AFRL-RI-RS-TR-2008-35
Final Technical Report
February 2008

HIGH-LEVEL VISION: TOP-DOWN PROCESSING IN NEURALLY INSPIRED ARCHITECTURES

Harvard University

Sponsored by
Defense Advanced Research Projects Agency
DARPA Order No. V026/00

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

NOTICE AND SIGNATURE PAGE

AFRL-RI-RS-TR-2008-35 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/	/s/
THOMAS E. RENZ Work Unit Manager	JAMES A. COLLINS, Deputy Chief Advanced Computing Division Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
FEB 08	Final	Sep 05 – Aug 07

4. TITLE AND SUBTITLE

HIGH-LEVEL VISION: TOP-DOWN PROCESSING IN NEURALLY INSPIRED ARCHITECTURE

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA8750-05-2-0270

5c. PROGRAM ELEMENT NUMBER
61101E

6. AUTHOR(S)

Stephen M. Kosslyn, Bruce Draper, Giorgio Ganis and Mark Knobel

5d. PROJECT NUMBER
BICA

5e. TASK NUMBER
00

5f. WORK UNIT NUMBER
01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Harvard University
1350 Mass Ave
Cambridge MA 02138

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Advanced Research Projects Agency
3701 North Fairfax Dr
Arlington VA 22203-1714

AFRL/RITC
525 Brooks Rd
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TR-2008-35

12. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA# WPAFB 08-0342

13. SUPPLEMENTARY NOTES

14. ABSTRACT

An exhaustive literature review of computationally relevant studies of high-level vision and mental imagery was conducted, and a qualitative theory of the processing subsystems in the brain and their interactions during visual object identification are summarized herein. A computational model that embodies these ideas was built with the intention that it could be "damaged" in various ways in order to observe its behavior while performing visual tasks that are analogous to those performed by the brain-damaged patients; this model is described in detail in this report. A second model of early/intermediate vision, which is intended to supplement and complement the work on late visual processing, has also been developed, and is described. Progress testing the models is reported. Additional deliverables include a binder with the abstracts of all the relevant literature (which can serve as a guide for further research) and an annotated copy of the software code for the IMPER model.

15. SUBJECT TERMS
Computational modeling of cognitive processes, biologically inspired architecture, visual processing, mental imagery, brain damage

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Tom Renz
U	U	U	UL	50	19b. TELEPHONE NUMBER (*Include area code*) N/A

Table of Contents

List of Figures

Summary

We delivered an exhaustive literature review of computationally relevant studies of high-level vision in the brain. We not only reviewed the major findings in the summary of our theory, but also delivered a binder that contains abstracts from the relevant research literature. This binder can serve as a guide for further research. 2) We conducted an analysis of the problems to be solved in each of the major processing phases in the brain during visual object identification (to further specify the nature of the processing subsystems that work together to identify objects and provide the information necessary for reaching and navigation). We focused on specifying the type of information that is sent, and on the circumstances in which specific contents (i.e., parameter values) must be specified. We have addressed this goal in two ways. First, we developed a qualitative theory of the processing subsystems and their interactions, which is summarized below. Second, we have built a computational model that embodies these ideas. In building this model, we focused on aspects of the empirical literature that—at first blush--seemed *inconsistent* with our theory. Because this literature was exclusively in the realm of dissociations that occur following brain damage, we built the model so that we could "damage" it in various ways and observe its behavior in performing tasks that are analogous to those performed by the brain-damaged patients. A report of the model itself is provided after our qualitative summary. We also report work on early/intermediate vision, which is intended to supplement and complement our work on late visual processing. 3) We provide a written description of the systems we developed, which specifies in detail how the computer models were implemented. 4) We provide a description of our progress in testing the model. We also provide a summary of the progress made in the early/intermediate vision system. 5) Finally, we provide a commented version of the code itself, in case other groups want to build on what we started. As promised, the code was written so that it can be easily executed on most Unix-based machines.

1. Introduction: Overview of Objectives and Progress

In this report we summarize how we have addressed each of the original objectives of the proposal, as noted in what follows:

1) We intended to conduct an exhaustive literature review of computationally relevant studies of high-level vision in the brain. We have done so. We not only review the major findings in the summary of our theory (see below), but also deliver a binder that contains abstracts from the relevant research literature. This binder can serve as a guide for further research.

2) We planned to conduct an analysis of the problems to be solved in each of the major processing phases in the brain during visual object identification (and to further specify the nature of the processing subsystems that work together to identify objects and provide the information necessary for reaching and navigation). We intended to focus on specifying the type of information that is sent, and on the circumstances in which specific contents (i.e., parameter values) must be specified. We have addressed this goal in two ways. First, we developed a qualitative theory of the processing subsystems and their interactions, which is summarized below. Second, we have built a computational model that embodies these ideas. In building this model, we first reviewed all of the literature that—at first blush--seemed *inconsistent* with our theory. Because this literature was exclusively in the realm of dissociations that occur following brain damage, we built the model so that we could "damage" it in various ways and observe its behavior in performing tasks that are analogous to those performed by the brain-damaged patients. A report of the model itself is provided after our qualitative summary.

3) We agreed to provide a written description of the system to be developed, specifying in detail how the computer model was implemented. This description is provided after the summary of the key ideas that are incorporated in the theory.

4) We agreed to provide a description of how the system operates in given circumstances. A summary of our progress in achieving this goal is provided after our overview of the theory.

5) We also agreed to provide a commented version of the code itself, in case other groups want to build on what we start. As promised, the code was written so that it can be easily executed on most Unix-based machines.

6) Finally, we said that we would indicate the next steps that should be taken to continue to develop this research program. The next steps would be: a) To discover whether the model can be scaled up. It currently operates only on a very small number of images, and we would like to expand that number dramatically. b) To provide a workable interface with low-level vision processing. We currently start with a parsed image, but would prefer to have a fully automated system. Dr. Draper has been working on this aspect of the project, and we would like to see his work fully integrated with the work on late vision that we summarize below. We also provide a summary of the work he has conducted, which follows the summary of the model that implements our theory. c) We would like to develop quantitative predictions about the effects of specific types of brain damage on behavior. The model is close to a state where this would be possible, but additional work would be required to refine the relevant aspects of the model.

2. Methods, Assumptions, and Procedures

2.1 Summary of the theory of visual processing in the brain.

Our project began with the observation that no animal could survive for long without perception. We must perceive the world not only in order to find food, shelter and mates, but also to avoid predators. Perception will fail if an animal does not register what is actually in the world. However, this simple observation does not imply that all processing during perception is "bottom up"—driven purely by the sensory input. Rather, bottom-up processing can be usefully supplemented by using stored information, engaging in processing that is "top down"—driven by stored knowledge, goals or expectations. In this

project we have explored the nature of top-down processing and its intimate dance with bottom-up processing. We begin by considering basic facts about the primate visual system, and then consider a theory of its functional organization, followed by novel proposals regarding the nature of different sorts of top-down processing.

The Structure of Visual Processing in the Brain

An enormous amount has been learned about visual processing by studying animal models. In particular, the macaque monkey has very similar visual abilities to those of humans, and the anatomy of its visual system appears very similar to ours. Studies of the monkey brain have revealed key aspects of the organization of the visual system, namely its hierarchical structure and the reciprocal nature of most connections between different visual areas of the brain. We briefly review key aspects of both characteristics of the brain below.

Hierarchical organization

Over the last several decades, researchers have provided much evidence that the primate visual system is organized hierarchically. In the early 60s and 70s, Hubel and Wiesel's electrophysiological findings, first in cats and then in non-human primates, strongly suggested a hierarchical relationship among early areas in the visual system; this inference was based on the increasing size and complexity of the receptive fields as one goes from striate cortex to areas farther along in the processing stream (Hubel & Wiesel, 1962, 1965, 1968, 1974). The earliest areas of the visual system are organized topographically; space on cortex represents space in the world, much as space on the retina represents space in the world (Felleman & Van Essen, 1991; Fox et al., 1986; Heeger, 1999; Sereno et al., 1995; Tootell et al., 1998; Van Essen et al., 2001). The higher-level areas are not organized topographically, but often represent information using population codes (e.g., Fujita, Tanaka, Ito, & Cheng, 1992; Miyashita & Chang, 1988; Tanaka, Saito, Fukada, & Moriya, 1991). In such codes, different neurons respond to complex visual properties, and shape is coded by the specific combination of which neurons are activated.

