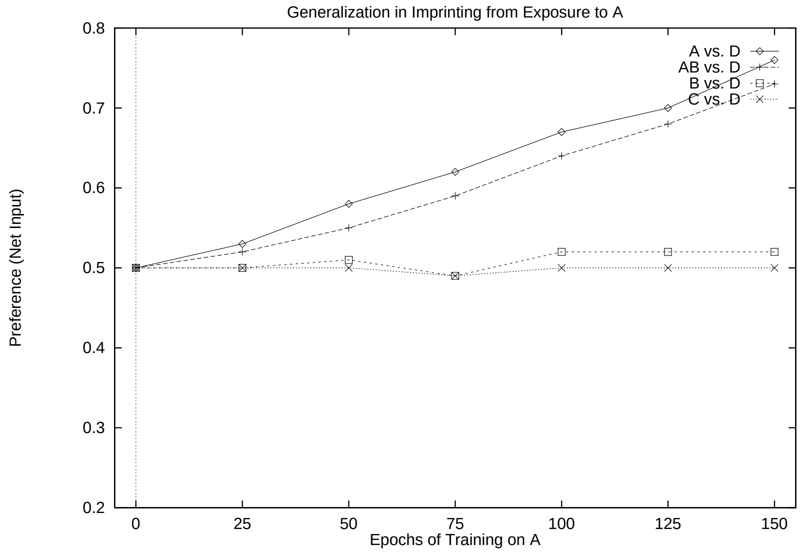

Figure 11: Simulation 4: Generalization of imprinting on stimulus A to stimulus AB (2/3 overlap with A), as revealed by a comparison with the measured preference for D. Note that stimulus B, which has a 1/3 overlap with A, does not experience significant imprinting from exposure to A, revealing a non-linearity in generalization

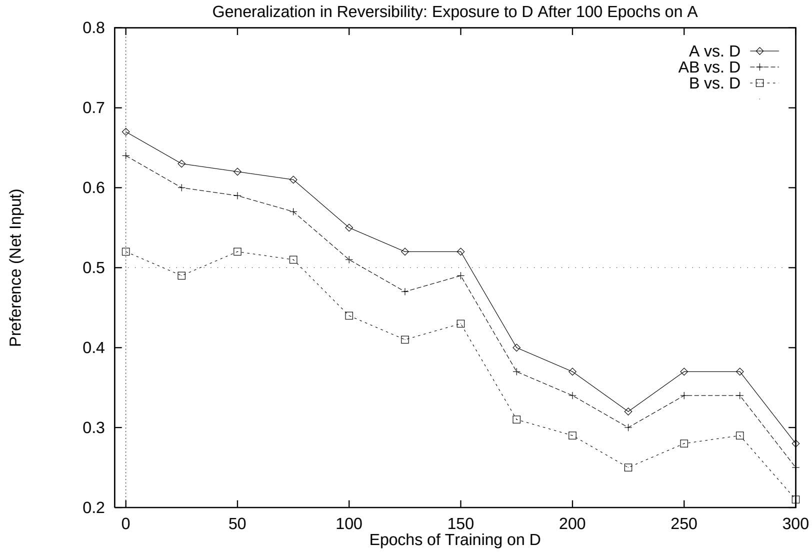

Figure 12: Simulation 4: Generalization of the reversibility effect from exposure to D after 100 epochs of exposure to A. Stimulus AB (2/3 overlap with A) shows both the initial preference, and the reversal of this preference from exposure to D. Note that stimulus B does not show the initial preference and remains less preferred than A or AB, due to non-linear generalization effects

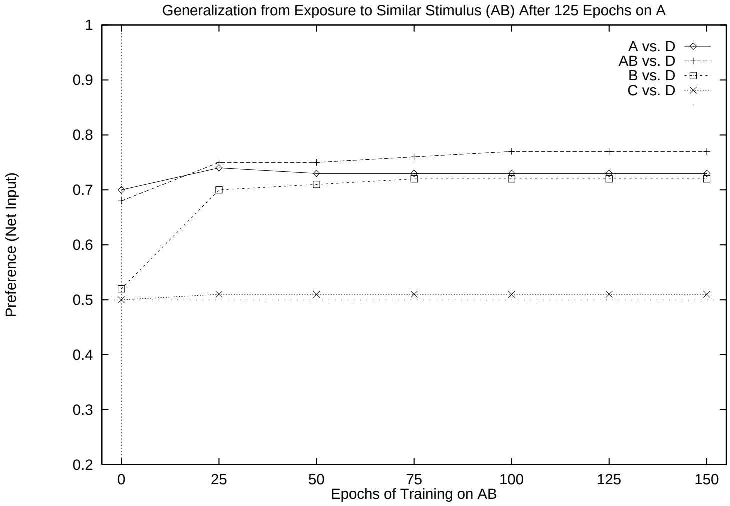

Figure 13: Simulation 4: Training on a highly similar stimulus (AB) to the original imprinting stimulus A increases the preference of A relative to a novel stimulus (D), instead of decreasing as with the case of a dissimilar training stimulus (D). While there is a reversal of preference for A relative to AB, this effect is minor relative to the increase in preference compared to the novel stimulus. Also, the AB stimulus, being equally similar to A and B, causes the system to generalize to B as well, which shows a similar level of preference relative to the novel stimulus as A.

