Recurrent processing during object reocognition

by

Dean R. Wyatte

B.S., Indiana University, 2007

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Master of Arts

Department of Psychology and Neuroscience

2010

This thesis entitled:
Recurrent processing during object reocognition
written by Dean R. Wyatte
has been approved for the Department of Psychology and Neuroscience

Randall C. O'Reilly

Prof. Tim Curran

Prof. Matt Jones

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol #10-0117

Wyatte, Dean R. (M.A., Cognitive Psychology)

Recurrent processing during object reocognition

Thesis directed by Prof. Randall C. O'Reilly

Although the rich bidirectional architecture of the ventral visual stream has been documented for some time, relatively little work has gone into understanding its function in object recognition. Recently, computational modeling work has suggested that the computations performed within a bidirectional architecture (recurrent processing) could be beneficial to object recognition by cleaning up ambiguity in input signals, thus providing robustness to degradations like occlusion that underspecify the visual stimulus. The research described here tests this claim by using visual masking to disrupt this recurrent processing. In one experiment and a series of accompanying modeling simulations, it is shown that there is a significant interaction between a mask and occlusion such that performance on an object categorization task is differentially impaired when a moderately occluded stimulus is masked compared to a relatively unoccluded one. Furthermore, the modeling simulations provide a mechanistic explanation of how recurrent processing resolves ambiguity as well as how masking interrupts this process. Together, the results of the experiment and modeling simulations suggest that object recognition is a dynamic process characterized by interactions between adjacent areas along the ventral visual stream.

Contents

Chapter

Figures

Figure

Chapter 1

Introduction

The mapping of cortico-cortical connections in the mammalian visual system has been shown to be almost completely bidirectional between adjacent sites (Felleman & Van Essen, 1991; Scannell, Blakemore, & Young, 1995; Sporns, Honey, & Kotter, 2007; see Sporns & Zwi, 2004, for a review). This pattern of recurrent connectivity has been investigated from a general computational perspective (e.g, Hopfield, 1982) and has even been implicated in early visual processes such as border ownership coding and figure-ground segmentation (Hupe et al., 1998; Hupe, James, Girard, & Bullier, 2001; Craft, Schutze, Niebur, & von der Heydt, 2007). However, it remains unclear whether the computations performed within recurrent circuits (hereafter referred to as recurrent processing) are similarly important for higher-level visual processes such as visual object recognition.

Perhaps one reason for the lack of understanding regarding the relationship between recurrent processing and object recognition is the acceptance of a standard model of object recognition in cortex (e.g., Riesenhuber & Poggio, 1999; Fukushima, 1980; Wallis & Rolls, 1997; Masquelier & Thorpe, 2007; see Riesenhuber & Poggio, 2000, for a review), which posits that a series of feedforward processing stages is sufficient for object recognition. Support for this standard model likely comes from its ability to explain data across multiple levels of analysis, from the single-cell neurophysiology of object recognition (Freedman, Riesenhuber, Poggio, & Miller, 2003) to humans' performance on real-world categorization tasks (Serre, Oliva, & Poggio, 2007). Despite the longstanding appeal of the standard model, it remains to be reconciled with the neuroanatomical data that indicate a large number of recurrent connections in the ventral visual stream. To address this this limitation, the standard model predicts that recurrent processing is not a necessary mechanism of object recognition and instead, backprojections probably play a secondary role to object recognition, serving "after-the-fact" functions such as feature-based attention (Riesenhuber & Poggio, 1999, 2000; Serre et al., 2007).

The goal of the present research is to investigate whether recurrent processing plays a more fundamental role in visual object recognition. In lieu of an established literature demonstrating such a relationship between recurrent processing and object recognition, computational principles that are not specific to vision such as attractor dynamics and mutual constraint satisfaction (Amit, Bruneland, & Tsodyks, 1994; Amit & Brunel, 1997; O'Reilly, 1998; O'Reilly & Munakata, 2000) are drawn upon to illustrate the plausibility of the theory. These computational principles have been embodied in a novel model of object recognition in cortex, the Leabra model of object recognition, which in turn, has generated explicit predictions regarding the nature of the relationship between recurrent processing and object recognition (O'Reilly, Wyatte, Herd, Mingus, & Jilk, in preparation). Specifically, the Leabra model predicts that recurrent processing could be important for cleaning up ambiguity in the input signal, such as when visual information is missing due to occlusion or deletion, by using high-level learned visual knowledge to strengthen weak or otherwise unreliable lower-level signals.

To test the Leabra model's prediction, an experiment was conducted that varied the amount of visual occlusion applied to stimuli during an object categorization task. The experiment was further characterized by incorporating visual masking, which has been suggested to disrupt recurrent processing (Fahrenfort, Scholte, & Lamme, 2007; see Lamme & Roelfsema, 2000, for a review). If recurrent processing does indeed rectify inputs that are made ambiguous due to occlusion, performance on the object categorization task should be impaired when a mask follows a stimulus that is moderately occluded such that diagnostic visual information is removed, but spared when a mask follows a stimulus with diagnostic visual information intact. This pattern of results, when compared with trials that do not contain a mask, should produce an interaction between the mask and occlusion. To anticipate the results of the experiment, a significant interaction was found between the mask and occlusion. Although the experimental data were subject to a floor effect, judicious analysis and accompanying modeling simulations in which the Leabra model was used to explicitly simulate the results of masking on occluded object recognition suggest that the interaction is valid.

The Leabra model also provided a mechanistic explanation for the results of the masking experiment that illuminates the dynamics of the object recognition process. Specifically, the identity of stimuli that were relatively unoccluded could be resolved rather quickly, without the need for much (if any) recurrent processing. However, recognizing the moderately occluded stimuli required much more extensive recurrent processing to strengthen the neural signals required for their identification, which took considerably longer than the relatively straightforward recognition of unoccluded stimuli. The mask was shown to interact selectively with the dynamics required to resolve the moderately occluded stimuli.

The rest of this thesis is structured as follows. Chapter 2 establishes the motivation for the experiment by reviewing the standard model of object recognition in cortex and previous demonstrations of recurrent processing in early visual processes, as well as the relevant background in the computational modeling of recurrent processes. In particular, the Leabra model of object recognition is reviewed and followed by a discussion on how visual masking can be used to test its predictions. Chapter 3 describes the methods of the experiment used to test the predictions of the model and Chapter 4 presents the results. Chapter 5 describes the accompanying modeling simulations which are used to validate the results of the experiment, as well as provide a mechanistic explanation of how masking interacts with recurrent processing during object recognition. Chapter 6 discusses the results of the experiment in the context of the modeling simulations and addresses the discrepancies between experimental subjects' data and the predictions of the model. Overall, the research presented here demonstrates a relatively tight coupling between the Leabra model and experimental data signifies the potential for using a biologically realistic model of object recognition for generating predictions for experiments regarding the dynamics of object recognition as well as providing a framework for interpreting their results.

Chapter 2

Relation to Previous Work

2.1 Object recognition in cortex

Object recognition in cortex is supported by the ventral visual stream, which consists of primary visual cortex (V1), extrastriate visual cortex (V2, V4), and inferior temporal (IT) cortex (Felleman & Van Essen, 1991; Ungerleider & Haxby, 1994; Riesenhuber & Poggio, 2002). These areas have been suggested to be arranged hierarchically such that processing takes place in early visual areas such as V1 prior to being processed by higher-level visual areas like IT. Furthermore, the receptive field structure of neurons in a given area grows in complexity as a function of its level in the hierarchy. Neurons in V1 respond best to simple stimuli like oriented bars of light and sinusoidal gratings within specific locations of the visual field (Hubel & Wiesel, 1962, 1968). Downstream neurons in V2 and V4 are tuned to slightly more complex features (e.g., contours, junctions) in a greater range of locations in the visual field (Desimone & Ungerleider, 1989). IT neurons continue this trend of discarding spatial information in favor of feature complexity, responding to moderately complex visual features regardless of their location in the visual field (Bruce, Desimone, & Gross, 1981; Logothetis, Pauls, & Poggio, 1995; Tanaka, 1996). IT neurons then project to prefrontal cortex (PFC), which has been shown to contain category-tuned neurons (Freedman, Riesenhuber, Poggio, & Miller, 2001), although its exact role in object recognition remains to be understood mechanistically (Miller, Nieder, Freedman, & Wallis, 2003).

Computational models of object recognition in cortex that implement the architecture of the ventral visual stream via a series of feedforward processing stages have collectively been referred to as the standard model of object recognition (Riesenhuber & Poggio, 2000). This class of models has been successful in explaining a wide range of phenomena across multiple levels of analysis. For example, Fukushima's (1980) *Neocognitron* demonstrated that interleaved layers of *S*-cells that coded for features in specific areas of the visual field in combination with *C*-cells that gradually corrected for small errors in the features' spatial location promoted the development of IT-like cells that were tuned to complex visual features regardless of their location in the input layer. Subsequent models using the same pattern of interleaved layers with gradual spatial pooling were similarly able to predict more idiosyncratic properties of IT neurons such as their invariant response to stimuli regardless of size and viewpoint (Wallis & Rolls, 1997; Riesenhuber & Poggio, 1999). The same basic architecture has also predicted the effects of category learning on IT and PFC neurons' tunings (Freedman et al., 2003) and has provided a highly accurate prediction of humans' distributions of errors on real-world categorization tasks (Serre et al., 2007).

Despite these successes, there remains a substantial disconnect between the feedforward architecture of the standard model of object recognition and the rich bidirectional anatomy of the ventral visual stream. Although backprojections have been assigned roles secondary to object recognition, such as feature-based attention (Riesenhuber & Poggio, 1999, 2000; Serre et al., 2007), intracranial recording data from the macaque cortex suggest that neural activation recurs in the ventral stream via backprojections well within the time course of normal object recognition, even when objects are viewed passively sans attention (Lamme & Roelfsema, 2000; see also Foxe & Simpson, 2002, for human electrophysiology). Furthermore, this more general recurrent activation has been demonstrated to occur regardless of whether the studied animals were awake or anesthetized (Roland, 2010), further dissociating it from secondary visual processes that might require conscious effort from the observer.

This pattern of recurrence in which higher-level visual areas feed back into lower-level visual areas early on during processing has been demonstrated to be important for other visual processes, such as figure-ground segmentation. For example, Hupe et al. (1998) found that inactivating macaque area V5/MT (a higher-level area in the dorsal visual stream that feeds back into primary visual and extrastriate cortex), significantly attenuated the responses of V1, V2, and V3 neurons during a task in which an illuminated bar was discriminated from a textured background. Subsequent research indicated that the latency of this recurrence was relatively short, with V5/MT inactivation having an effect on V1-V3 neurons within 10 ms of their first response, and that the effect often lasted until the neurons finished responding (Hupe et al., 2001). These results, although occurring partially outside the bounds of the ventral visual stream, suggest that one effect of recurrent processing could be to amplify and strengthen a neural representation.