The work by Felleman and Van Essen (1991) charted the hierarchical organization of the entire visual system. They compiled a matrix of known anatomical connections among areas, and showed that the pattern of connectivity could best be accounted for by a hierarchical structure with multiple parallel streams. The striate cortex was at the bottom of the entire hierarchy, and the inferotemporal (area TE) and parahippocampal (areas TH and TF) cortex were at the top of the *ventral stream* (which is specialized for object vision, registering properties such as shape and color; (Desimone & Ungerleider, 1989).

The big picture of cortical organization provided by Felleman and Van Essen has been generally confirmed by computational analyses of the same dataset (Hilgetag, O'Neill, & Young, 1996), as well as by additional empirical approaches, such as those based on measuring the proportion of projecting supragranular layer neurons labeled by a suitable retrograde tracer (Vezoli et al., 2004). This neuroanatomical picture of a hierarchically organized visual system was also confirmed by data from single unit recording studies of higher-level visual areas. For instance, area TE in the inferotemporal cortex has been shown to contain neurons with extremely large receptive fields (often encompassing the entire visual field), which are tuned to complex combinations of visual features (such as combination of shape fragments and textures); in contrast, neurons in lower-level areas, such as V4, have smaller receptive fields and are tuned to simpler feature combinations (Tanaka, 1996).

Connections among areas

A considerable amount is now known about the connections among visual brain areas, and the evidence suggests that different connections are used in bottom-up and top-down processing.

Feed-forward connections and bottom-up processing. A set of contiguous neurons in area V1 have contiguous receptive fields (i.e., regions of space in which they will respond to stimuli). A set of contiguous neurons in area V1 in turn projects to a single neuron in area V2, and this neuron has a larger receptive field than any of those neurons that feed into it. This many : 1 mapping continues up the hierarchy until the receptive fields become so large that the areas are no longer topographically organized.

The neuroanatomical findings and the properties of the receptive fields have given rise to numerous models that emphasize the feed-forward nature of the ventral stream (Fukushima, 1988; Riesenhuber & Poggio, 1999; VanRullen, Delorme, & Thorpe, 2001; Wallis & Rolls, 1997). Electrophysiological findings that document the fast onset of neural responses to visual stimuli at all levels in the ventral stream (i.e., the mean latency of neurons in area TE at the highest level of the hierarchy is just over 100 ms after stimulus onset) provided additional impetus for these models (Lamme & Roelfsema, 2000). In these models, objects are identified during a feed-forward pass throughout the ventral stream hierarchy, with increasingly complex information being extracted at higher levels in the system. For instance, in Riesenhuber and Poggio's model of the ventral stream, the units farther upstream (from area V1 to TE) are tuned to increasingly complex features, all the way to units that are tuned to specific views of objects.

Feedback connections and top-down processing. Crucially for the topic at hand, and consistent with the connectivity pattern reported in earlier work by Rockland and Pandya (1979) and others, Felleman and Van Essen described not only feed-forward connections among areas in the primate visual system but also widespread feedback connections. They found a striking regularity in the pattern of laminar origin of feed-forward and feedback connections: whereas feed-forward connections originate in the supragranular layers (often layer III) and terminate in layer IV in the target area, feedback connections originate from neurons in layer VI and IIIA of the projecting area and end in layer I of the target area. Indeed, numerous other anatomical studies in non-human primates have confirmed that there are massive feedback connections at many levels in the visual system, including from areas that are not traditionally considered visual areas (Barone, Batardiere, Knoblauch, & Kennedy, 2000; Budd, 1998; Clavagnier, Falchier, & Kennedy, 2004; Rockland & Pandya, 1979; Salin & Bullier, 1995). For instance, area V1 has been shown to receive direct feedback connections from many extrastriate regions (including V2, V3, V4, TEO, TE), as well as from non-visual areas, including the frontal eye fields, area 36, areas TH/TF, STP and even auditory cortex.

The feedback connections are not simply the inverse of feed-forward connections. Whereas the feed-forward connections display a lovely many-to-one mapping as they ascend the hierarchy, there is nothing of the sort for the feedback connections. Instead, the feedback connections do not appear to be precisely targeted, but rather often appear to meander (e.g., Budd, 1998). Evidently, the feedback connections are not simply "replaying" information sent downstream.

Neuroanatomically inspired models of top-down processing. A class of models of object vision has incorporated the finding of feedback connections in the visual system (Grossberg & Mingolla, 1985; Li, 1998; Mumford, 1992; Ullman, 1989, 1995). Generally, these models assume that feedback connections provide a mechanism by which top-down processing can occur, allowing relatively abstract information stored in higher-level visual areas to influence and constrain processing in lower-level visual areas. To illustrate the basic idea of why top-down processing is needed, researchers have created binarized photographs. In such photos, grayscale pixels are replaced with white if their brightness value is above a chosen threshold, or replaced with black if it is below this value. Because binarized images are highly degraded, pure bottom-up processes typically cannot organize them correctly into their constituent parts, and often one needs to use previously acquired knowledge about objects to identify the objects in them (see Figure 1).

Figure 1. *This binarized picture illustrates the problems encountered by purely bottom-up approaches to vision. It is very difficult to parse correctly the fox at the center of the picture using purely bottom-up processing. Using top-down processing to exploit constraints imposed by knowledge of the shape of foxes makes the task much easier.*

The models just mentioned rest on algorithms that allow an interplay between stored information and on-line input. For instance, Mumford posits that higher-level visual areas try to find the best fit with the information they receive from lower level visual areas by using the more abstract knowledge they store (e.g., a representation of a shape). The feedback connections allow higher-level visual areas to reconstruct the visual input in lower-level visual areas, based on such best fit. The mismatch between the reconstructed visual input and the original input in lower-level visual areas (i.e., information not explained by the current fit in higher-level areas) is then sent forward, which can trigger another top-down processing cycle.

This class of models of top-down processing in the ventral stream has typically ignored the role of areas outside the ventral stream, or has only postulated unspecified extra-visual inputs. However, many non-visual areas in the frontal and parietal lobe are connected to areas in the ventral stream (e.g., Petrides, 2005). Another class of models, in contrast, focuses on top-down influences that these non-visual areas exert on areas in the ventral stream. For instance, the model of prefrontal function by Miller and Cohen (2001) has focused on the role of the prefrontal cortex in biasing processing in areas in the ventral stream.

Traditionally, the different classes of models have been pursued independently, although some of the terminology has overlapped. Unfortunately, the term "top-down processing" in vision has been used loosely in the neuroscientific literature to refer to a disparate range of phenomena. For instance, it has been used in the context of the neural effects of visual attention (Hopfinger, Buonocore, & Mangun, 2000), memory retrieval (Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999), in the context of phenomena such as illusory contours (Halgren, Mendola, Chong, & Dale, 2003), and so on.

In the remainder of this report we develop explicit distinctions between different types of visual top-down processes; these distinctions are cast within the context of a broad theory of the visual system that incorporates both bottom-up and top-down processes (Kosslyn, 1994) as well as the role of non-visual areas. Our aim is to make explicit some of the assumptions regarding top-down processes that are implicit in the literature, and propose a first-order taxonomy, rather than to provide an exhaustive review of the top-down processing literature. In the following section we briefly summarize our functional theory of the visual system and its operations during visual object identification, relying on the background already provided, and then we proceed to describe how different types of top-down processing may operate within this system.