the other. Again, this results directly from the fact that the same units are becoming active in the presence of both stimuli, and their common weights are enhanced through training. However, technically a reversal does take place, since AB does become slightly preferred to A after exposure to AB, even though the preference for A also increases.

The other interesting effect shown in the figure is that the retraining on AB generalizes to stimulus B. This derives from the fact that stimulus B overlaps by 2/3 with AB, which causes generalization to B by the same principle that causes training on A to generalize to AB. Thus, initially quite distinct objects can be made to elicit the same preference from the network, and activate the same units, by providing a "bridge" between them. This network was originally exposed to A, and then the bridge was formed from exposure to AB. It is possible that this ordering of exposure to the objects could be responsible for the bridging effect, and different orderings (e.g., exposing to A and B, then AB) would give different effects.

Several control simulations were run to determine if the order of training was important. The first involved initial training on A and B (in random order, with both swept across the input in each epoch, with a "delay" between each, similar to Simulation 1), and testing for the AB preference. Since we know from Simulation 1 that this kind of training will result in the formation of different sets of units responding to either A or B, this might affect the generalization to AB, which has 2/3 overlap with both A and B. However, despite the enhanced competition from the distinct A and B imprinted units relative to the original network with only A imprinted units, AB was preferred at the same level as the original network. In a similar test, we trained another network initially with A and C (methods as before), and then tested its preference for B, which overlaps 1/3 with each of these stimuli. In this case, we found no evidence of generalization, even though 2/3 of the features belonging to B were imprinted on (1/3 from A, 1/3 from C). Finally, we tested a network where A, AB, and B were all imprinted initially, and found no important differences from the original network.

Another effect of using stimulus AB for retraining is that the self-terminating sensitive period effect found when D was the retraining stimulus does not occur. This is a direct consequence of the fact that many of the same units in the model that were active for A are active for AB, so the recruitment and tuning processes that led to the self-termination of the sensitive period do not have the same effect. Instead, the preference for A will be maintained or even increased relative to novel stimuli because of the further training on the units that mediate the preference for A.

Predictions

The findings on generalization lead to a series of predictions about the behavioral reactions of chicks to stimuli with differing degrees of similarity, and for the neural firing in IMHV that presumably subserves these phenomena.

- By varying stimulus similarity systematically along one or more dimensions, one can determine the width of the *generalization window* in chicks. We predict that the shape of this window will be non-linear with respect to stimulus similarity, with a rather sharp boundary to the range of stimuli that generalize and those that do not. The effect in the model was due to the presence of inhibition within each layer, which is a reported property of the circuitry in IMHV.
- Further, we predict that objects within the generalization window will activate many of the same principal neurons (Tömböl et al., 1988 type 1 or 2) in IMHV, while those not in the window will not. This finding would establish the neural basis of generalization as predicted from our model.
- As a behavioral effect, we predict that secondary training on a stimulus that is very similar to the original imprinting stimulus, but also similar to another stimulus, will result in a "widening" of the imprinting effect, so that all three stimuli will now show a similar level of preference to a novel stimulus. An example of this could involve using color and shape features, as was done in Bolhuis & Horn (1992) with red and blue objects of the same and different shapes. In this case, the initial training stimulus might be a red square, which would cause imprinting on this stimulus and generalization to a blue square, but not to a blue circle. However, retraining on the blue square would cause the bird to generalize to the blue circle as well.

  While this prediction has yet to be systematically studied, some existing evidence is at least consistent with it. Ryan & Lea (1989) exposed chicks to a stimulus composed of four table tennis balls. For some chicks the balls were all white, while for others the balls were all brown. In one group of chicks the stimulus was gradually changed by changing the color of one ball every four days until all of the balls were of the opposite color. In other groups all of the balls were changed in one step. The results of preferences tests indicated that the chicks with gradually changing stimulus had developed an equally strong preference for both colors of the stimulus. The authors argue that these chicks had undergone *category enlargement*, whereas the chicks exposed to the sudden change had not.