How might recurrent processing benefit object recognition? Given the receptive field structure of ventral visual stream, which grows in complexity at higher levels, it is possible that high-level visual feature representations in IT cortex could strengthen the response of lower level visual areas that code for the features' constituents. Similarly, knowledge of visual categories represented in PFC could strengthen visual feature representations of those categories in IT. Regardless, the bidirectional architecture of the ventral visual stream and data on recurrent processing during figure-ground segmentation provides reason to speculate that object recognition is a much more dynamic process than the standard feedforward model has posited.

2.2 Modeling recurrent processing during object recognition

The pattern of bidirectional connectivity in the ventral visual stream is consistent across multiple species of mammals, each of whose environment requires very different visual demands. Furthermore, bidirectional organization is not limited to the visual cortices, but is apparent across many neural systems (Sporns & Zwi, 2004). Taken together, these pieces of evidence suggest that recurrent processing serves a highly general function, not necessarily specific to vision.

One of the first accounts of the general computational properties of recurrent circuits was demonstrated by Hopfield (1982), in which he described a neural network whose connectivity pattern formed cycles between nodes. The behavior of the Hopfield network was governed by an "energy function" that contained minima that produced stable network states. The implication of this modeling work was that network behavior was a dynamic process that could ultimately converge on these stable states if they were to become partially active. This theory of dynamic, attractor-based processing has been more rigorously applied to neurophysiology by Amit and colleagues (Amit et al., 1994; Amit & Brunel, 1997) to demonstrate that such networks are plausible given the architecture and neurobiology of the brain and are capable of capturing quantitative properties of actual single-cell recordings of a visual working memory task.

O'Reilly and colleagues have described a similar principle in their Leabra theory of the neocortex, referred to as multiple constraint satisfaction (O'Reilly, 1998; O'Reilly & Munakata, 2000), in which neural network attractor states are a compromise between low-level perceptual representations and high-level conceptual representations. Similar to the Hopfield network's energy function, the overall activation in Leabra networks, which is determined by both feedforward and feedback weights, is pulled toward these attractor states over the course of processing. Multiple constraint satisfaction can thus be viewed as being advantageous over other processing strategies like strictly feedforward activation propagation because it allows all brain areas to weigh in on the decision process in proportion to their learned importance via recurrent processing dynamics.

Given that recurrent processing dynamics have successfully been implemented in a Leabra model of figure-ground segmentation (Vecera & O'Reilly, 1998), the Leabra model of object recognition (Figure 2.1) was developed to determine whether the same principles could be successfully applied to object recognition. The Leabra model employs the same general computational principles of recurrent connectivity demonstrated by Hopfield's and Amit and colleagues' attractor-based neural networks, but does so in a hierarchical manner constrained the neuroanatomy of the ventral visual stream. Thus, the Leabra model can be viewed as an extension of the standard model of object recognition, but with the key addition of excitatory backprojections between hierarchically adjacent visual areas.

To test whether the bidirectional connectivity of the Leabra model was beneficial to object recognition, O'Reilly et al. (in preparation) degraded visual inputs using an occlusion paradigm that removed Gaussian shaped areas of information (Figure 2.2a). This occlusion manipulation had the effect of introducing ambiguity into the recognition process by rendering inputs underspecified

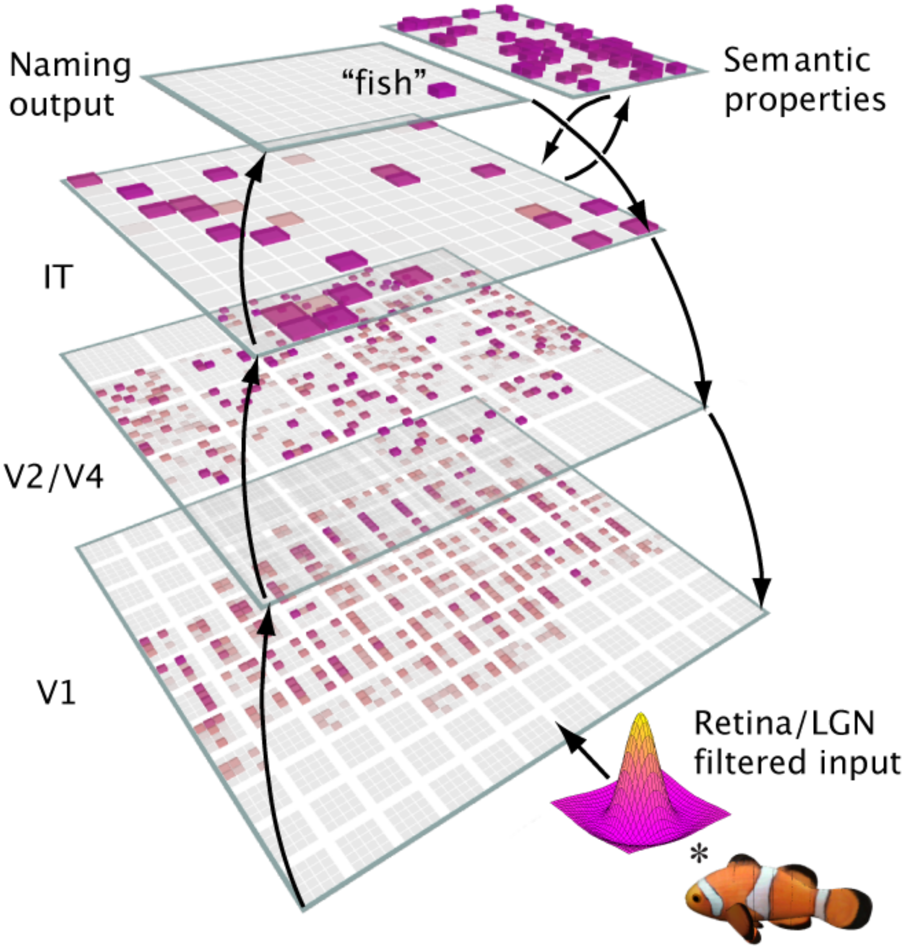

Figure 2.1: The Leabra model of object recognition embodies a general theory of the neocortex with a fundamental assumption that brain areas are bidirectionally connected. Like the standard model of object recognition, the Leabra model is partially characterized by a series of a feedforward processing stages corresponding to canonical areas along the ventral visual stream. However, the Leabra model makes a significant departure from the standard model with its key addition of excitatory backprojections between hierarchically adjacent visual areas. This bidirectional architecture has been suggested to support recurrent processing dynamics (O'Reilly, 1998; O'Reilly & Munakata, 2000), which could potentially prove important for object recognition. The Leabra model has also demonstrated the effects of this bidirectional architecture on the nature of learned representations, such as how higher-level "semantic" areas can constrain the learning of visual features (see O'Reilly et al., in preparation, for more information).

with respect to trained exemplars. The performance of the Leabra model was compared with other candidate models lacking backprojections in a test of object category generalization. The results of this test (Figure 2.2b) indicated that the bidirectional architecture of the Leabra model provided a significant advantage over a strictly feedforward architecture. Furthermore, the magnitude of this advantage tended to increase as a function of the amount of occlusion applied to the inputs.

The results of O'Reilly et al.'s simulations suggest that recurrent processing is indeed beneficial to object recognition. Specifically, recurrent processing appears to resolve ambiguity in input signals by strengthening ambiguous representations using top-down, learned signals from valid categories to reinforce their associated visual features and and constrain the overall recognition process. This recurrent dynamic has been suggested to exist in theoretical models of high-level vision (e.g., Ullman, 1995; Hochstein & Ahissar, 2002), but thus far has not been demonstrated mechanistically in a way that is consistent with the neurophysiological data on object recognition (Lamme & Roelfsema, 2000). Computational models of object recognition that have implemented bidirectional connectivity have thus far been limited to generative models in which the top-down connections provide a net inhibitory effect that cancels out representations that are consistent with bottom-up inputs during learning (Rao & Ballard, 1997; Hinton, 2010). In contrast, the Leabra model's top-down connections are purely excitatory, fundamentally no different than bottom-up connections, and provide additional benefits like ambiguity resolution in addition to enabling learning to take place in a biologically plausible way.

Although O'Reilly et al.'s simulations did not explicitly test whether recurrent processing was the basis of the Leabra model's advantage over the model's that operated in a strictly feedforward manner, their results indicated that the model's bidirectional architecture did provide a selective advantage in recognition performance when images were moderately occluded. Importantly though, recurrent processing requires a bidirectional architecture to form recurrent circuits between hierarchically adjacent processing sites. The Leabra model's prediction is still testable though, given that there is a way to prevent backprojections from engaging in visual processing during object recognition.

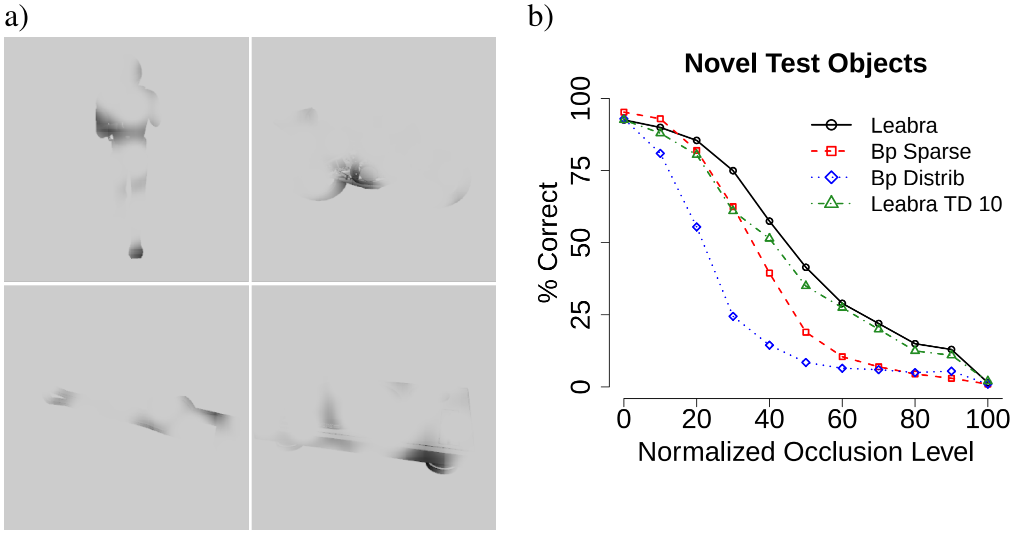

Figure 2.2: O'Reilly et al. (in preparation) tested the effects of a occlusion on object recognition to determine if the bidirectional architecture of the Leabra model provided an advantage over a strictly feedforward model. **a)** The occlusion paradigm removed Gaussian shaped areas of information from object stimuli, rendering them ambiguous at moderate levels of occlusion (shown here, 50% occlusion). **b)** The performance of the Leabra model was compared with candidate other candidate models lacking backprojections on a test of object category generalization. The results of this test indicate that the bidirectional architecture of the Leabra model provides a significant advantage over a strictly feedforward architecture and furthermore, the magnitude of this advantage is increasingly apparent at increasing levels of occlusion. **Bp Sparse:** Feedforward, backpropagation-trained model including optimized parameters and k-Winners-Take-All (kWTA) competitive inhibitory dynamics that facilitate the forming of sparse distributed representations in line with those learned by the Leabra model. **Bp Distrib:** Feedforward, backpropagation-trained model without optimized parameters or kWTA dynamics. **Leabra TD 10:** Leabra model with backprojections scaled to 10% of their trained strength.