A Theory of the Functional Organization of Late Visual Processing in the Primate Brain

We propose that there are two general kinds of visual processes, "early" and "late." Early visual processes rely on information coming from the eyes, whereas late processes rely on information stored in memory to direct processing. We must distinguish between early and late processes and the specific brain areas involved in vision: late processes can occur even in areas that are involved in the first stages of bottom-up processing (Lamme & Roelfsema, 2000). Low-level visual areas are involved in both early (bottom-up) and late (top-down processing).

Vision, and more specifically object identification, is not a unitary and undifferentiated process. Indeed, similarly to memory operations such as encoding and recall, which are carried out by many subprocesses (Schacter, 1996; Squire, 1987), object identification is carried out by numerous subprocesses (for example, those involved in figure-ground segregation, in shifting attention, in matching input to stored information)). Our theory posits a specific set of component visual processes, with an emphasis on those involved in late visual processing; we call these components *processing subsystems*. A processing subsystem receives input, transforms it in a specific way, and produces a specific type of output; this output in turn serves as input to other subsystems.

Figure 2 illustrates the most recent version of our theory of processing subsystems. . Although this diagram appears to imply sequentiality, the theory does not in fact assume that each processing subsystem finishes before sending output to the next. Rather, the theory posits that all processes are running simultaneously and asynchronously, and that partial results are continually being propagated through the system. Moreover, we assume that what shifts over time is how intensively a given process is engaged. Thus, the theory posits processing subsystems that operate in cascade, often operate on partial input, and send new outputs to other subsystems before they have completed processing (Kosslyn, 1994).

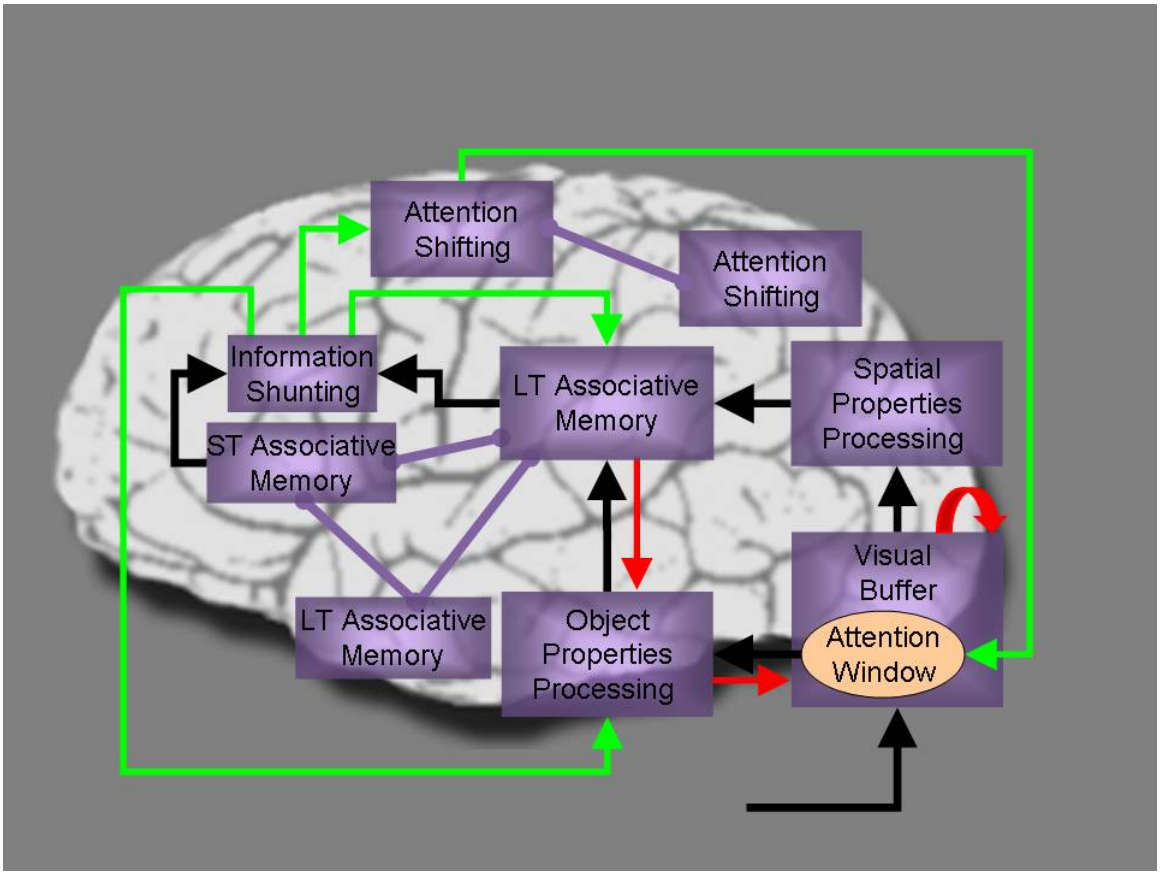

Figure 2. *The functional architecture of the visual system (subcortical structures are not shown for simplicity). Note that each box represents a structure or process (e.g., the Visual Buffer) that is implemented in multiple areas (and often can itself be decomposed into more specialized processes, not discussed here). The Associative Memory subsystem is divided into long-term (LT) and short-term (ST), but this distinction turned out not to be useful in the tasks we implemented in our model. Some subsystems (Associative Memory and Attention Shifting) are implemented by spatially distant brain regions. These subsystems are connected by arrowless lines; for simplicity, inputs and outputs from these subsystems are only indicated for one of them. Connections that implement reflexive top-down processing are indicated by red arrows whereas those that implement strategic top-down processing are indicated in green. Note that the reflexive connections can also be engaged by strategic top-down processing (via inputs from the information shunting subsystem). Visual input from the lateral geniculate enters the visual buffer via the black arrow at the bottom.*

Processing subsystems used in visual object identification

The process of object identification can be conceptualized as the search for a satisfactory match between the input and stored memories. The specific set of subsystems involved, as well as their timecourse of engagement, depends in part on the properties of the incoming visual stimulus. In all cases, however, the same architecture is used – the same types of representations and the same processes are used, but sometimes in different orders or more or less intensively. The first step in describing our theory is to summarize the processing subsystems, at the most general level, as follows.

Visual Buffer. In primates, vision is carried out by a multitude of cortical areas, at least 32 in monkeys (Felleman & Van Essen, 1991) and probably even more in humans (Sereno & Tootell, 2005). Many of these areas, including V1 and V2, are organized topographically. That is, these areas use space on cortex to represent space in the world. The specific pattern of activation in these areas reflects the geometry of the planar projection of a stimulus; in addition, focal damage to these areas causes scotomas, that is, blind spots at the spatial location represented by the damaged cortex. In our theory, the subset of topographically organized areas in the occipital lobe implement what we refer to as the *visual buffer* .

However, even if we think of the set of these topographically organized areas as a single functional entity, these areas are hierarchically organized themselves, and have somewhat different functional properties. We stress that although these areas are at the bottom of the visual hierarchy, and thus carry out bottom-up processes necessary for vision, they are also affected by top-down processes originating both within other portions of the visual buffer itself and from areas outside this structure.

Attention Window. Not all the information in the visual buffer can be fully processed. A subset of the information in the visual buffer is selected by an *attention window*, based on location, feature or object of interest, for further processing (Brefczynski & DeYoe, 1999; Cave & Kosslyn, 1989; Posner & Petersen, 1990; Treisman & Gelade, 1980). There is good evidence that this attention window can be covertly shifted (Beauchamp, Petit, Ellmore, Ingeholm, & Haxby, 2001; Corbetta & Shulman, 1998; Posner, Snyder, & Davidson, 1980) and it can also be split—at least in some specific circumstances--to include non-adjacent regions in the visual space (McMains & Somers, 2004).