- Retraining with a probe stimulus can be used as a better test of what the animal considers to be the same object. With a simple generalization test (typified by the first generalization simulations reported in figure 11), the preference difference between the stimulus which was generalized to (AB) and the one that was not (B) was a matter of around 20 *Preference Percentage Points*, but the pattern of change with more imprinting on A was similar (both increased, with B increasing only slightly). However, in a retraining test where the animal is trained with a probe stimulus after imprinting, the probe that generalizes to the imprinting stimulus shows a completely different pattern of results than the one that does not. Retraining on a stimulus that generalizes can actually increase the preference of the initial imprinting stimulus to a novel object, whereas retraining on a stimulus that does not generalize decreases this preference. Thus, a simple generalization test produces quantitative differences, while the retraining test produces qualitative differences. While it is not an issue with our simple model, these qualitative differences should be easier to detect in noisy empirical data than quantitative ones.

  Further, the retraining effect makes a stronger case for the idea that the same IMHV neurons mediating the preference are active for both the initial and generalized stimuli, since a the same level of preference in the simple generalization test could in theory be caused by different active populations of neurons in IMHV responding with equal vigor. Generalization of retraining, however, requires that learning in the set of neurons activated by the retraining stimulus affect the response to the initial stimulus, meaning that the same set of neurons are active for the two stimuli.

Simulation 5: Temporal Contiguity and Blending

Behavior

In a phenomenon related to generalization, a number of authors have reported that blending can occur when two objects are presented in close temporal or spatial contiguity. For example, Chantrey (1974) varied the inter-stimulus-interval (ISI) between the presentation of two objects from 15 seconds to 30 minutes. By subsequently attempting to train chicks on a visual discrimination task involving the two objects he was able to demonstrate that when the objects had been presented in close temporal proximity (under 30 seconds gap between them) they became *blended* together to some extent. This result was replicated by Stewart et al. (1977) who also controlled for the total amount of exposure to the stimuli in a 15 second ISI condition as compared to a 30 minute ISI condition. Such blending effects may also depend on the physical similarity of the two objects (Stewart et al., 1977, Bateson, pers. comm.), with physically dissimilar objects paradoxically showing greater temporal blending than similar ones.

Simulations

According to our assumptions about the specific nature of the learning taking place in IMHV, temporal contiguity should be a very important variable in determining what the network considers to be the same object. In the "identity from endurance" algorithm, an object is *defined* as that which coheres over time. If this kind of learning is indeed taking place in IMHV, then one would expect to find the kinds of blending effects found by Chantrey (1974) due to close temporal contiguity.

In order to demonstrate that blending effects can result from the spatial invariance learning algorithm employed in the model, a set of simulations were run that manipulated the "delay" between the presentation of two different objects. In actuality, the simulations used the initialization of the activation state as a proxy for a delay long enough to allow the neural activation states to decay to values near resting, or at least retain little information about previous states. The exact length of time that this corresponds to in the chick is not known. Given this situation, we manipulated the delay as a binary variable, running different networks where we either presented the stimuli (one sweep across the input layer) and initialized the activation state between each (delay), or presented them without an initialization between (no delay). Note that all simulations to this point involving multiple stimuli per epoch have been run with a delay.

Several approaches could be taken in order to measure blending. One would be to simply record the preference for each of the objects relative to the same novel object, and consider those that have roughly the same preference level to be *blended*. However, this approach fails to distinguish the possibility that different populations of units are responding to the different stimuli with the same level of preference. As was mentioned earlier, a behavioral measure that circumvents this problem is to retrain on each of the different stimuli, and see if the other ones are enhanced or diminished by this retraining. If they are enhanced, then it is very likely that the same units are responding to the different stimuli, indicating a blending across stimuli.

However, in a simulation, it is possible to look directly at the units and their weights to determine what the units have encoded. If a unit has weights that have been enhanced from features belonging to two or more different stimuli, then this unit has a representation that blends the distinction between these objects. If the unit only has strong weights from features for one stimulus, then the unit has a representation that differentiates between stimuli. Thus, to measure blending, we simply count the number of units that have non-differentiating representations. In order to allow for noise in the weights, we take an arbitrary

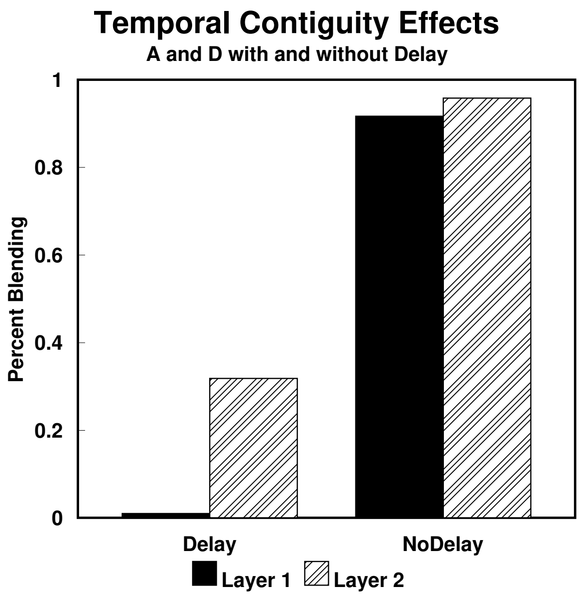

Figure 14: Simulation 5: Temporal contiguity manipulation (delay vs. no delay) affects number of blended representations, so that the no delay condition causes considerably increased blending of the two stimuli. Blending is measured by examining the receptive fields of the units in each of the two simulated IMHV layers.

threshold at .36788 ($1/e$) times the strength of the strongest weight into a unit as a cutoff for considering that weight to be enhanced.