2.3 Testing predictions about recurrent processing during object recognition

It has been suggested that one way to prevent backprojections from engaging in visual processing is to use a visual masking paradigm. In a visual masking experiment, a first stimulus is followed by a second stimulus with some, usually short, temporal delay. At very short delays, the second stimulus impairs the visibility of the first stimulus and in extreme cases, renders it imperceptible (Macknik & Livingstone, 1998). This retroactive effect of the mask is counterintuitive since prevalent theories of object recognition that assume a strictly feedforward processing architecture require that stimuli are recognized in a “first in, first out” manner, and thus, if anything, the first stimulus should override any subsequent perception of the mask. These feedforward theories have attempted to work around this problematic prediction by positing that the mask somehow catches up with the encoding of the original stimulus and impairs it in some way, such as by inhibition from lateral connections within a processing area (Francis, 1997; Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005).

However, single-cell recording studies of masking have indicated that the neurophysiological changes that prevent perception of the masked stimulus could occur well after the initial stimulus encoding is finished. For example, Rolls and Tovee (1994) found that face-selective neurons in IT normally respond to a stimulus of a face within 75 ms and then exhibit a sustained response for 200-300 ms even in the absence of the stimulus. When the stimulus of the face was followed by a pattern mask, the sustained response of the neurons was drastically shortened, whereas the initial encoding response was left relatively intact. The amount of the reduction of the sustained response was highly correlated with the latency of the mask such that faster mask latencies resulted in a greater reduction of the neurons sustained response. Subsequent investigations indicated that this reduction in IT neurons’ sustained response greatly reduces the amount of information available to the recognition process (Rolls, Tovee, & Panzeri, 1999).

How does a reduction in neurons’ sustained responding disrupt recurrent processing? Lamme and Roelfsema (2000) suggests that masking creates a mismatch between feedforward- and feedbackbased activation. Specifically, activation associated with the initial stimulus propagates through the ventral visual stream in a feedforward manner, just as it would during standard object recognition. This activation ultimately reaches processing sites that contain backprojections to lower-level areas through which recurrent processing acts to strengthen the representation of the stimulus in the absence of its physical display. If a mask stimulus is encoded at a short latency following the initial stimulus, its activation will propagate into the circuit where the recurrent processing is occurring. This simultaneous processing of both the initial stimulus and the mask within the same recurrent circuit could decrease the visibility of the original stimulus. Accordingly, functional neuroimaging research has indicated that masking causes a dynamic decoupling in the functional connectivity (i.e., co-activation) between low-level and high-level visual areas, which is correlated with visibility (Haynes, Driver, & Rees, 2005).

Although visual masking has not been specifically used in conjunction with the occluded object recognition paradigm used in O'Reilly et al.'s (in preparation) modeling simulations, visual masking has been used to investigate the time course of the neural events leading up to explicit object recognition (i.e., conscious recollection of the object's category). Functional neuroimaging has revealed that the strength of activation in object-selective areas is both reduced by masking and positively correlated with subjects' ability to explicitly name an object (Grill-Spector, Kushnir, Hendler, & Malach, 2000; Bar et al., 2001). Interestingly, the strength of the neural activation or signal associated with explicit recognition in these experiments was asymptotic as it approached levels associated with explicit recognition, suggesting that the process of explicit recognition is iterative. It is likely that this asymptotic pattern of activation is indicative of a recurrent processing dynamic such as evolution of activation toward the attractor state associated with the explicit recognition process.

The foremost goal of the present research was to determine whether visual masking is a suitable paradigm for investigating the relationship between recurrent processing and object recognition. Given that O'Reilly et al.'s Leabra model predicted that a bidirectional architecture provides a selective advantage over a strictly feedforward one when recognizing occluded objects, masking can potentially be used to prevent backprojections within a bidirectional architecture from engaging in visual processing. To test this prediction, an experiment was conducted that varied the amount of occlusion applied to stimuli during an object categorization task. If recurrent processing does indeed serve to disambiguate occluded inputs, performance should be impaired when a mask follows a stimulus that is occluded to underspecification, but spared when a mask follows a stimulus that is sufficiently specified. This pattern of results, when compared with trials that do not contain a mask, should produce an interaction between the mask (present or absent) and occlusion (moderate or low).

Additionally, the Leabra model was used to explicitly simulate the results of the masking experiment. Instead of simply measuring the model's performance on a test of occluded object recognition and comparing its performance to a class of feedforward models a la O'Reilly et al., the inputs to the Leabra model were swapped with an image of a mask while it processed the objects. The results of this simulation are qualitatively similar to the predictions of O'Reilly et al. and provide a good quantitative fit of subjects' actual data. Furthermore, examining the dynamics of the model gave insight into the overall dynamics of the object recognition process and provided a mechanistic framework within which the results of the experiment could be interpreted.

Chapter 3

Experimental Methods

A total of 19 subjects from the University of Colorado Boulder participated in the experiment as part of their introductory psychology course credit (11 males, 8 females, mean age of 18.89 years). All subjects reported normal or corrected-to-normal vision and gave informed consent prior to the experiment in accordance with human subjects policy at the University of Colorado Boulder.

3.1 Stimuli

During the experiment, subjects were required to categorize images from six real-world occurring object categories – *cannon, car, fish, gun, key,* and *trumpet* (Figure 3.1a). The categories and their comprising images were taken from the *CU3D 100* dataset (O'Reilly et al., in preparation). The specific categories used in the experiment were chosen due to their sharing a horizontal axis of orientation, preventing subjects from using coarse orientation information as a cue for category membership. The images were processed with the SHINE toolbox (Willenbockel et al., 2010) to convert their colorspace to grayscale and to normalize luminance across categories. During the experiment, the images subtended approximately 16 degrees of visual angle.

Images of seven exemplars from each category were used during the full experimental session. Prior to beginning the actual experiment, subjects were shown images of two views (one left-facing, one right-facing) of two exemplars from each category (24 total images) to familiarize themselves with the basic visual category structure of the dataset. Images of twenty views of the

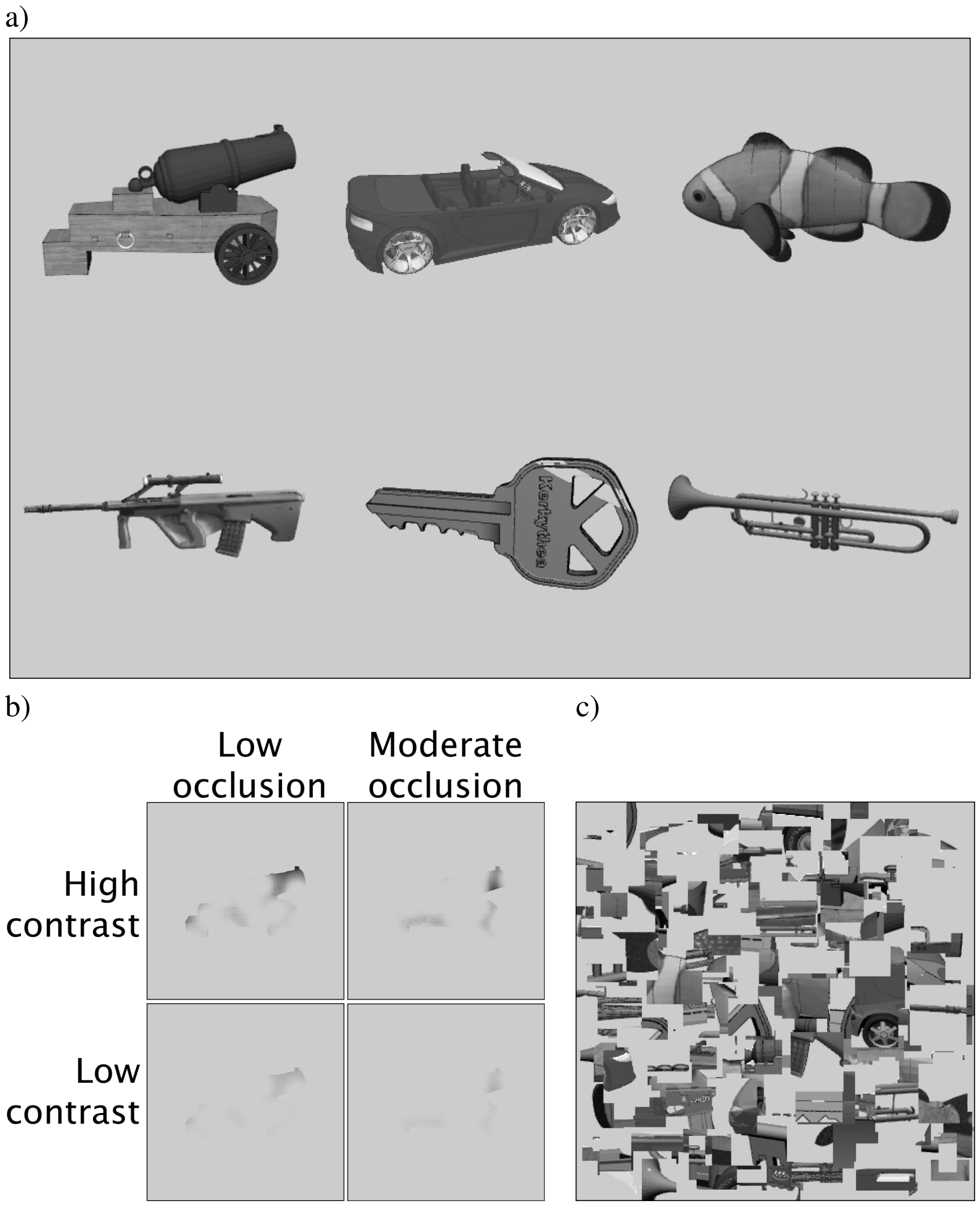

Figure 3.1: Stimuli were taken from the *CU3D 100* dataset (O'Reilly et al., in preparation), which consists of three-dimensional views of object exemplars from 100 real-world occurring categories **a)** One view of one exemplar from the six object categories used in the actual experiment. **b)** Occlusion and contrast manipulations were crossed across two levels: low and moderate occlusion and high and low contrast **c)** Masks were constructed by sampling patches of the images used in the actual experiment and randomly assembling them into a new image.

remaining five exemplars from each category were used during the experiment itself (600 total images). During the familiarization phase, subjects were informed that the specific images they were viewing would not be used in the experiment itself in order to discourage the memorization of visual features associated with each category. The familiarization phase was self-paced, but subjects never took more than 1 minute in practice.