Object-Properties-Processing. As discussed earlier, many connections run from the topographically organized areas of the occipital lobe to other areas of the brain, giving rise to parallel processing streams. The ventral stream runs from the occipital lobe to the inferior temporal lobe (Desimone & Ungerleider, 1989; Haxby et al., 1991; Kosslyn, 1994; Mishkin, Ungerleider, & Macko, 1983; Ungerleider & Mishkin, 1982), where visual memories are stored (e.g., Fujita, Tanaka, Ito, & Cheng, 1992; Tanaka, Saito, Fukada, & Moriya, 1991). The visual memories are stored, at least in the monkey brain, using a *population code* by which nearby cortical columns do not store information about nearby points in two-dimensional space, but information about nearby points in feature space (e.g., Fujita et al., 1992; Miyashita & Chang, 1988; Tanaka et al., 1991). These areas, implementing what we will refer to as the *object-properties-processing subsystem,* not only store information about shape and shape-related properties of objects and scenes, such as color and texture, but also match input to such stored information. Visual *recognition* of an object occurs when the content of the attention window matches a stored visual representation in this system.

Spatial-Properties-Processing. In addition to being able to determine the identity of objects we can also determine their location in space. Our visual system accomplishes this by allocating different resources to extract and process information about properties that are inherent to objects (such as their shape or color) versus information about their spatial properties (such as their size and location). The object-properties-processing subsystem essentially ignores spatial information, and produces—position-invariant object recognition (Gross & Mishkin, 1977; Rueckl, Cave, & Kosslyn, 1989). In contrast, spatial processing captures the very information discarded during object processing. Spatial processing is accomplished by the dorsal stream, a pathway that runs from the occipital lobe to the posterior parietal lobe (e.g., Andersen, Essick, & Siegel, 1985; Haxby et al., 1991; Kosslyn, Thompson, Gitelman, & Alpert, 1998; Ungerleider & Mishkin, 1982). In our theory, these posterior parietal regions embody what we refer to as the *spatial-properties-processing* subsystem.

According to our theory, the spatial-properties-processing subsystem not only registers spatial properties, but also constructs *object maps,* which indicate the locations of objects or parts of objects in space (cf., Mesulam, 1990). The cortex that implements at least part of the spatial-properties-processing subsystem is topographically organized, and thus at least some of the representations used in this subsystem depict the locations of objects in space (Sereno, Pitzalis, & Martinez, 2001).

Associative Memories. Outputs from both the object-properties-processing and spatial-properties-processing subsystems converge on associative memories. Associative memories specify links among representations. Our theory posits two classes of associative memories. On the one hand, *short-term associative memory* structures maintain information on-line about which objects are in specific locations (Rao, Rainer, & Miller, 1997; Wilson, Scalaidhe, & Goldman-Rakic, 1993). These memory structures are implemented in the dorsolateral prefrontal regions. On the other hand, *long-term associative memory* structures store more enduring associations among stored categories, characteristics, situations, and events. If the outputs from the object-properties-processing and spatial-properties-processing subsystems match a stored representation in long-term associative memory, the information associated with it is accessed, leading to object *identification*. For instance, if the shape matches that of a cat, one can access information that it is a mammal, likes to drink milk, and sometimes sleeps much of the day. If no good match is found, the best-matching representation is used as a hypothesis of what the viewed object might be (we will discuss this process in detail shortly). Long-term associative memory is implemented in Wernicke's area, the angular gyrus, classic "association cortex" (e.g., Area 19, Kosslyn, Thompson, & Alpert, 1995), and parts of the anterior temporal lobes (e.g. Chan et al., 2001).

Information Shunting. In our theory, when the match of the input to representations in long-term associative memory is poor, then representations of distinctive visual parts and attributes of the best-matching object are retrieved and used by an information shunting subsystem (cf. Gregory, 1970; Neisser, 1967; Neisser, 1976) to guide top-down search. Thus, by means of this process, the visual system actively seeks information to test hypotheses about the visual input. The information shunting subsystem operates in two related ways: First, it sends information to other subsystems, enabling them to shift the focus of attention to the likely location of distinctive parts or attributes. Second, simultaneously, the information shunting subsystem *primes* representations of these parts and attributes in the object-properties-processing subsystem (cf. Kosslyn, 1994; McAuliffe & Knowlton, 2000; McDermott & Roediger, 1994), which facilitates the ease of encoding these representations. The information shunting subsystem is implemented by one or more parts of dorsolateral prefrontal cortex (DLPFC) (e.g., see Damasio, 1985; Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999; Luria, 1980; Petrides, 2005; Posner & Petersen, 1990). However, the DLPFC is a very large region, and so it is unlikely that implementing the information shunting subsystem is its only function; moreover, different regions of DLPFC in principle may implement specialized components of the information shunting subsystem, with each operating only on specific types of information (e.g., location versus shape).

Attention Shifting. Shifting the focus of attention to a new location or to a new attribute involves a complex subsystem, which is implemented in many parts of the brain (including the superior parietal lobes, frontal eye fields, superior colliculus, thalamus and anterior cingulate (see Corbetta, 1993; Corbetta & Shulman, 1998; LaBerge & Buchsbaum, 1990; Mesulam, 1981; Posner & Petersen, 1990). The attention shifting subsystem can shift the location of the attention window both covertly and overtly, such as occurs when we move our eyes, head or body to look for new information.

Operation of the subsystems working together

If an object is seen under optimal viewing conditions and is familiar, recognition (i.e., the match to visual representations in the object-properties-encoding subsystem) and identification (i.e., the match to representations in long-term associative memory) proceed very quickly and may be carried out entirely via bottom-up processes. However, if an object cannot be recognized and identified relatively quickly via bottom-up processing because it is unfamiliar or is seen under impoverished viewing conditions, then there is time for top-down processing to unfold. In this case, information about attributes of the best-matching object is accessed from long-term associative memory, and then is used to direct attention to the location where a distinctive part or attribute should be found (and to prime the object's representations in the object-properties-encoding subsystem), and a new part or attribute is encoded into the visual buffer,

beginning a new processing cycle. If the new part or attribute matches the primed representation in the object-properties-processing subsystem, this part or attributes is recognized, and this may lead to a good match in long-term associative memory-- and the object will have been identified. If not, either additional parts or attributes of that object are sought or another hypothesis is generated (e.g., by taking the next best-match in long-term associative memory) to guide a new search for a distinctive part or attribute.

Varieties of Top-Down Processing

In describing our theory, at different points we have invoked distinct types of top-down processing. In most theories, these different types are conflated, and simply referred to as different instances of the same kind of activity. However, we propose here that these processes are in fact distinct, and that each operates only in specific circumstances. In the following, we will outline a first-order taxonomy, and discuss some of the empirical evidence that supports it. Our theory makes an initial distinction between two broad classes of top-down mechanisms.

Strategic top-down processing

Strategic top-down processing relies on "executive control mechanisms" (which provide input to the information-shunting subsystem) to direct a sequence of operations in other brain regions, such as is used to engage voluntary attention or to retrieve stored information voluntarily. In the case of visual object recognition, strategic top-down processing is recruited when the initial encoding is not sufficient to recognize an object. In such circumstances, the best-matching information is treated as a hypothesis, which then is used both to shift the focus of attention and to prime representations in the object-properties-processing subsystem to facilitate the encoding of a sought part or characteristic.

For example, if a picture of a degraded object (degraded perhaps because it is partially hidden by another object, or is in poor lighting) is presented in a familiar visual context, then the best-matching representation in long-term associative memory is treated as an "hypothesis," which is used by the information shunting subsystem to direct attention to possible parts or characteristics that would identify the degraded object (Ganis, Schendan, & Kosslyn, submitted; Kosslyn, Thompson, & Alpert, 1997). To illustrate, if a chair is presented in the context of a kitchen, the associations stored in long-term associative memory can be used to identify the chair even if only small parts of it are visible above and behind the table.

We also note that strategic top-down processing can also take place in the absence of any visual input, such as in some cases of visual imagery. In these cases, the information-shunting subsystem directs a sequence of operations entirely driven by endogenously generated information, which leads to the retrieval of stored information in the absence of an external stimulus (see Kosslyn, Thompson, & Ganis, 2006).