Figure 14 shows the level of blending for training on stimuli A and D with either a delay (zeroing the activations) or no delay between them. Clearly, the absence of a delay causes almost total blending in the representations that form, so that the network would be unable, based on the active units, to distinguish whether stimulus A or stimulus D was present in the environment.

The explanation for this effect derives from the hysteresis in layers 1 and 2. The hysteresis causes units in these layers that become active to remain so, and the Hebbian learning rule will enhance weights to whatever pattern is active while the unit is also active. Thus, units that become activated when one stimulus is presented will tend to remain active for the other stimulus, causing weights to encompass both stimuli. Since blending has been found with short delays in empirical testing with chicks, this lends support to the idea that hysteresis occurs in IMHV, as temporal effects would not be expected without it.

The effect of similarity of stimuli on this blending process may be important, so that more similar objects will be subject to greater blending effects. This was tested in the model by training with stimulus A in conjunction with AB or B, and comparing the results to those found with stimulus D. Figure 15 shows that the effect of similarity on blending is not the same for both delay and no delay conditions. In the delay condition, the more similar the stimuli were, the more blending occurred. This is the predicted direction of the effect. However, in the no delay condition, the more similar the two stimuli, the *less* blending occurred. Indeed, the level of blending in the no delay condition was inversely proportional to the level of blending in the delay condition.

The inverse relationship between blending in the delay and no delay conditions suggests a kind of *ceiling effect* for blending. That is, if there is a relatively high degree of blending in the delay condition due to stimulus similarity, it is difficult for the additional blending force of the no delay condition to have any effect. Indeed, it appears that the system becomes *inoculated* against further blending by the presence of intrinsic blending. The explanation for this effect, like that for the reversibility effects described above, has to do with the covariance formulation of the learning rule, the presence of hysteresis, and the overlap of the similar stimuli.

When the stimuli overlap, weights that are enhanced for one stimulus will enhance the chance of being activated for the other stimulus. However, as we saw with the recruitment process described previously, the longer a unit is active, the more it decreases its weights to each individual element that activates it, in accordance with the conditional probability formulation of the weight update rule (see Appendix A and Rumelhart & Zipser, 1986). Thus, whenever a unit in layer 1 becomes active longer due to increasing weights from overlapping regions and increased hysteresis, the conditional probability for any particular input unit (which has a fixed probability of being active given by the stimuli) necessarily goes down, and so does the weight. So, the initial increase in activity for a unit gets balanced out by this decreasing force on the individual weights, causing this unit to become less active subsequently.

The pattern of increased-then-decreased activity results in a recruitment effect, since the unit tunes its weights to the stimuli while active, and gives other units an opportunity to do so when it becomes inactive. These other units subsequently experience the same up-then-down pattern of activation. The overall effect of the enhanced recruitment in this situation is to increase the chances that a unit becomes sensitive to one stimulus or the other, instead of both. Thus, paradoxically, increasing overlap in the input causes increased pressure to form differentiated representations in layers 1 and 2. This increased differentiation pressure explains why the blending in the delay condition for stimuli A - AB is only around 50% even though AB is known to show a high degree of generalization to A. Given that the tendency of generalization to cause blending is so strong in this case anyway, the additional pressure from the no delay condition results in only a small increase in blending. With A - B, the generalization pressure is less, and so is the *inoculization* effect from the overlap between them. Thus, the no delay has a larger effect, nearly doubling the level of blending. Finally, the distinct stimuli A - D have no *inoculization* effect at all, and blending is nearly complete in the no delay condition.