Occlusion was manipulated using the same method as in O'Reilly et al. First, a filter was constructed that comprised a circle with a radius of 5% of the image size whose edges were softened with a Gaussian. This filter was applied to the image at a random location by taking a weighted average between the background gray intensity of the image and the pixel intensities at the location of application. Two levels of occlusion were used in the experiment. During low occlusion trials, the filter was applied 29 times to the image and during moderate occlusion trials, the filter was applied 73 times. In both cases, application of the filter was an iterative process such that the filter could be applied to the same location more than once.

Contrast was independently manipulated to ensure that the potential interaction between the mask and occlusion was not due to issues of scale dependency as well as to provide an orthogonal degradation from which the interaction between the mask and occlusion could be dissociated (see Bogartz, 1976; Newell & Dunn, 2008, for discussions of these issues). Two levels of contrast were used in the experiment. During high contrast trials, the contrast of the image remained at its original value. During low contrast trials, the contrast of the image was scaled to 25% of the original range. The background gray level was held constant during the contrast reduction process.

The masks were constructed by sampling patches of the images used during the actual experiment and assembling them into a new 320 x 320 image, subtending the same visual angle (16 degrees) as the raw images. The size of the sampled patch varied between 16 x 16 and 64 x 64 pixels and was randomly sampled from a region surrounding the bounding box of the object in each image (to prevent sampling the background). The resulting image patches were placed at random into a new image with the same background gray level as the original images. Like the occlusion algorithm, the patches were placed into the new image in an iterative manner and were allowed to overlap. A total of 416 samples were taken across the 600 source images. A total of 600 masks were pre-generated for use in the experiment.

Examples of the *CU3D* stimuli with occlusion applied and contrast reduced as well as an example of one of the masks used in the experiment can be viewed in Figure 3.1b-c.

3.2 Procedure

During the experiment, subjects were seated approximately 45 cm from a monitor running at a resolution of 1024 x 768 at 120 Hz. Stimuli were presented using the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997) to synchronize their display with the monitor's refresh interval. The monitor was calibrated using a luminance photometer to correct for nonlinearities in its Gamma response function (see Brainard, Pelli, & Robson, 2002).

The experiment was characterized by eight trial types, reflecting the factorial crossing of the stimuli manipulations with the masks – low or moderate occlusion, high or low contrast, and mask absent or mask present. On each trial, the subject was presented with a fixation cross for 300 ms, followed by the object stimulus. On trials during which the mask was absent, the object stimulus remained visible for 100 ms. On trials during which the mask was present, the object stimulus was replaced after 100 ms by a randomly selected mask which remained visible for an additional 100 ms. Subjects were then presented with a response screen that contained the six category names. Subjects' responses were collected via a QWERTY keyboard using the *S, D, F, J, K,* and *L* keys. The arrangement of the category names on the response screen was isomorphic with the placement of subjects' fingers on the keyboard to facilitate their responding without having to explicitly recall the key associated with their response. Subjects were required to respond within 5000 ms. The response screen remained visible until the subject responded. The ordering of events within a single trial is depicted in Figure 3.2.

All trial types were presented randomly in blocks of 50 trials. The experiment consisted of 1000 total trials. Subjects received feedback after each trial regarding only whether their response was correct or incorrect (i.e., if incorrect, the correct category was not given as an additional source

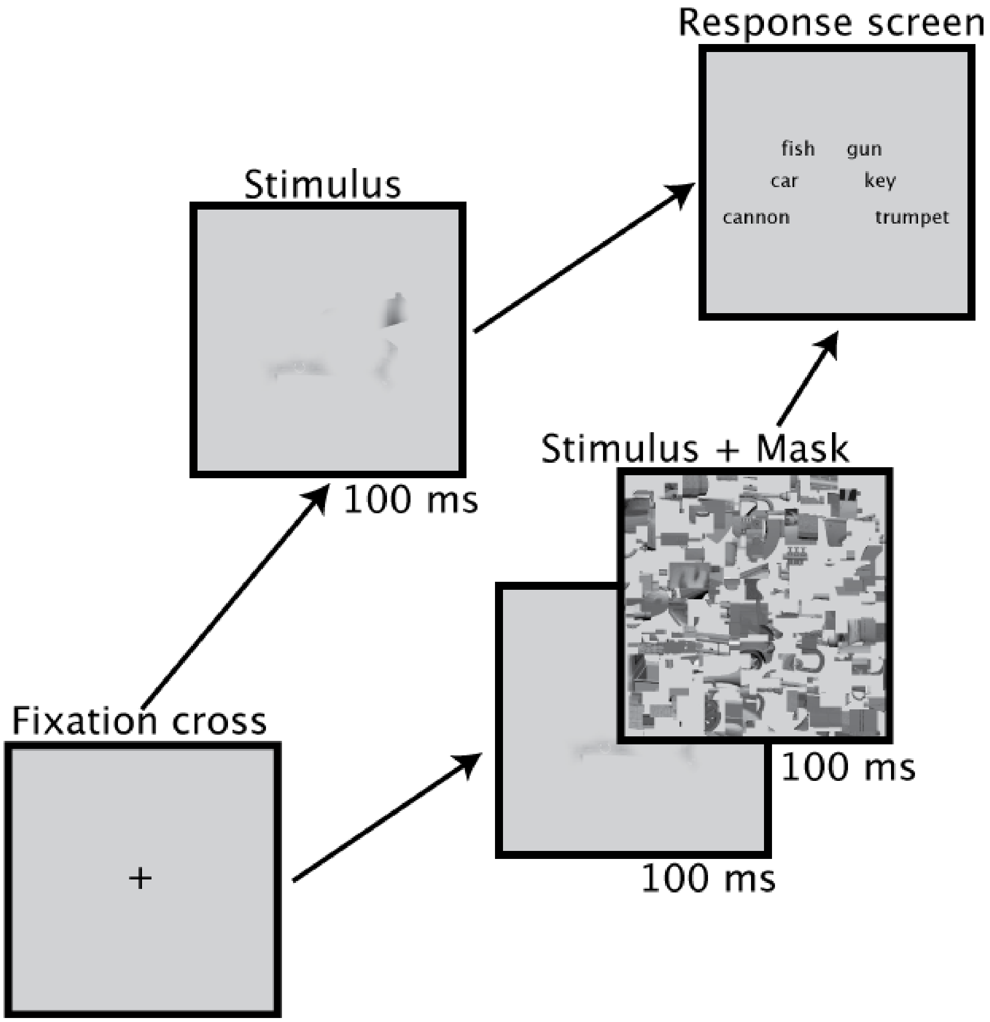

Figure 3.2: Experimental trials consisted of either a degraded object stimulus in isolation, or a degraded object stimulus followed by a mask. On trials that contained a mask, the object stimulus was replaced by the mask after 100 ms. All trials ended with a response screen that contained the names of the six categories used in the experiment presented in a rough topological ordering with respect to the placement of subjects' fingers on the keyboard.

of feedback).

Chapter 4

Results

The results of the experiment are plotted in Figure 4.1a-b. A total of three subjects were excluded from statistical analysis – two subjects for response levels that were well below 1.5 times the interquartile range of the data in the putatively easiest condition (i.e., mask absent, low occlusion, high contrast) and one subject for being unable to complete the full experiment. The remaining 16 subjects were included in the final analysis.

4.1 Statistical analysis

Subjects' data from the experiment were submitted initially to a 2 x 2 x 2 repeated measures ANOVA that assumed the presence of the mask (absent or present), occlusion (low or moderate), and contrast (low or high) as within-subjects factors. There were significant main effects of all three factors. The mask significantly impaired subjects' performance ($F(1,15) = 138.46$, $p < .001$), as did occlusion ($F(1,15) = 3906.20$, $p < .001$) and contrast ($F(1,15) = 525.12$, $p < .001$). The mask exhibited a significant interaction with contrast ($F(1,15) = 17.09$, $p < .001$), but not occlusion ($F(1,15) = .40$, $p = .537$). The interaction between occlusion and contrast was also significant ($F(1,15) = 6.79$, $p = .02$). Given the magnitude of the main effects, it is likely that the data were subject to a floor effect in conditions that were extremely degraded. Accordingly, the interaction between all three factors was significant ($F(1,15) = 17.87$, $p < .001$), indicating that the nature of the interaction between two of the factors (e.g., the mask and occlusion) was different depending on the level of the remaining factor (e.g., contrast).