Two classes of evidence for strategic top-down processing have been reported, from non-human primates and humans. We summarize important examples of each class of research in what follows.

Strategic top-down processing in non-human primates. Although it is very difficult to obtain direct neural evidence for top-down processing because, at minimum, it entails recording neural activity from multiple sites in awake animals, some of this sort of direct evidence is available in nonhuman primates.

One of the most compelling studies that demonstrates strategic top-down processing in action was reported by Tomita and collaborators on the retrieval of visual paired associates in monkeys (Tomita et al., 1999). In this study, monkeys first learned a set of paired-associates, and then the posterior parts of the corpus callosum was severed; after this surgery, the only remaining communication path between hemispheres was via the anterior parts of the corpus callosum, which connect the prefrontal cortex in the two hemispheres. Thus, following surgery, a visual stimulus presented to the left hemisphere (right visual

hemifield) could only affect neural activity in right inferotemporal cortex by means of an indirect route, which drew on the left prefrontal cortex, the anterior corpus callosum, and the right prefrontal cortex (the corresponding sets of structures hold for right hemifield presentation). As expected, the presentation of a cue to one hemifield resulted in robust bottom-up activation of stimulus-selective inferotemporal neurons in the contralateral hemisphere (with an average latency of 78 ms). Crucially, presentation of the same cue to the other hemifield (ipsilateral) resulted in delayed activity in neurons that were selective for the probe or the paired associate (with an average latency of 178 ms).

After the main experiment, fully severing the callosum abolished responses to the ipsilateral stimuli but left the responses to the contralateral stimuli intact, which showed that the previous results were not due to subcortical influences. These results provide good evidence that the prefrontal cortex sends top-down signals to inferotemporal neurons during the retrieval of visual information.

Fuster and his collaborators reported another study that documented strategic top-down processing in nonhuman primates; this study relied on a reversible cooling technique that temporarily inactivates an area by using cooling probes (Fuster, Bauer, & Jervey, 1985). In this study, spiking activity was recorded from single neurons in monkey inferotemporal cortex during a delay match-to-sample task with colors, while the prefrontal cortex was inactivated bilaterally via cooling probes. Before inactivation, during the delay period, neurons in inferotemporal cortex showed a sustained response with a clear preference for the color to be remembered. However, inactivation of prefrontal cortex by cooling impaired this selectivity profile. The critical finding was that inactivation of prefrontal cortex impaired the monkey's performance in this task. These data indicate that top-down signals from prefrontal cortex are necessary for the maintenance of delayed, stimulus-specific activity in the inferotemporal cortex when no external stimuli are present.

Another study that documented strategic top-down processing in nonhuman primates was reported by Moore and Armstrong (Moore & Armstrong, 2003). This study was designed to investigate the effects of electrical stimulation of the frontal eye fields on neural activity in area V4. When electrical stimulation of the frontal eye fields is strong enough, it produces systematic saccades to specific locations; the specific target location of the saccade depends on which specific part of the frontal eye fields is stimulated. The researchers recorded from neurons in V4 that had receptive fields at the location where specific stimulation of the frontal eye field directed a saccade. The activity of the V4 neurons to preferred and non-preferred visual stimuli was also recorded without stimulation or with subthreshold stimulation of the frontal eye fields (subthreshold stimulation is not sufficient to elicit a saccade). The results showed that the responses of V4 neurons were enhanced when a preferred visual stimulus was within the neuron's receptive field and the frontal eye fields were stimulated subthreshold (i.e., without generating a saccade) compared to when there was no stimulation. Crucially, this effect was not present without a visual stimulus or with a non-preferred visual stimulus in the neuron's receptive field. Furthermore, the effect was not present if the receptive field of the neuron did not cover the end location of the saccade that would be elicited by suprathreshold stimulation of the frontal eye fields. Thus, electrical stimulation of the frontal eye fields simulated the operation of covert attentional shifts on neural activity in V4 (note that the eye movements were monitored carefully and trials during which the monkey was not fixating were excluded from the analyses). This finding provides direct evidence that strategic top-down signals can originate in the frontal eye fields, and then can affect activation in at least one area that we include in the visual buffer.

Another study, reported by Buffalo and collaborators (Buffalo, Fries, Landman, Liang, & Desimone, 2005), charted the timecourse over which areas in the ventral stream are engaged during strategic top-down processing. In this study, the researchers recorded single-unit and multi-unit activity in areas V4, V2 and V1, both during the peripheral presentation of visual stimuli and during a visual attention task. Firing rates were compared for conditions when the animal paid attention to a stimulus inside versus outside the receptive field of the neuron (while the monkey maintained fixation). As found

in other studies, paying attention to a stimulus inside the receptive field of neurons in areas V1, V2, and V4 increased firing rates. The onset of the attentional effects revealed a striking pattern: area V4 showed the earliest onset (240 ms post-stimulus onset), area V2 was next (370 ms), and area V1 was last (490 ms). When the researchers presented simple visual stimuli, they now found exactly the reverse order of activation onset. These results strongly suggest that strategic top-down signals trickle down from higher-level visual areas that receive these signals from prefrontal or parietal areas.

Strategic top-down processing in humans. In humans, the evidence for strategic top-down processes is largely indirect, because of limitations of the non-invasive neuroimaging techniques employed. Nonetheless, such evidence is consistent with that from nonhuman primates. We will discuss briefly three functional magnetic resonance imaging (fMRI) studies that document how strategic top-down processes affect neural activation in the ventral stream in humans.

Kastner and colleagues (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999) investigated the mechanisms by which visual attention affects activation in occipital cortex in the presence of multiple stimuli. The rationale for this study was grounded in a finding from single-cell studies of monkeys; this finding showed that responses of neurons in areas V2 and V4, to an otherwise effective stimulus, decrease when a second stimulus is presented in the neuron's receptive field. However, this decrease in response is eliminated if the animal pays attention to the first stimulus, ignoring the other stimuli in the receptive field (Reynolds, Chelazzi, & Desimone, 1999). To test the hypothesis that these same effects can occur in the human visual system, Kastner and collaborators presented four images in the upper right visual quadrant, either simultaneously (SIM) or in sequence (SEQ) in independent trials. In the "attend" condition (ATT), the participants were asked to maintain fixation and to pay attention to the images presented peripherally at a given spatial location and ignore the others. In the control condition, participants were asked simply to maintain fixation and ignore all visual stimuli in the periphery (UNATT). During ATT trials, 11 seconds before the onset of the visual stimuli there was a small cue at fixation, telling participants to direct attention covertly to the appropriate location in the visual field, waiting for the visual stimuli to be presented. Using this methodology, the researchers could monitor brain activity during attention in the absence of visual stimulation.

The results revealed increased activation during the expectation period (before the onset of the visual stimuli) in several visual areas, including V1, V2/VP, V4, and TEO. The increased baseline activation was retinotopically specific, and it was strongest in visual areas TEO and V4. This increase in baseline firing rates in the absence of visual stimuli is very similar to that found in non-human primates during attentional tasks (e.g., Luck, Chelazzi, Hillyard, & Desimone, 1997). In addition, Kastner and colleagues found greater activation in extrastriate visual areas in the response to visual stimuli during the ATT condition, relative to the UNATT condition. Furthermore, as expected, the increase was larger in the SIM than in the SEQ condition, because the inhibitory effects of multiple stimuli were not as strong when the stimuli were presented sequentially. Finally, the researchers did not observe an effect of type of presentation in area V1, probably because the receptive fields in area V1 are too small to encompass more than one of the stimuli used in this experiment.