Predictions

Although the results on the effects of similarity and temporal contiguity are somewhat counter-intuitive, the mechanisms which cause them are quite general and should be present

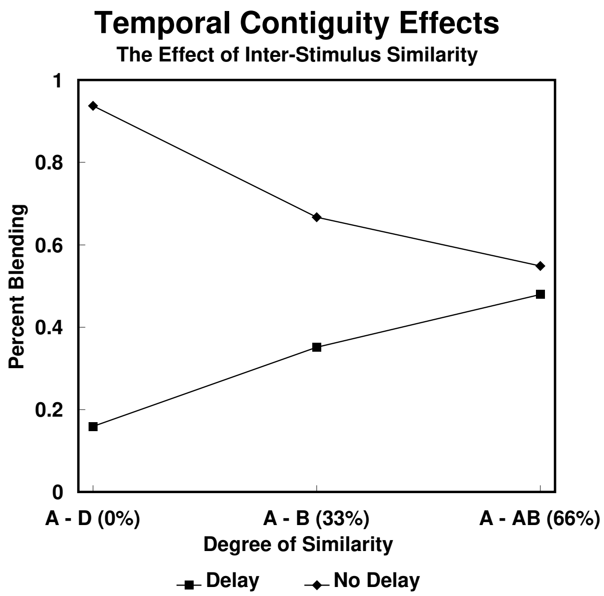

Figure 15: Simulation 5: Temporal contiguity manipulation (delay vs. no delay) interacts with similarity of the stimuli, so that in the delay condition, more similar stimuli (AB, B) show greater blending than dissimilar stimuli (D). The no delay condition causes increased blending in inverse proportion to the amount of blending in the delay case. This data is the average of both layers 1 and 2, but each layer individually shows the same pattern.

in the chick IMHV. In the model, they are the result of the same mechanisms that produced the reversibility and sensitive period effects, and are particular to the specific kind of object recognition learning rule hypothesized to be taking place in the chick IMHV. Specifically, our predictions are as follows:

- The difficulties in extending the Chantrey (1974) findings to other stimuli and experimental conditions (Stewart et al., 1977) may be due to the relative similarity of the stimuli used. According to our model, when very dissimilar stimuli are used, temporal blending will be observed. When stimuli are used which share many features, evidence for temporal blending will be more difficult to detect. Some support for this prediction comes from the experiments of Stewart et al. (1977), in which they failed to find temporal blending effects with stimuli that differed only in color.

- The temporal blending effect should be apparent in neural recording studies. With two dissimilar training stimuli, a short gap condition should result in more cells responsive to both stimuli. In contrast, a long gap condition with the same stimuli should reveal more neurons responsive to one or the other stimulus. This difference should not be apparent when two very similar stimuli are used.

Discussion

Through a series of simulations and analysis of the mechanisms behind them, we have been able to explain some of the behavioral phenomena surrounding chick imprinting with a simple network model. Further, this model is based on certain properties of the region of the chick brain known to be involved in imprinting, providing an account of the phenomena that spans from the neural to the behavioral.

Several important assumptions have been made in constructing the model. First, in applying a neural network model of translation invariant object recognition to imprinting phenomena in the chick, we have assumed that area IMHV is involved in the process of object recognition. Further, we have assumed that the response properties of neurons in IMHV are sufficient to determine the preference behavior of chicks, since our preference measures in the model were taken directly from the units corresponding to those in IMHV. However, we do not wish to imply that IMHV is the only neural structure which influences the preferences of chicks. Indeed, it is well established that there are other areas of the chick brain which are involved in both imprinting (Horn, 1985; Horn & Johnson, 1989) and in passive avoidance learning (Kossut & Rose, 1984; Rose, 1991). One proposal is that IMHV influences motor output by the selective inhibition of non-specific predispositions to approach medium-sized objects and peck at small objects (Johnson, 1991).

Once IMHV is viewed as mediating object recognition, it is possible to provide a functional role for the kind of connectivity found in this area. The pattern of connectivity is hypothesized to cause hysteresis in the activation states of the principal neurons in this region. As a result of this hysteresis, invariant visual object representations will develop. However, hysteresis also plays a major role in virtually all of the behavioral effects found in the simulations, from terminating the sensitive period to causing blending for stimuli seen in close temporal proximity. Since many of these effects have been supported by existing empirical evidence, and they can all be explained by the existence of hysteresis, this provides reasonably strong evidence for the proposed account of IMHV.

Aside from the specific predictions made regarding each of the different behavioral effects found in imprinting, we feel the model leads to several broader conclusions. To the extent that our account of IMHV fits the behavioral and neural data, it provides support for the computational model of object recognition upon which our account is based. Given the challenging nature of the object recognition problem from a computational standpoint, the existence of a simplified animal model of object recognition would provide a valuable tool for further exploration and understanding how neural systems recognize objects.

Further, the sensitive period issues explored in our model may have utility outside the sphere of visual imprinting in birds. For example, several authors have pointed out similarities between some of the phenomena associated with imprinting, and observations about the development and plasticity of the primate cortex (Rauschecker & Marler, 1987). In particular, there may be strong parallels between the reversibility effects examined in this paper, and the extent of recovery of visual acuity following monocular occlusion in the kitten (e.g., Mitchell, 1991). If monocular occlusion occurs for long enough, then no recovery in measurable vision is possible, however if the occlusion is short full recovery can occur.