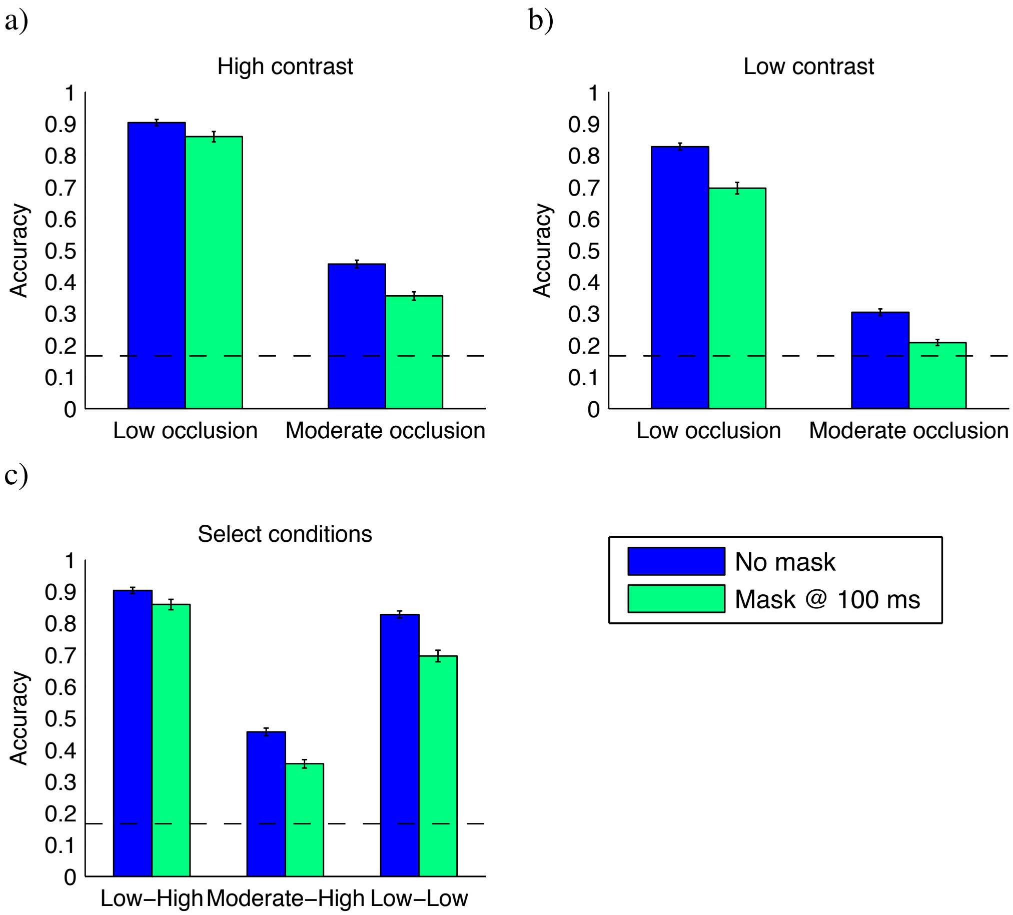

Figure 4.1: The data from the experiment reflected significant main effects of all three factors – the mask, occlusion, and contrast. Of particular interest though is whether there was an interaction between the mask and occlusion. This interaction was not significant, but the results of the ANOVA indicated a significant interaction between all three factors that prompted further analysis. **a)** Under high contrast, there was a significant interaction between the mask and occlusion such that the mask had a greater impairment on performance in the moderate level of occlusion. **b)** Under low contrast, the interaction between the mask and occlusion was also significant, but had the opposite effect of that under high contrast. However, it is likely that the data were subject to a floor effect. Chance performance on the task (1/6 or 16.67%) is indicated by a black dotted line. **c)** To interpret the data without being subject to the floor effect, three conditions were selected for analysis on the grounds that they were least likely to suffer from the floor effect – low occlusion under high contrast, moderate occlusion under high contrast, and low occlusion under low contrast. *t*-tests that compared the effect of the mask across these conditions indicated a significant interaction between the mask and occlusion as well as a significant interaction between the mask and contrast. Error bars on all plots indicate standard error of the mean.

To further investigate the three-way interaction between the mask, occlusion, and contrast, the data from the high contrast and low levels of contrast were analyzed independently. Under high contrast, the interaction between the mask and occlusion was significant ($F(1,15) = 5.38$, $p = .03$) such that the mask had a greater impairment on performance in the moderate level of occlusion. Under low contrast, the interaction between the mask and occlusion was also significant ($F(1,15) = 6.46$, $p = .02$). However, the nature of the interaction was opposite of that under high contrast, such that the mask had a greater impairment on performance in the low level of occlusion. Although this pattern of results could imply that the interaction between the mask and occlusion is non-monotonic, it is again likely that the observed effects are due to a floor effect.

To interpret the data without being subject to the floor effect, a repeated measures *t*-test was used to test the significance of the effect of the mask under three conditions – low occlusion under high contrast, moderate occlusion under high contrast, and low occlusion under low contrast (Figure 4.1c). These tests omit the condition from the full factorial design that was most likely subjected to the floor effect – moderate occlusion under high contrast. The results of these tests indicated that the effect of the mask was significant across all three conditions (for low occlusion under high contrast, moderate occlusion under high contrast, and low occlusion under low contrast, $t(15) = .3.01$, $p = .01$; $t(15) = 6.17$, $p < .001$; $t(15) = 11.75$, $p < .001$, respectively). To test the interaction between the mask and occlusion, the differential effect of the mask was compared between the low occlusion, high contrast condition and the moderate occlusion high contrast condition. The interaction between the mask and occlusion was indeed significant ($t(15) = -2.32$, $p = .03$).[1] The interaction between the mask and contrast was tested in the same way, except the difference was taken between the low occlusion, high contrast condition and the low occlusion, low contrast condition, and was also found to be significant ($t(15) = -5.73$, $p < .001$).

[1] Note that this comparison is equivalent to the *F*-test of the interaction between the mask and occlusion under high contrast alone. Hence, $t(15) = -2.32^2 = F(1,15) = 5.38$.

4.2 Discussion

The statistical analysis of the results indicated that there was a significant interaction between the mask and occlusion such that subjects' performance on the categorization task was differentially impaired when the moderately occluded stimulus was masked compared to a relatively unoccluded one. This finding is consistent with the basic predictions from the Leabra model described in O'Reilly et al. (in preparation) in which the bidirectional architecture of the Leabra model provided a selective robustness to occlusion compared to a strictly feedforward architecture. In the context of the present experiment, the Leabra model can be viewed as consistent with subjects' performance in the conditions that did not contain a mask, allowing recurrent processing to resolve the ambiguity in the moderately occluded stimuli. When these same conditions were masked, subjects' performance was more consistent with the feedforward models used in O'Reilly et al.'s simulations, suggesting that the mask prevented recurrent processing from assisting the recognition of the occluded stimulus.

These results, however, appeared to be subject to a floor effect, preventing the interpretation of the full interaction between the mask and occlusion across multiple levels of contrast. Thus, it is still possible that recurrent processing is restricted to a subset of the accuracy scale (i.e., the subset contained under the high contrast condition), opposed to a general process that resolves ambiguity across the full accuracy scale. To gain insight into this issue with interpretation, the Leabra model was used to explicitly simulate the results of the experiment. After finding a set of parameters that provided a good fit of subjects' data, the model was used to simulate the broader parameter space to determine whether the interaction between the mask and occlusion was a valid effect or an artifact of the accuracy scale.

Chapter 5

Modeling Simulations

The Leabra model used in the modeling simulations was identical to that described in O'Reilly et al. (in preparation), except with minor augmentations (described here) that constrained its recognition to the six categories of objects used in the experiment.

5.1 Procedure

Prior to simulating the results of the experiment, the Leabra model was pre-trained across images from all 100 categories from the *CU3D 100* dataset. This pre-training encompassed the entire dataset, opposed to only the six valid categories from the experiment, due to anecdotal evidence of the Leabra model exhibiting better performance when trained on images from a large number of categories, compared to fewer (presumably because doing so accounted for more variance in the space of possible inputs).

During the pre-training, there was no need to normalize the images for luminance across categories (as was the case in the experiment) since the model was unable to the make use of luminance information as a cue for category membership. The images were however converted from RGB to grayscale prior to being presented to the model. Additionally, images were presented to the model during both the pre-training and the subsequent 6-way categorization test with small variations in foveal position, scale, and planar rotation. These small variations, which were important for the model's ability to learn an invariant representation similar to that coded by IT neurons (Wallis & Rolls, 1997; Riesenhuber & Poggio, 1999), were within the range that subjects would be expected to experience due to head movements during the experiment.

The pre-training included all views of all except two exemplars from each of the 100 categories (14840 images total) and proceeded for 1000 epochs of 500 images each. The model's performance was evaluated using a reduced version of the 6-way categorization test used in the experiment, in which the model's ability to generalize the correct category of the twenty views of the two novel exemplars (240 images total) was recorded. The specific exemplars that were included in the pre-training and the categorization test sets (referred to here as training/testing splits) was randomized each time the pre-training was performed. A total of five training/testing splits were used.

Separate simulations were performed for each of the eight trial types used in the experiment – low or moderate occlusion, high or low contrast, and mask absent or mask present. Each of these simulation types used the same five training/testing splits and corresponding weights from the pre-training. In order to perform the 6-way categorization test using the weights learned from the pre-training, the model was augmented to restrict its response to the six valid categories from the experiment (*cannon, car, fish, gun, key,* and *trumpet*) by biasing their associated response units with additional excitatory activation (10% of their maximum activation).

During the categorization test, occlusion and contrast were manipulated in the same manner as described in the Experimental Methods chapter. Masks were also constructed in the same manner, except that pre-generating a large number of masks was unnecessary due to the model being prevented from learning during the 6-way categorization test. A total of 100 masks were pre-generated for use in the simulations.

On each trial of the categorization test, the object stimulus was degraded using the same levels of occlusion and contrast as during the experiment (for occlusion, low = 20%, moderate = 50%; for contrast, high = 100%, low = 25%). For simulations during which the mask was absent, the object stimulus was presented to the model's inputs and the model was iterated for 50 processing cycles after which the category associated with the most active output unit was recorded as the model's response. For simulations during which the mask was present, presentation of the object stimulus proceeded in the same way, but the model's inputs were subsequently remapped to a randomly chosen mask prior to the model reaching its 50th processing cycle. Any activation associated with the object stimulus that had been established prior to the remapping was preserved and allowed to interact with any new activation that was established as a result of processing the mask. The mask image remained presented to the model for the remainder of the 50 processing cycles, after which the model's response was recorded.

No assumptions were made about how the time integration represented by a single cycle of the model's processing mapped onto the passage of time in the physical world. Thus, to find the equivalent of presenting a mask with a latency of 100 ms, the processing cycle that the mask was mapped into the model's inputs was varied to find the best fit of subjects' data. Occlusion and contrast were held constant at their values used in the experiment during this fitting procedure.

5.2 Results and discussion

The simulations indicated that mapping the mask into the Leabra model's inputs at the 30th cycle of processing produced the best fit of subjects' data under high contrast (Figure 5.1a). Again, this simulated mask latency lacks theoretical importance since there is no relationship between the model's processing cycles and the passage of time in the real world. However, the results of the simulation indicated that the model could provide a good quantitative fit of the effect of masking in subjects' data. Specifically, the model produced the same interaction between the mask and occlusion such that the mask had a greater impairment on performance in the moderate level of occlusion.