Kastner and collaborators also examined activation in areas outside the ventral stream, to gather evidence about the sources of the modulation of visual areas. They found that areas in the frontal lobe (specifically, the frontal eye fields and supplementary eye fields) and in the parietal lobe (the inferior parietal lobule and superior parietal lobule) showed robust increases in activation during the expectation period. The frontal eye fields, supplementary eye fields and the superior parietal lobe areas did not display increased activation following presentation of the visual stimuli, which suggests that they were not driven by bottom-up information but rather were involved in strategic top-down processing. This result is consistent with the fact that the frontal eye fields and the supplementary eye fields have rich efferent (i.e., feedback) connections to areas in the ventral stream and posterior parietal cortex (Felleman & Van Essen, 1991). Furthermore, this result is consistent with findings that these areas are engaged

during covert shifts in attention (Beauchamp et al., 2001; Corbetta, 1993; Corbetta & Shulman, 1998), as described earlier.

Another fMRI study, performed by Ress, Backus, & Heeger confirmed and complemented Kastner et al.'s findings. In this study, participants were asked to respond when an annulus (inner radius = 3 deg/outer radius = 6 deg) appeared in the center of the screen. In the main condition, the contrast of the stimuli was adjusted for each participant so that correct detection occurred only on 75% of the trials. Target trials were intermixed with catch trials, during which no stimulus was presented. A slow event-related paradigm was used to allow the analysis of activation during individual trials. Results showed that the regions in areas V1, V2, and V3 representing the annulus exhibited a robust BOLD response both during annulus-present and annulus-absent trials (note that the annulus-absent trials resemble the UNATT condition in the experiment by Kastner et al.). In fact, when the stimulus was presented at the lowest contrast levels (near-threshold), activation during the annulus-present trials was only slightly stronger than during the annulus-absent trials. Crucially, BOLD activity predicted the participant's performance on the detection task: the greater the activity, the more likely the participant would correctly detect the presence or absence of the annulus. Furthermore, the annulus-absent BOLD response was much smaller during an independent condition that used blocks of trials with higher contrast stimuli (which were easier to detect), which suggests that this response reflected neural activation required to perform the more difficult detection task. Strategic top-down processes might increase sensitivity in low-level visual areas to incoming visual stimuli by pushing the neuron into a higher gain region of its operating range, where smaller differences in the input produce relatively larger differences in response (Ress, Backus, & Heeger, 2000). Because higher firing rates are costly metabolically, it makes sense to have a mechanism that can increase sensitivity by increasing baseline firing rates only during difficult detection situations.

Although the study by Kastner et al. (1999) found parietal and frontal activation during the expectation period, the results do not provide direct evidence that these regions are actually a source of strategic top-down influences. Furthermore, Ress and collaborators did not sample the entire brain, and therefore were unable to provide direct evidence as to source of the modulation of activation they observed in striate and extrastriate cortex. A more recent fMRI study used dynamic causal modeling to investigate the direction of influences among brain areas during visual mental imagery and visual perception (Mechelli, Price, Friston, & Ishai, 2004). The key to this study is the comparison between mental imagery and perception. According to our theory, during visual imagery the information shunting subsystem retrieves stored representations of the structure of an object in long-term associative memory and sends information to the object- and spatial-properties processing subsystems to activate the corresponding modality-specific representations. According to the theory, this activation process is identical to the priming that occurs during top-down hypothesis testing in perception; however, now the priming is so strong that activation propagates backwards, and an image representation is formed in the visual buffer. The visualized shapes and spatial relations are retained (which is equivalent to holding them in "working memory"), and they can be inspected and identified in an image by the same attentional mechanisms used to inspect objects and locations during perception.

Mechelli et al. compared blocks of trials in which participants formed visual mental images of faces, houses, or chairs with blocks of trials in which participants actually viewed the same objects. The researchers first measured intrinsic connectivity during visual perception and visual imagery, that is, the influence brain regions have on each other as a result of being in visual perception or visual imagery modes (regardless of the visual category); such intrinsic connectivity was then used as a baseline to quantify functional connectivity changes brought about by the experimental manipulation (visual category) within those modes. The intrinsic connectivity during visual perception revealed paths from occipital cortex and superior parietal cortex to ventral temporal cortex, whereas that during visual mental imagery revealed paths from superior parietal cortex and the precuneus to ventral temporal cortex. Analyses on the category specificity of the changes in functional connectivity showed that during visual

perception there was an increase in functional connectivity from low-level visual cortices to the regions of ventral temporal cortex selective for the corresponding stimuli. For instance, the functional connectivity between inferior occipital cortex and the ventrotemporal region that responded the most to faces increased the most during the presentation of blocks of trials containing faces (compared to blocks of trials containing houses or chairs). In contrast, during visual mental imagery, the researchers found a selective increase in functional connectivity from prefrontal cortex and parietal cortex to these regions in ventral temporal cortex. However, only the strength of the path from prefrontal cortex was modulated by stimulus category (i.e., faces, houses, or chairs). For instance, the functional connectivity between prefrontal cortex and the ventrotemporal region that responded the most to faces increased the most during the presentation of blocks of trials containing faces (compared to blocks of trials containing houses or chairs). Thus, the analysis of functional connectivity changes during visual imagery (compared to visual perception) suggests the existence of two types of strategic top-down influences on the ventral stream. The first one is an influence from the parietal cortex that is not modulated by stimulus category (i.e., it is the same regardless of the category of the visualized stimulus) and may reflect the operation of attentional mechanisms, whereas the second one is a category-specific signal that may be involved in the reconstruction of visual information in category-specific areas in the ventral stream.

Reflexive top-down processing

We have so far been focusing on strategic top-down processing, which is under voluntary control. We also propose a second major class of top-down processing, which is automatic. Such *reflexive top-down processing* occurs between areas that are bidirectionally connected in the visual buffer, the object-properties-processing subsystem, and in long-term associative memory. Crucially, reflexive top-down processing is triggered by bottom-up signals without the intervention of the information shunting subsystem in the prefrontal cortex.

Reflexive top-down processing in non-human primates and in humans. The best evidence for reflexive top-down processing comes from studies in non-human primates that investigated how stimulus-driven neural activity in area V1 is affected by activity in higher-level visual areas. Similar processes probably take place in higher levels of the visual hierarchy and in long-term associative memory as well.

Some of the more compelling experiments use reversible inactivation techniques in anesthetized animals; these techniques rely on cooling or application of the inhibitor GABA. One critical finding is that inactivation of area V2 changes the response properties of neurons in area V1, generally making them less selective (Payne, Lomber, Villa, & Bullier, 1996; Sandell & Schiller, 1982). However, it is difficult to know how the anesthesia affected these results. Thus, findings obtained in awake monkeys, such as from the study reported by Lee and Nguyen (Lee & Nguyen, 2001), are probably more compelling. In Lee and Nguyen's study, activity was recorded from neurons in area V1 and V2 while monkeys viewed illusory contours (such as the ones produced by the four black disks in Figure 3) as well as corresponding real contours. The presentation paradigm was slightly different from that of previous studies that had failed to observe responses to illusory contours in V1 neurons (von der Heydt, Peterhans, & Baumgartner, 1984). In this new paradigm, four black disks centered at the corners of an imaginary square were first presented for 400 ms. Then they were suddenly replaced with four corner disks in the same position, giving the impression that a white square had appeared in front of them, generating partial occlusion. There were also numerous control conditions, including squares defined by real contours.

The results of Lee and Nguyen's study showed that, as expected, neurons in V1 responded vigorously to real contours of appropriate orientation, with a response onset of about 50 ms relative to stimulus onset. These same neurons, at least those in the superficial layers of V1, also responded to illusory contours with the same preferred orientation. However, the neurons responded more weakly to illusory contours, and, crucially, they began to respond about 55 ms later than they did to the real contours. The neural responses in area V2 to illusory contours were generally stronger than those in area

V1 and began around 65 ms post-stimulus, which was about 40 ms *before* the V1 response. One plausible interpretation of this finding is that V2 neurons, with their larger receptive field size compared to area V1 neurons, integrate more global information and can aid the contour completion process in V1 by sending feedback. The advantage of sending information back to area V1 is that V1 maintains a higher-resolution version of the visual input (because of its small receptive field sizes), whereas higher-level visual areas have access to more abstract and global information required to parse the visual input into meaningful parts (e.g., surfaces) (Lee & Mumford, 2003; Lee & Nguyen, 2001). In addition to feedback processes, these findings may also reflect properties of recurrent circuits within V1 itself, which may take as long to carry out some contour completion operations as feedback from area V2 (Girard, Hupe, & Bullier, 2001).