A general issue that arises with a model such as the one presented in this paper concerns the extent to which it can be considered a "neural" model or purely an abstract psychological model. Clearly, the units in our model have nothing like the detail of a real neuron, or even of a relatively simplified point neuron model. For this reason we have been unable to make any quantitative predictions about patterns of firing of neuron types within IMHV, or about the temporal parameters which would produce blending effects. On the other hand, we have not simply adopted an "off the shelf" PDP model either. Instead, we have attempted to capture in a simplified computational framework a relatively few critical features of IMHV's neural structure that relate directly to its role as an object recognition system. The resulting model is simple enough to make the analysis of its behavior possible, furthering our understanding of the system and allowing us to gauge the generalizability of the results to the biological system. Thus, we have been able to make several qualitative predictions which are open to experimental investigation.

However, a more realistic model would be able to address the kinds of representations that form in each of the two identified types of IMHV excitatory principal neurons. There are two general categories of representational schemes that are consistent with the present model of IMHV. However, each of these schemes will affect the quantitative behavior of the system. The first kind of division of labor among the PN's holds that both types are functionally equivalent, and that the kinds of representations (receptive fields) formed in one are the same as in the other. Another division holds that one type (probably Tömböl et al., 1988 type 2), receives inputs from HA, and forms more localized representations compared to type 1 PN's, which form completely invariant representations. This distinction between the two types would suggest a more gradual invariance algorithm, of the type advocated in Mozer (1991) and O'Reilly (1992), and is the kind used in the present model. Whether these kinds of representations are likely to form given the biological parameters of IMHV, and the computational implications thereof could both be studied in the more realistic model. However, our initial experience with a more detailed point neuron model revealed that many of the variables crucial for the behavior and stability of the model are still unknown for chick IMHV.

Even in the more abstract model, there are important simplifications made that could potentially be captured at this level. The most obvious of these was the decision to use only translating images across the retina, and not to include objects rotating and moving in depth. While these relatively neglected aspects of invariant object recognition deserve further investigation, the basic invariance learning algorithm employed in our model should work for any form of visual transformation, not just translation. Given the trade-off between a more complete model of visual object recognition on the one hand and a simpler, more easily understood and analyzed system on the other, we opted for the latter. However, the pattern of weights that need to develop for representations that are invariant across rotation and size changes is less intuitively obvious than that for translation invariance. For this reason, it would be both educational and possibly revealing of limitations in the algorithm to test these other invariancies. Work on this is presently in progress.

*

Appendix A: Hebbian Weight Update Rule

The weight update rule used throughout was as follows:

$$\frac{1}{\lambda}\Delta w_{ij} = \begin{cases} (1.0 - w_{ij}) & a_j > 0, a_i > 0 \\ -w_{ij} & a_j > 0, a_i < 0 \\ 0 & otherwise \end{cases} \tag{1}$$

where λ is the learning rate.

The rule depends on the signs of the pre and postsynaptic terms (a_i and a_j, respectively), as the inhibition within a layer will drive all but a single unit into the negative activation range. Without hysteresis, this unit would always be the one with the largest net input from the current input pattern, making this formulation in combination with the lateral inhibition roughly equivalent to the Competitive Learning scheme (Rumelhart & Zipser, 1986). The hysteresis can cause an already-active unit that might not have the largest amount of input from the current pattern to remain active, which is the basis of the translation invariance learning, and in this way the system differs from Rumelhart & Zipser (1986). Note that it differs also from the Földiák (1991) scheme in that the hysteresis or trace component of the activation is implemented in the activation function through the influence of lateral inhibition and mutual excitation, and not directly in the weight update function.

*

Appendix B: IAC Activation Function

The IAC activation function (in a slightly modified form) contains an input and a decay term (with the relative contribution of these two factors controlled by modifying the level of the decay term, α) controlled by a rate parameter λ:

$$\Delta a_i = \lambda(f(net_i) - \alpha(a_i - rest)) \tag{2}$$

where $f(net_i)$ is the following input function:

$$f(net_i) = \begin{cases} net_i(max - a_i) & net_i > 0 \\ net_i(a_i - min) & net_i < 0 \end{cases} \tag{3}$$

and net_i is the net input to the unit:

$$net_i = \sum_j o_j w_{ji} + \sum_{l, l \neq i} o_l w_{li} \tag{4}$$

with l indexing over the other units in the same layer with inhibitory connections, and o_j representing the positive-only activation value of unit j (0 otherwise).

References

Artola, A., Brocher, S., & Singer, W. (1990). Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, 347:69–72.

Bateson, P. (1966). The characteristics and context of imprinting. *Biological Reviews*, 91:177–220.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2):115–147.

Bolhuis, J. J. (1991). Mechanisms of avian imprinting: A review. *Biological Reviews*, 66:303–345.

Bolhuis, J. J. & Bateson, P. (1990). The importance of being first: A primacy effect in filial imprinting. *Animal Behaviour*, 40:472–483.