The model was less successful in fitting subjects' data under low contrast (Figure 5.1b). Specifically, the model's recognition performance on trials during which the mask was absent was lower than that of subjects, regardless of the level of occlusion. One likely reason for this prediction error is due to the model failing to implement the full range of contrast operations performed during early processing in cortex (e.g., Enroth-Cugell & Robson, 1966). This inherent disconnect between the model and biology makes the mapping between the model's contrast response function and that

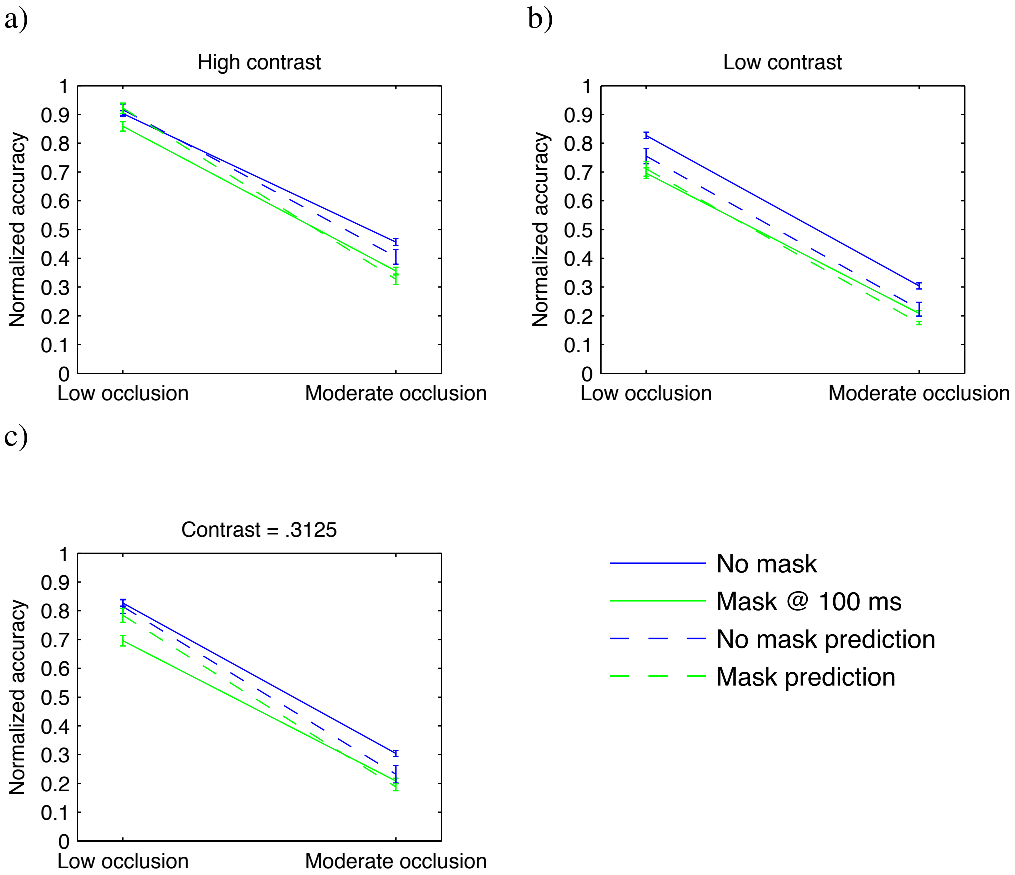

Figure 5.1: Experimental data were simulated by mapping the mask into the Leabra model's inputs at the 30th cycle of processing. Occlusion and contrast were held constant at the values used in the actual experiment. **a)** This fitting procedure successfully produced the interaction between the mask and occlusion observed under high contrast. **b)** The fitting procedure was less successful in fitting subjects' data under low contrast. Possible reasons for this prediction are discussed in the main text. **c)** To better account for subjects' data under low contrast, contrast was varied across a broader range of values while maintaining the simulated mask latency of 30 processing cycles. A contrast of 31.25% of the original range provided a reasonable fit of subjects' data. Error bars on the simulated data indicate the standard error of mean accuracy averaged across 5 random sets of training/testing splits.

of human subjects unknown. If this disconnect is the source of the prediction error, then allowing contrast to vary as a free parameter during the fitting procedure should produce a more accurate prediction of subjects' data. Additional simulations indicated that a contrast of 31.25% of the original range (with the simulated mask latency and levels of occlusion held at their values from the original simulations) provided a better fit of subjects' data under low contrast (Figure 5.1c).

To gain insight into the continuous nature of the interaction between the mask and occlusion as well as to generate predictions for future experiments, additional simulations were conducted for a single training/testing split that varied the cycle that the mask was mapped into the model's inputs, the level of occlusion, and the level of contrast across a broader range of their overall parameter space (Figures 5.2-5.3).

The first thing worth noting from this series of simulations is that the differential impairment of the mask at higher levels of occlusion exists across multiple levels of contrast (Figure 5.2a-d). This finding suggests that subjects' data from the experiment were indeed subject to a floor effect and that there exists a true interaction in low contrast. Interestingly, the model produces a qualitatively similar floor effect at the levels of occlusion and contrast used in the experiment with a short mask latency. However, most combinations of occlusion, contrast, and fast simulated mask latency give rise to an interaction between the mask and occlusion, providing a reasonable subset of the parameter to space from which to select values for future experiments.

Finally, an interaction between the mask and contrast was not observed, even across multiple levels of occlusion (Figure 5.3a-d). Instead, the effect of the mask and contrast reduction sum linearly to impair performance. This finding stands in opposition to subjects' actual data, which do indicate a significant interaction between the mask and contrast. Reasons for this discrepancy are discussed in further detail in the General Discussion chapter.

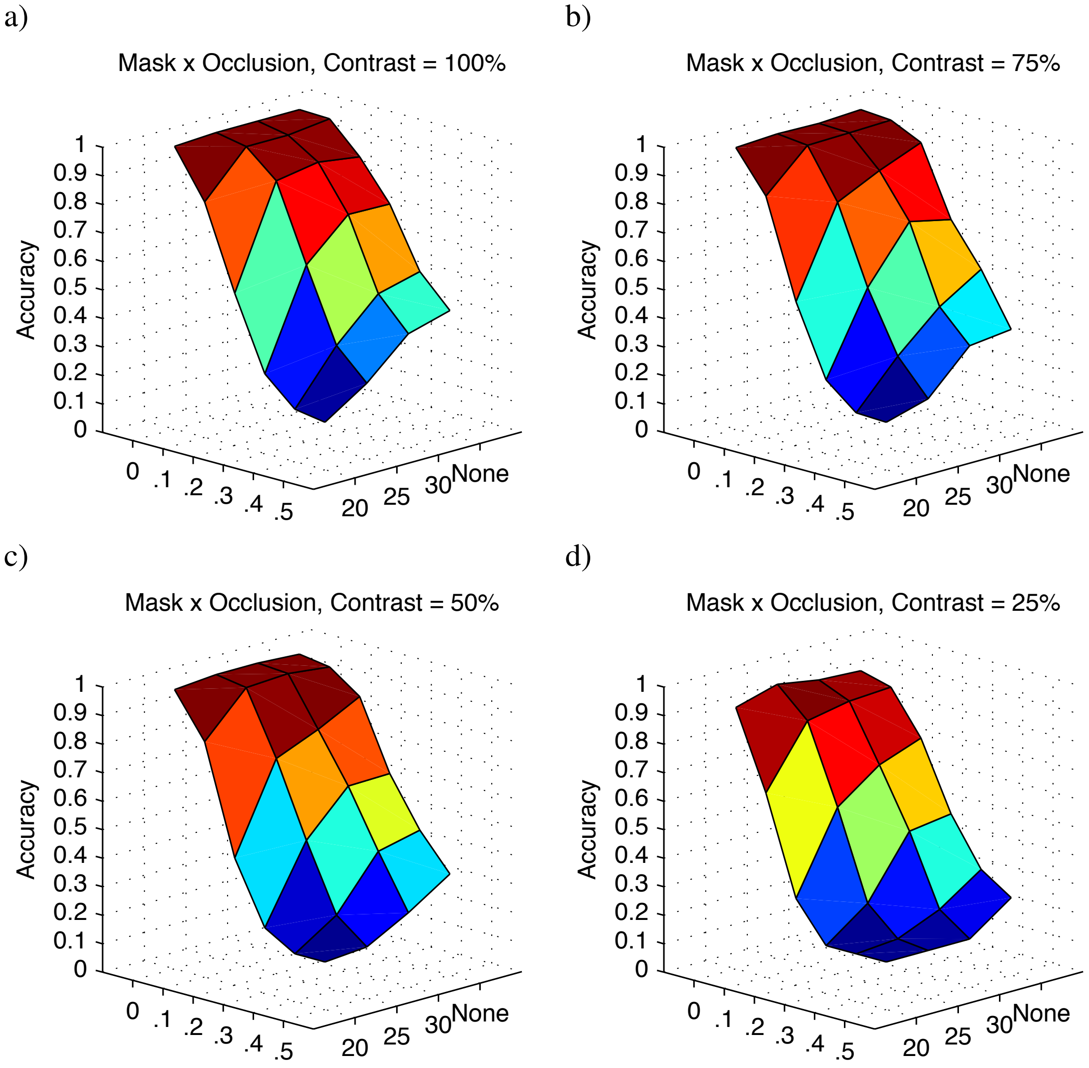

Figure 5.2: The cycle that the mask was mapped into the model's inputs, the level of occlusion, and the level of contrast were varied across a broader range of their overall parameter space. **a)-d)** Surface plots indicate that interaction between the simulated mask latency (right horizontal axis) and occlusion (left horizontal axis) exists across multiple levels of contrast. A floor effect, qualitatively similar that found in subjects' data, can be observed at the upper end of the occlusion spectrum under low contrast (25%) with relatively short simulated mask latencies. All data reflect accuracy on the categorization test from a single training/testing split.