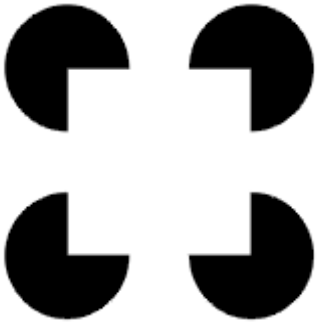

Figure 3*. An example of illusory contours (modal contours) used to study feedback from V2 to V1.*

Although Lee and Nguyen (2001) did not record from neurons outside of areas V1 and V2, the completion processes documented by this study are probably also affected by feedback from higher-level visual areas, such as inferotemporal cortex, perhaps on a different timescale. Indeed, monkeys with inferotemporal lesions have been shown to be severely and permanently impaired at shape discriminations based on illusory contours (Huxlin, Saunders, Marchionini, Pham, & Merigan, 2000). Furthermore, this putative role of feedback from higher-level areas is consistent with results from neuroimaging studies of humans, although the limitations of the non-invasive techniques make inferences more difficult to draw. For instance, Halgren and colleagues (2003) recorded **Magnetoencephalography** (MEG) while participants viewed arrays of shapes defined by illusory contours versus arrays that contained similar stimuli without illusory contours. The MEG activation to illusory contours was localized to the cortical surface by using a linear estimation approach that included noise-sensitivity normalization (Dale et al., 2000; Liu, Belliveau, & Dale, 1998). The results revealed multiple waves of activation in occipital polar cortex that suggested the operation of feedback loops. Specifically, following an earlier activation in the occipital pole around 100 ms after stimulus onset, a second wave of activation between about 140 and 190 ms after stimulus presentation spread from object-sensitive regions in the anterior occipital lobe back to foveal parts of areas V3, V3a, V2, and V1.

Finally, when reviewing the cognitive neuroscience literature on illusory contours, (Seghier & Vuilleumier, 2006) suggested that there may be two distinct feedback processing stages unfolding during the first 200 ms post-stimulus: the first involves interactions between areas V1 and V2, and the second involves feedback to areas V1 and V2 from higher-level visual areas, such as lateral occipital complex

(probably homologous to some object-sensitive inferotemporal regions in monkeys—which corresponds to our object-properties-processing subsystem). According to our definition of reflexive top-down processing, both processing stages would be examples of reflexive top-down processes, even though they take place among different sets of areas in the ventral stream.

Modulating interpretation

We further propose that both strategic and reflexive top-down processing can operate by altering the way earlier activation is interpreted. We can distinguish two types of such processing, which are directly analogous to changing the parameters d' and β in classical signal detection theory.

Changing sensitivity. On the one hand, higher-level areas can increase or decrease the sensitivity (corresponding to d') of the neurons that implement subsystems earlier in the processing stream (e.g., by increasing the baseline firing rates), making them more likely to detect the information to which they are selective. For example, in our theory, the attention shifting subsystem has the effect of priming some regions of the visual buffer (i.e., focusing the "attention window"). In addition, the information shunting subsystem passes information from long-term associative memory to the object-properties-processing subsystem; this information has the effect of increasing the sensitivity of neural populations in this subsystem for the expected parts or characteristics.

This sort of "anticipatory" priming is strategic; a comparable kind of reflexive priming can occur in the presence of a highly constraining context. In such a case, associations in long-term associative memory would be activated by the input, and may have the effect of reflexively providing feedback to increase the sensitivity to objects that are associated with the context (such as a nose in the context of a face). By the same token, sensitivity can also be reduced, for instance, to filter out unwanted stimuli. Some of the studies discussed earlier (Kastner et al., 1999; Ress et al., 2000) that found baseline increases due to expectation in the absence of visual information illustrate this type of process taking place in multiple areas in the ventral stream.

Changing decision criterion. On the other hand, feedback can alter how much information is necessary to make a decision. For example, to the farmer gathering his cows at dusk, a passing shadow may be sufficient to recognize a cow. Such changes in criterion (β in classical signal detection theory) could affect two kinds of processing:

First, they could affect simple detection thresholds. For example, they could alter how much activation of specific neurons in area V4 is necessary to register that one is viewing a particular color. Similarly, if one is expecting to see a handle on a cup, the threshold for that part could be lowered to the extent that only a small portion of the handle would be required to trigger the corresponding representations in the object-properties-processing and long-term associative memory subsystems.

Second, top-down processing could alter the threshold *difference* in activation required to decide between two or more alternatives. For example, a "winner take all" process probably takes place in the object-properties-processing subsystem, where representations of objects are mutually inhibitory – so that only one representation of an object is activated at a time. Top-down processing could affect how much relative difference in activation is required for one representation to "win" over its competitors (e.g., Kosslyn, 1994; Miller & Cohen, 2001). For example, when viewing a kitchen scene, reflexive top-down processing from long-term associative memory may bias representations of "chair" so that they are not only activated by less input (i.e., their thresholds are lowed), but they need not be activated much more strongly than representations of other objects.

Supplementing input

In addition to altering how input is interpreted (via changing sensitivity or decision criterion), either strategic or reflexive top-down processing could have its effects by actually filling in missing information. In neural network models, such processing is called "vector completion" (e.g., Hopfield,

1982). According to our theory, if representations in the object-properties-processing subsystem are primed strongly enough, feedback connections from the areas that implement this subsystem to the areas that implement the visual buffer can force activation in these early visual cortical areas. This activation in turn corresponds to a high-resolution visual mental image; all such imagery relies on such completion operations. Such images can be strategic, as when one intentionally tries to visualize, or reflexive, as occurs when a partially degraded object is seen and one "automatically" fills in missing contours. As an illustrative example, say that you see just the very tip of the nipple on a baby bottle sticking out from under a cloth. The portion of the tip (its shape, size, texture and color) is sufficiently distinctive that the "baby bottle" representation in the object-properties-processing subsystem is activated. However, the input is too degraded to be sufficient for recognition. In this case, top-down processes may "complete" the image, allowing one to "see" the remainder of the bottle (as a specific shape under the cloth), at the proper size and orientation to fit the visible features. If that image cannot be "fit" to the input, then the object must be something else. If the input image can be so "completed," that would be evidence that the activated representation is appropriate.

We stress that the process of supplementing input is distinct from even an extreme case of modulating interpretation, either via altering sensitivity or criterion: no matter how much we increase sensitivity for a certain visual attribute or lower our decision threshold, missing information is not filled in. Completion involves actually adding information to a representation earlier in the processing sequence. Some of the studies discussed earlier (Mechelli et al., 2004; Tomita et al., 1999) are examples of strategic top-down processes that perform some form of pattern completion.

The distinction between modulating interpretation and supplementing input corresponds to the distinction between "inspecting" a pattern in a visual mental image and "generating" the image in the first place (Kosslyn, 1994). The former process relies on mapping the input to a specific output, which is the interpretation; whereas the latter relies on using one representation to create another, which need not be fully interpreted in advance. For example, when asked what shape are Mickey Mouse's ears, most people report that they visualize the cartoon character. The process of visualizing relies on strategic top-down processing, where information in the visual buffer is supplemented – creating a pattern that depicts the object. Once formed, this representation can then be interpreted; one can classify the ears as "round." At the outset, however, the pattern was not necessarily interpreted in this way; such prior interpretation is not a prerequisite for generating an image (see Kosslyn et al., 2006).