Bolhuis, J. J. & Horn, G. (1992). Generalization of learned preferences in filial imprinting. *Animal Behaviour*, 44:185–187.

Bolhuis, J. J., Johnson, M. H., Horn, G., & Bateson, P. (1989). Long-lasting effects of IMHV lesions on social preferences in domestic fowl. *Behavioral Neuroscience*, 103:438–441.

Bradler, J. & Barrionuevo, G. (1990). Heterosynaptic correlates of long-term potentiation induction in hippocampal CA3 neurons. *Neuroscience*, 35(2):265–271.

Bradley, P., Davies, D., & Horn, G. (1985). Connections of the hyperstriatum ventrale in the domestic chick *gallus domesticus*. *Journal of Anatomy*, 140:577–589.

Brown, M. W. & Horn, G. (1992). Neurones in the intermediate and medial part of the hyperstriatum ventrale (IMHV) of freely moving chicks respond to visual and/or auditory stimuli. *Journal of Physiology*, 452:102P.

Chantrey, D. (1974). Stimulus pre-exposure and discrimination learning by domestic chicks: Effect of varying interstimulus time. *Journal of Comparative and Physiological Psychology*, 87:517–525.

Cherfas, J. & Scott, A. (1981). Impermanent reversal of filal imprinting. *Animal Behaviour*, 30:301.

Cofoid, D. & Honig, W. (1961). Stimulus generalization of imprinting. *Science*, 134:1692–1694.

Collingridge, G. & Bliss, T. (1987). NMDA receptors - their role in long-term potentiation. *Trends In Neurosciences*, 10:288–293.

Davey, J. & Horn, G. (1991). The development of hemispheric asymmetries in neuronal activity in the domestic chick after visual experience. *Behavioural Brain Research*, 45:81–86.

Davies, D., Taylor, D., & Johnson, M. H. (1988). Restricted hyperstriatal lesions and passive avoidance learning in the chick. *The Journal of Neuroscience*, 8:4662–8.

Douglas, R. J. & Martin, K. A. C. (1990). Neocortex. In Shepherd, G. M., editor, *The Synaptic Organization of the Brain*, chapter 12, pages 389–438. Oxford University Press, Oxford.

Einsiedel, A. A. (1975). The development and modification of object preferences in domestic white leghorn chicks. *Developmental Psychobiology*, 8(6):533–540.

Farah, M. J. (1990). *Visual Agnosia.* MIT Press, Cambridge, MA.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3(2):194–200.

Frégnac, Y., Shulz, D., Thorpe, S., & Bienenstock, E. L. (1988). A cellular analogue of visual cortical plasticity. *Nature*, 333:367–370.

Hoffman, H. & Ratner, A. (1973). A reinforcement model of imprinting. *Psychological Review*, 80:527–544.

Horn, G. (1985). *Memory, Imprinting, and the Brain: An inquiry into mechanisms.* Clarendon Press, Oxford.

Horn, G., Bradley, P., & McCabe, B. J. (1985). Changes in the structure of synapses associated with learning. *The Journal of Neuroscience*, 5:3161–8.

Horn, G. & Johnson, M. H. (1989). Memory systems in the chick: Dissociations and neuronal analysis. *Neuropsychologia*, 27:1–22.

Hubel, D. & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154.

Jaynes, J. (1956). Imprinting: the interaction of learned and innate behavior. I. development and generalization. *Journal of Comparative and Physiological Psychology*, 49:200–206.

Jaynes, J. (1958). Imprinting: the interaction of learned and innate behavior. IV. generalization and emergent discrimination. *Journal of Comparative and Physiological Psychology*, 51:238–242.

Johnson, M. H. (1991). Information processing and storage during filial imprinting. In Hepper, P., editor, *Kin Recognition*, pages 335–357. Cambridge University Press, Cambridge.

Johnson, M. H. & Horn, G. (1986). Dissociation of recognition memory and associative learning by a restricted lesion of the chick forebrain. *Neuropsychologia*, 24:329–340.

Johnson, M. H. & Horn, G. (1987). The role of a restricted region of the chick forebrain in the recognition of conspecifics. *Behavioural Brain Research*, 23:269–275.

Kertzman, C. & Demarest, J. (1982). Irreversibility of imprinting after active vs. passive exposure to the object. *Journal of Comparative and Physiological Psychology*, 96:130–142.

Klopfer, P. H. (1967). Stimulus preferences and imprinting. *Science*, 156:1394–96.

Klopfer, P. H. & Hailman, J. P. (1964a). Basic parameters of following and imprinting in precocial birds. *Z. Tierpsychol*, 21:755–62.

Klopfer, P. H. & Hailman, J. P. (1964b). Perceptual preferences and imprinting in chicks. *Science*, 145:1333–4.