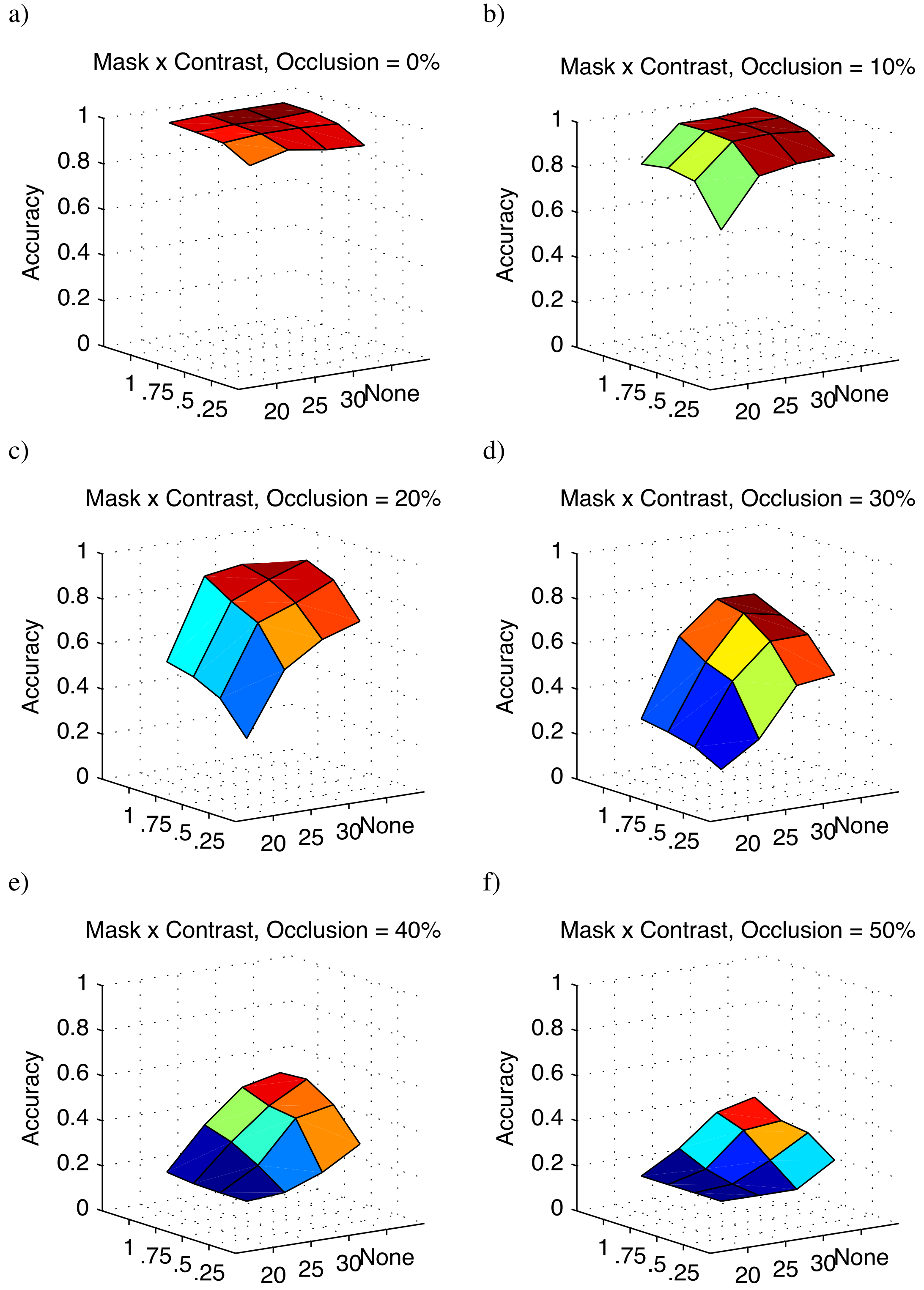

Figure 5.3: **a)-f)** Plotting the surface over the effect of contrast reduction (left horizontal axis) and the simulated mask latency (right horizontal axis) indicates that the model did not predict an interaction between the mask and contrast. All data reflect accuracy on the categorization test from a single training/testing split.

Chapter 6

General Discussion

Visual object recognition in cortex has traditionally been characterized as the net computational result of processing performed in the feedforward pathway of the ventral visual stream (Riesenhuber & Poggio, 1999; Fukushima, 1980; Wallis & Rolls, 1997; Serre et al., 2007; Masquelier & Thorpe, 2007; Riesenhuber & Poggio, 2000). Despite the predominance of this view, neuroanatomical data indicate that the majority of cortico-cortical connections in the ventral visual stream (as well as many other neural systems) are bidirectional (Felleman & Van Essen, 1991; Scannell et al., 1995; Sporns et al., 2007; Sporns & Zwi, 2004). O'Reilly et al. (in preparation) recently proposed a novel model of object recognition, the Leabra model of object recognition, that demonstrated how the bidirectional processing dynamics within recurrent circuits can resolve ambiguity in visual inputs during object object recognition such as when visual stimuli are occluded.

The present experiment used visual masking, which has been suggested to disrupt recurrent processing in cortex (Lamme & Roelfsema, 2000) to test the predictions of the Leabra model regarding the role of recurrent processing during object recognition. The results of the experiment indicated that the presence of the mask differentially impaired the recognition of visual stimuli that were moderately occluded compared to visual stimuli that were less occluded. This result, characterized by an interaction between the mask and occlusion, was present under high contrast in the experiment, but not under low contrast. The accompanying modeling simulations, which explicitly simulated the effect of masking on the Leabra model's ability to recognize objects, indicated that the lack of the interaction between the mask and occlusion under low contrast was due to a floor effect, extending the “true” interaction between the mask and occlusion across multiple levels of contrast. Together, these results suggest that recurrent processing is indeed beneficial to object recognition when inputs are occluded or otherwise underspecified.

Mechanistically, recurrent processing between hierarchically adjacent areas has been shown to strengthen the fidelity of the encoded representation. For example, figure-ground segmentation is normally characterized by high-level visual areas providing excitatory feedback to earlier visual areas. Thus, inactivating these high-level visual areas has been demonstrated to result in a weakened the response in earlier visual areas (Hupe et al., 1998; Hupe et al., 2001). A similar weakening dynamic has been observed during electrophysiological recordings of figure-ground segmentation when subjects are shown a visual mask (Fahrenfort et al., 2007). The modeling simulations in which the Leabra model was used to explicitly simulate the results of masking demonstrate the same strengthening and weakening dynamics of recurrent processing during object recognition.

During normal object recognition, the strength of the neural activation signal in cortex associated with a given object has been shown to be correlated with subjects’ ability to explicitly name the object’s category (Grill-Spector et al., 2000; Bar et al., 2001). Object recognition works much the same way in the Leabra model. Over the course of processing, the Leabra model’s overall activation pattern evolves toward toward its final pattern. The model is most successful at object recognition when this final activation pattern closely matches a stored pattern associated with the given object’s category, strongly activating the appropriate category name unit. When objects are relatively unoccluded and well-specified, the overall activation pattern rapidly converges on a stored pattern and activates the correct category name (Figure 6.1a). Occluding an object underspecifies it is an input, such that its associated activation pattern doesn’t initially resemble any stored patterns. However recurrent processing can ultimately recover the correct pattern that by strengthening the unreliable, weak representation over the course of processing, recovering the correct category name (Figure 6.1b). When an occluded object is masked, the model is often unable to recover the correct pattern because the representation associated with the mask biases the overall activation toward a different final pattern (one associated with the mask), effectively halting the ability of recurrent processing to recover the correct category name (Figure 6.1c). Recognition is spared when unoccluded objects are masked because the resulting representation is sufficiently specified such that the model converges, activating the correct category name, prior to the onset of the mask.

The dynamics demonstrated by the model are synonymous with computational principles described elsewhere in the literature such as attractor dynamics and mutual constraint satisfaction (Amit et al., 1994; Amit & Brunel, 1997; O'Reilly, 1998; O'Reilly & Munakata, 2000) and are consistent with theories of high-level vision that incorporate top-down processes (e.g., Ullman, 1995; Hochstein & Ahissar, 2002). Again, the use of top-down connections in the Leabra model, which are fundamentally excitatory and give rise to recurrent interactions, can be contrasted with that of other bidirectional models of object recognition that implement a recognition-by-synthesis approach to vision (Rao & Ballard, 1997; Hinton, 2010). This class of models uses inhibitory top-down connections to form a generative model of the input that effectively cancels with the actual bottom-up inputs, leaving a residual error that can be applied as a learning signal. Altogether, the Leabra model provides a mechanistic implementation of recurrent processing during object recognition as well as a novel demonstration of how masking disrupts this processing.

6.1 Open questions

One open question with the present research is concerned with why subjects' experimental data indicated an interaction between the mask and contrast, yet the Leabra model failed to predict such an interaction. One explanation for this discrepancy comes from the fact that the model does not capture the full range of contrast operations that occur during early visual processing. Specifically, the model uses a Difference of Gaussians (DoG) operation to approximate the contrast enhanced signal produced by retinal ganglion cells (Enroth-Cugell & Robson, 1966; Young, 1987). However, mammalian V1 likely implements additional constraints on top of this contrast-coded representation. For example, human contrast thresholds are highly dependent on the spatial frequency of the visual signal, leading to the proposition that early visual processing of contrast

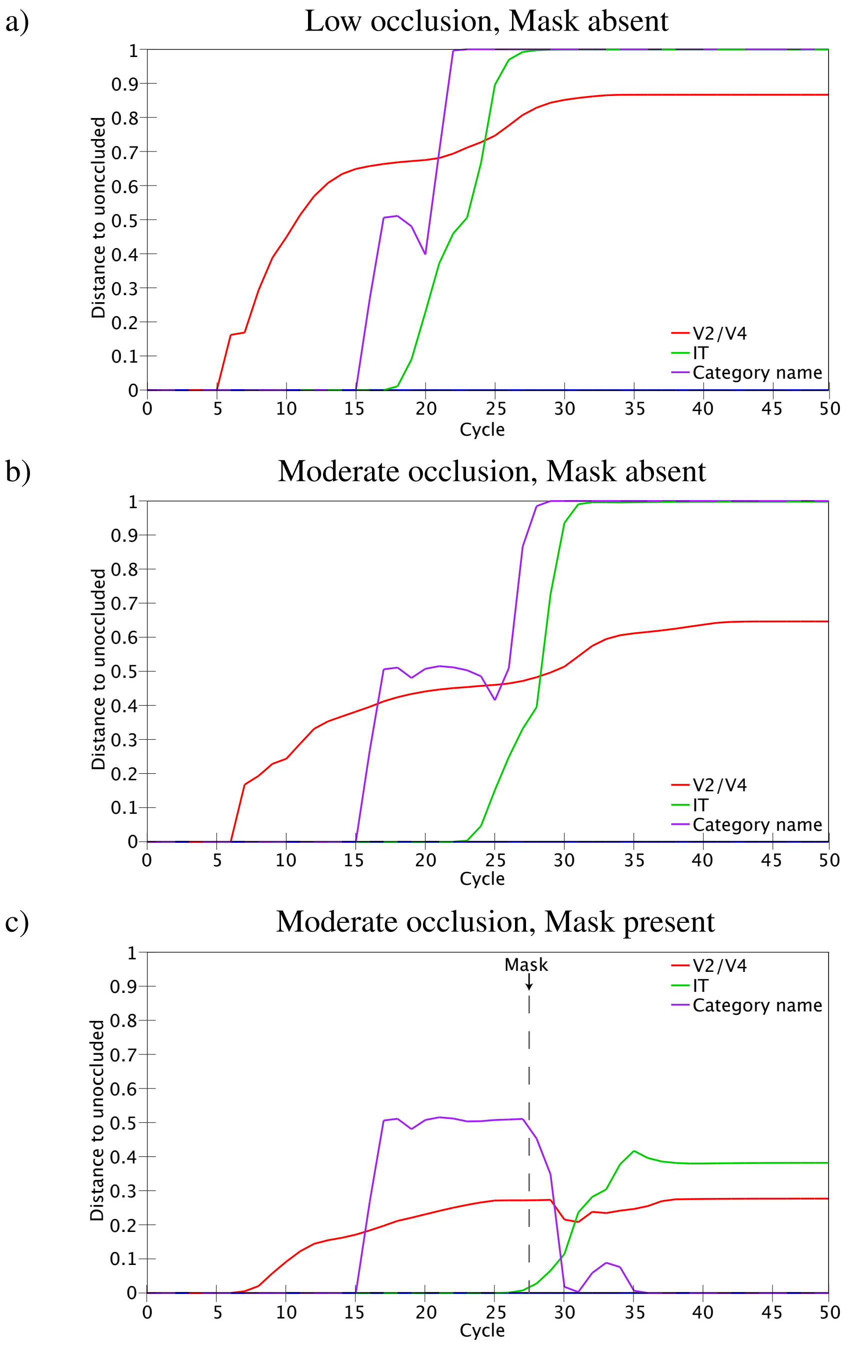

Figure 6.1: The Leabra model produces the full range of dynamics over the entire object recognition process, in addition to the net result of recognition. These plots were constructed by computing the distance of the activation pattern in each of the model's processing area's for an occluded stimulus to the activation pattern that would be associated with the identical, but unoccluded stimulus during each cycle of processing. **a)** When a stimulus is relatively unoccluded, the model rapidly recovers the category associated with the stimlus. **b)** When the stimulus is occluded, recurrent processing can gradually strengthen the weak representation over the course of processing, ultimately recovering the correct category of the stimulus. **c)** When an occluded stimulus is masked, recurrent processing is unable to recover its category because lower-level processing areas become increasingly susceptible to bias from the mask.

may operate in parallel channels organized by spatial frequency (Campbell & Robson, 1968). The model captures this to idea some extent, but only implements two such channels – relatively low and relatively high spatial frequencies. Furthermore, the model renormalizes the result of processing in these V1 channels. This normalization, while important for the mathematics of the Leabra equations (O'Reilly & Munakata, 2000), is likely a vast simplification of the complexity of the distributed, graded encoding performed by biological V1 neurons. It is difficult to determine whether these mathematical simplifications can account for the model's failure to produce an interaction between the mask and contrast without scaling up the complexity of the model up to match that of the real brain.

A different interpretation of subjects' interaction between the mask and contrast is that degrading an image by reducing its contrast has a qualitatively similar effect to degrading an image by occluding it. Under this view, degradations to what appear to be two independent stimulus dimensions actually affect a common, underlying psychological dimension (see Newell & Dunn, 2008). This psychological dimension, while intangible, can roughly be conceptualized as the perceived amount of ambiguity in the mental representation. To account for the pattern of interactions within this framework, the Leabra model would require a means of converting extrinsic stimulus values (pixel intensities) to the intrinsic psychological values associated with ambiguity. Presently, early visual processing and/or the learned weights of the model permit this conversion for degradations of occlusion (hence, the model's prediction of an interaction between the mask and occlusion) but not for degradations of contrast. It is unclear whether this conversion for contrast is impossible due to the aforementioned limitations of contrast operations in the model or if ambiguity resolution in the human brain involves a brain region outside of the ventral visual stream that monitors and resolves conflict, like the anterior cingulate cortex (Botvinick, Braver, Barch, Carter, & Cohen, 2001), to modulate recurrent processing.

Although the Leabra model can be augmented to address these questions and better account for subjects' data, it may also be possible to gain insight into the nature of recurrent processing and object recognition by using additional experimental methods chapter. For example, eventrelated components have been identified that appear to index psychological difficulty, opposed to a simple correlations with the visibility of a stimulus (Philiastides, Ratcliff, & Sajda, 2006; Philiastides & Sajda, 2007). Similarly, psychopharmacology has suggested that certain types of benzodiazapenes, such as lorazepam, interrupt only the visual-specific aspects of recurrent processing (Beckers, Wagemans, Boucart, & Giersch, 2001; van Loon, Scholte, & Lamme, 2010). These experimental methods provide a promising way in which to determine whether an interaction between the mask and occlusion is qualitatively different from an interaction between the mask and contrast and could help determine the set of conditions that is appropriate for studying recurrent processing.

6.2 Conclusion

Overall, the research presented here provides a suitable demonstration of recurrent processing during object recognition. The results of the experiment indicate that the recognition of objects that are degraded to underspecification is indeed a dynamic process that masking is capable of interrupting. The accompanying modeling simulations, in addition to partially accounting for the results of the experiment, further provide a mechanistic explanation of the dynamics of the recognition process.

Despite the issues in isolating the nature of the ambiguity that recurrent processing resolves, the cumulative results of the work presented here suggest that object recognition can be characterized by a number of interacting computations that evolve over time depending on extrinsic (and potentially intrinsic) conditions. This position is a significant departure from the standard model of object recognition that posits that a single series of feedforward computations is sufficient for explaining the capabilities of the brain when it comes to recognizing visual objects (Riesenhuber & Poggio, 1999; Wallis & Rolls, 1997; Freedman et al., 2003; Serre et al., 2007; Masquelier & Thorpe, 2007).

Interestingly, the results of the modeling simulations indicate that under normal viewing conditions (i.e., unoccluded, high contrast), the dynamics of the recognition process converge rapidly, suggesting that the bulk of the recognition process is subserved by the model's feedforward pathway of connectivity. Importantly however, the architecture of the Leabra model – and by extension, the brain – does not constrain processing to feedforward computations only. A bidirectional architecture appears necessary for resolving ambiguity in input signals and does so by strengthening the appropriate overall representation and constraining the recognition process. Furthermore, detailed accounts of the time course of object recognition estimate the onset of recurrent activation into the ventral visual stream to be within the bounds of the recognition process (Lamme & Roelfsema, 2000; Foxe & Simpson, 2002) in addition to any after-the-fact processes like feature-based attention. Ultimately, research that focuses on a tight coupling between biologically realistic computational modeling and experimental work on recurrent processing and object recognition will help illustrate the full dynamics of the object recognition process and the relative contributions of recurrent processing.

References

Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cerebral cortex, 7(3), 237–252.

Amit, D. J., Bruneland, N., & Tsodyks, M. V. (1994). Correlations of cortical hebbian reverberations: theory versus experiment. The Journal of neuroscience, 14(11), 6435–6445.

Bacon-Mace, N., Mace, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: backward masking and natural scene categorisation. Vision research, 45(11), 1459–1469.

Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B. R., & Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. Neuron, 29, 529–535.

Beckers, T., Wagemans, J., Boucart, M., & Giersch, A. (2001). Different effects of lorazepam and diazepam on perceptual integration. Vision research, 41(17).

Bogartz, R. (1976). On the meaning of statistical interactions. Journal of Experimental Child Psychology, 22(1), 178–183.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. Psychological Review, 108, 624–652.

Brainard, D., Pelli, D., & Robson, T. (2002). Display characterization. In J. Hornak (Ed.), Encyclopedia of imaging science and technology (pp. 172–188). Wiley.

Brainard, D. H. (1997). The psychophysics toolbox. Spatial vision, 10(4), 433–436.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. Journal of neurophysiology, 46(2), 369–384.

Campbell, F. W., & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. The Journal of physiology, 197(3), 551–566.

Craft, E., Schutze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. Journal of Neurophysiology, 97(6), 4310–4326.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller, & J. Grafman (Eds.), Handbook of neurophysiology, vol. 2 (Chap. 14, pp. 267–299). Amsterdam: Elsevier.

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. The Journal of physiology, 187(3), 517–552.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. Journal of cognitive neuroscience, 19(9), 1488–1497.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex, 1, 1–47.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from v1 to frontal cortex in humans. a framework for defining ”early” visual processing. Experimental Brain Research, 142, 139–150.

Francis, G. (1997). Cortical dynamics of lateral inhibition: metacontrast masking. Psychological review, 104, 572.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science, 291, 312–316.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. The Journal of neuroscience, 23, 5235–5246.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36, 193–202.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. Nature neuroscience, 3(8), 837–843.

Haynes, J.-D., Driver, J., & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between v1 and fusiform cortex. Neuron, 46(5), 811–821.

Hinton, G. E. (2010). Learning to represent visual input. Philosophical transactions of the Royal Society of London, 365, 177–184.

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. Neuron, 36(5), 791–804.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences of the United States of America, 79, 2554–2558.

Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology, 195(1), 215–243.

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.

Hupe, J. M., James, A. C., Girard, P., & Bullier, J. (2001). Response modulations by static texture surround in area v1 of the macaque monkey do not depend on feedback connections from v2. Journal of Neurophysiology, 85, 146–163.

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by v1 v2 and v3 neurons. Nature, 394(6695), 784–787.

Lamme, V., & Roelfsema, P. (2000). The distinct modes of vision offered by feedforward and recurrent processing. Trends in Neurosciences, 23(11), 571–579.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Current biology, 5(5), 552–563.

Macknik, S. L., & Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. Nature neuroscience, 1(2), 144–149.

Masquelier, T., & Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. PLoS computational biology, 3(2), e31.

Miller, E. K., Nieder, A., Freedman, D. J., & Wallis, J. D. (2003). Neural correlates of categories and concepts. Current opinion in neurobiology, 13, 198–203.

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: testing psychological models using state-trace analysis. Trends in cognitive sciences, 12(8).

O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. Trends in Cognitive Sciences, 2(11), 455–462.

O'Reilly, R. C., & Munakata, Y. (2000). Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. Cambridge, MA: The MIT Press.

O'Reilly, R. C., Wyatte, D., Herd, S. A., Mingus, B., & Jilk, D. J. (in preparation). Bidirectional biological object recognition.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. Spatial vision, 10(4), 437–442.

Philiastides, M., & Sajda, P. (2007). Eeg-informed fmri reveals spatiotemporal characteristics of perceptual decision making. Journal of Neuroscience, 27.

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. The Journal of neuroscience, 26(35), 8965–8975.

Rao, R. P. N., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Computation, 9, 721–763.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature Neuroscience, 3, 1199–1204.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. Nature neuroscience, 3 Suppl, 1199–1204.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. Current Opinion in Neurobiology, 12, 162–168.

Roland, P. (2010). Six principles of visual cortical dynamics. Frontiers in Systems Neuroscience, 4(28), 1–21.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proceedings: Biological Sciences, 257(1348), 9–15.

Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: information analysis. Journal of cognitive neuroscience, 11(3).

Scannell, J., Blakemore, C., & Young, M. P. (1995). Analysis of connectivity in the cat cerebral cortex. Journal of Neuroscience, 15, 1463–1483.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proceedings of the National Academy of Sciences of the United States of America, 104(15), 6424–6429.

Sporns, O., Honey, C. J., & Kotter, R. (2007). Identification and classification of hubs in brain networks. PloS one, 2(10), 1–14.

Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. Neuroinformatics, 2(2), 145–162.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annual review of neuroscience, 19, 109–139.

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. Cerebral cortex, 5(1), 1–11.

Ungerleider, L. G., & Haxby, J. V. (1994). “What” and “Where” in the human brain. Current Opinion in Neurobiology, 4, 157–165.

van Loon, A. M., Scholte, H. S., & Lamme, V. A. F. (2010). Lorazepam reduces stimulus visibility by impairing recurrent processing in visual cortex. VSS 2010: The 10th annual meeting of the Vision Sciences Society. Naples, FL.

Vecera, S. P., & O’Reilly, R. C. (1998). Figure-ground organization and object recognition processes: an interactive account. Journal of experimental psychology. Human perception and performance, 24, 441–462.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. Progress in neurobiology, 51(2), 167–194.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the shine toolbox. Behavior research methods, 42(3), 671–684.

Young, R. A. (1987). The gaussian derivative model for spatial vision: I. retinal mechanisms. Spatial vision, 2(4), 273–293.