Kohsaka, S., Takamatsu, K., Aoki, E., & Tsukada, Y. (1979). Metabolic mapping of chick brain after imprinting using [14C]2-deoxy-glucose technique. *Brain Research*, 172:539–544.

Kossut, M. & Rose, S. (1984). Differential 2-deoxyglucose uptake into chick brain structures during passive avoidance training. *The Journal of Neuroscience*, 12:971–977.

Lorenz, K. (1935). Der Kumpan in der Umwelt des Vogels. *Journal of Ornithology*, 83:137–213, 289–413.

Lorenz, K. (1937). The companion in the bird's world. *Auk*, 54:245–73.

Marr, D. (1982). *Vision*. Freeman, New York.

McCabe, B. J., Cipolla-Neto, J., Horn, G., & Bateson, P. (1982). Amnesic effects of bilateral lesions placed in the hyperstriatum ventrale of the chick after imprinting. *Experimental Brain Research*, 48:13–21.

McCabe, B. J. & Horn, G. (1988). Learning and memory: Regional changes in N-methyl-D-aspartate receptors in the chick brain. *Proc. Natl. Acad. Sci. USA*, 85:2849–53.

McClelland, J. L. (1981). An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review*, 88(5):375–407.

McCloskey, M. & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In Bower, G. H., editor, *The Psychology of Learning and Motivation, Vol. 24*, pages 109–164. Academic Press, Inc., San Diego, CA.

Mitchell, D. E. (1991). The long-term effectiveness of different regimens of occlusion on recovery from early monocular deprivation in kittens. *Philosophical Transactions of the Royal Society (Lond.) B*, 333:51–79.

Mozer, M. C. (1991). *The Perception of Multiple Objects: A Connectionist Approach*. MIT Press, Cambridge, MA.

O'Reilly, R. C. (1992). The self-organization of spatially invariant representations. Parallel Distributed Processing and Cognitive Neuroscience PDP.CNS.92.5, Carnegie Mellon University, Department of Psychology.

Payne, J. & Horn, G. (1984). Differential effects of exposure to an imprinting stimulus on 'spontaneous' neuronal activity in two regions of the chick brain. *Brain Research*, 232:191–193.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317:314–319.

Rauschecker, J. & Marler, P. (1987). *Imprinting and Cortical Plasticity: Comparative Aspects of Sensitive Periods*. Wiley, New York.

Rose, S. (1991). How chicks make memories: The cellular cascade from C-fos to dendritic remodelling. *Trends In Neurosciences*, 14:390–397.

Rumelhart, D. E. & Zipser, D. (1986). Feature discovery by competitive learning. In Rumelhart, D. E., McClelland, J. L., & PDP Research Group, editors, *Parallel Distributed Processing. Volume 1: Foundations*, chapter 5, pages 151–193. MIT Press, Cambridge, MA.

Ryan, C. & Lea, S. (1989). Pattern recognition, updating, and filial imprinting in the domestic chicken (*gallus gallus*). In Commons, M., Herrnstein, R., Kosslyn, S., & Mumford, D., editors, *Models of Behaviour: Behavioural Approaches to Pattern Recognition and Concept Formation. Quantitative Analyses of Behavior, vol. 8*, pages 89–110. Lawrence Earlbaum Associates, Inc., Hillsdale, NJ.

Salzen, E. A. & Meyer, C. (1968). Reversibility of imprinting. *Journal of Comparative Physiology*, 66:269–75.

Sejnowski, T. J. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, 4:303–321.

Shapiro, L. & Thurston, K. (1978). The effect of enforced exposure to live models on the reversibility. *The Psychological Record*, 28(479-485).

Sluckin, W. (1972). *Imprinting and Early Learning, 2nd edn.* Methuen, London.

Sluckin, W. & Salzen, E. A. (1961). Imprinting and perceptual learning. *Quarterly Journal of Experimental Psychology*, 13:65–77.

Stanton, P. K. & Sejnowski, T. J. (1989). Associative long-term depression in the hippocampus induced by hebbian covariance. *Nature*, 339:215–218.

Stewart, D., Capretta, P., Cooper, A., & Littlefield, V. (1977). Learning in domestic chicks after exposure to both discriminanda. *Journal of Comparative and Physiological Psychology*, 91:1095–1109.

Tömböl, T., Csillag, A., & Stewart, M. G. (1988). Cell types of the hyperstriatum ventrale of the domestic chicken *gallus domesticus*: A golgi study. *Journal für Hinforschung*, 29(3):319–334.

Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In Ingle, D. J., Goodale, M. A., & Mansfield, R. J. W., editors, *The Analysis of Visual Behavior*. MIT Press, Cambridge, MA.

Yuille, A. L. (1990). Generalized deformable models, statsitical physics, and matching problems. *Neural Computation*, 2(1):1–24.