In sum, our theorizing set the stage for implementing a new computational model, which incorporates novel distinctions regarding the nature and types of top-down processing.

Formulating the "Binder"

Finally, we used the theory and distinctions to guide an encyclopedic literature review, the results of which are provided in a binder. We not only looked at the citations in all of the articles cited above and followed up relevant articles, but also conducted searches on Pubmed and PsychInfo, as well as Google. We sought to find all research that pertained to each of the subsystems and their anatomical and functional connections. This literature stretched from neurophysiology and neuroanatomy of animal models to human neuroimaging to relevant studies of behavior. The result is the binder, which is organized according to the distinctions we draw in the theory. We also include a section on representative computational modeling (this is not all-inclusive, which would have gone beyond the scope of our project). The articles on other computational models serve not only to set the stage for our own modeling, but provide a context for understanding the value and importance of the studies cataloged in the remainder of the binder. The results we found provided strong constraints on our own computational model, which we intended to be consistent with the available literature.

3. Results and Discussion I: Computational Modeling of Late Vision and Mental Imagery (IMPER model)

3.1 IMPER model specifications.

We have made progress in implementing the theory in a running computer simulation model. Note that the focus of this model is at the systems level: We had to model all the components in sufficient detail to have a working model, but the main issue is the extent to which the same system, as a whole, can carry out both perception and imagery operations. The following is the specification for all components of the IMPER (IMagery and PERception) model, which consists of the processing subsystems summarized in the first section of this report (see Figure 2): the visual buffer (VB), object properties processing system (OPPS), spatial properties processing system (SPPS), associative memory (AM), information shunting system (IS), and attention shifting system (AS). Each component corresponds transparently to processing subsystems postulated in our theory. IMPER models both forward and backward information flow between these processing subsystems and tries to capture essential aspects of both strategic and reflexive top-down processing. Although the underlying brain systems send continuous signals, these processes are modeled in discrete time steps.

Visual Buffer

The role of the visual buffer (VB) is to hold the image information that is being processed by the rest of the system. The attention window (AW) is the current focus of the VB and can either store the whole image (at a limited resolution) or focus on specific details in high-resolution. The VB does some basic level processing of the image as a whole (output for the spatial properties processing system, SPPS) and then further analyzes the contents of the AW (output for the object properties processing system, OPPS).

Input

From Experiment:

- Current image input from early vision (if any)

From OPPS:

- High curvature points on the most likely object model (in center relative coordinates)
- Confidence in the likelihood of the object model

From SPPS:

- Center point of the figure (in absolute coordinates)
- Standard deviation of the center

From AS (attention shifting system):

- AW position (in absolute VB coordinates)
- AW size (in absolute VB coordinates)

Output

To OPPS:

- Set of high curvature points from the current image

To SPPS:

- Center of image (in absolute VB coordinates)
- Four spread points (top, down, left, and right of center)
- Center of AW

- Size of AW

Representation and processing

The visual buffer (VB) holds the current image, whether perceived or imagined. In the case of perception, it does early processing of the input image, simulating the processing done in early visual areas. Furthermore, the VB extracts essential image information for both the spatial (SPPS) and object (OPPS) properties processing systems. It also receives reflexive feedback from each of these systems, which affects how it processes the current image.

The VB also receives input from the attention shifting system (AS), which tells it where to locate the attention window (AW). The location and scope of the AW affects the VB processing for the OPPS and SPPS, although part of the SPPS processing occurs preattentively (i.e., the total content of the visual buffer is processed, not just the contents of the AW). The AW has a scope/resolution tradeoff, where it can see the whole image at low resolution, or specific details at high resolution. What follows is a step by step description of the algorithms used in all stages of processing in the VB.

Early perceptual processing. In the case of perceiving an image, the first processing step of the VB consists of separating the figure from ground in the AW. The procedure is as follows:

1) *Find peak luminances.* Looking at the histogram of the pixel counts for each luminance level of a black and white version of the image, separate out the peak values from the other local maxima. The pixel counts at the local extrema are multiplied by the distance from the nearest equal pixel counts in the histogram (or half the number of luminance levels in the case of global extrema). The idea is not only to find the highest peaks, but to find locally high peaks even if they are not very high relative to the global maximum. Peak luminances are those that are above average on the measure: pixel count * distance.
2) *Find cutoff luminance level:* Next the algorithm looks for the largest distance between the peaks. It assumes that this largest separation is between the "black" and "non black" areas. The cutoff point is determined to be the trough (i.e., local minimum) immediately following the lower luminance peak bordering the largest interpeak distance just found.
3) *Figure ground determination:* Takes the cutoff value determined in the previous step and sets all "black" areas (i.e., luminance less than cutoff) as ground and all "non-black" areas as figure.
4) *Fill in and touch up:* at this point, the object may still have "black" areas inside white areas; the fill-in process looks at all areas that are walled in on at least M (max 4, parameter set to 3) sides by N-thick (parameter set to 5 pixels) white pixels and counts these as the final figure areas of the image.

Using the resulting figure, the following image descriptions are calculations for output to SPPS: center and spread. The algorithms for each measure follow below:

Center. All points (pixels) in the image are assigned a probability of being the center based on what proportion of the total figure they are to the right of and below in the x and y directions. The most likely center is to the right of and below exactly half of the figure. The rest of the points are calculated by the (normalized) distance between their proportions and the ideal proportions. These conditional probabilities (i.e., probability of each point being the center based on the image input) are then multiplied by the prior probabilities, which are sent as feedback to the VB from the SPPS. By default the prior probabilities are equal for each point. If the SPPS sends a feedback center point to the VB, then the priors are calculated from a two-dimensional Gaussian function centered on this feedback point with a standard deviation also provided by the SPPS. This produces an exponential drop-off in priors probabilities for points in the image as they deviate from the expected center. The highest probability point (priors * conditional probabilities) is returned to the SPPS.

Spread. The spread calculation takes the previously determined center and find the points in the figure that are horizontally and vertically most distant from the center. These four points: top, bottom, left, and right are sent to the SPPS as the spread of the image.

The VB also sends the location and size of the AW to the SPPS.

The VB then continues with the analysis of the contents of the AW. First, if necessary, it moves the AW's location and adjusts its size as specified by the input from the AS (described below). The VB extracts the high curvature points from the figure resulting in a compact image shape description requiring only a few points. The following computations are used in this process:

1) *Determine figure outline:* the outline of the figure is calculated as only those pixels in the image which border a non-figure ("black) pixel on any one of four sides.
2) *Calculate curvature:* The curvature, k, at any given pixel in the figure's outline is calculated by multiplying the turning angle by the distance between points used to determine the turning angle (this distance is a parameter set to 5 pixels; Feldman & Singh, 2005; Ghosh & Petkov, 2006).
3) *Calculate prior probabilities:* If the OPPS has not sent feedback about the likely high curvature points, all points on the outline are assigned equal prior probability of being a high curvature point. If the OPPS has provided candidate points, then all points are assigned an equal share of 1.0-confidence probability, and those points on the outline nearest those provided by the OPPS gain an additional probability equal to confidence/N, where N is the number of these points (a parameter in the model set to 12).
4) *Finding high curvature points:* To force the high curvature points to be more distributed around the outline of the model, their distance from one another is also a factor in the decision. The process is iterative and starts by selecting the most likely high curvature point (as determined by priors * conditional probabilities). After each new point is selected (up to a total of N), the distance between all the remaining points in the outline and the closest point already selected to be a high curvature point is calculated. The original likelihoods (priors * conditional probabilities) is weighted by this distance measure.

The center value calculated earlier is subtracted from the set of high curvature points making their coordinates all relative to the center (i.e., position invariant). They are then sent to the OPPS for further calculation.

Noise. Gaussian noise is added to both the internal calculation in the VB, reflecting noise in both the input, early visual processing, and in VB processing. Specifically, noise is added to the internal calculations of center and curvature, and to the outputs sent to the SPPS and OPPS.

Object Properties Processing System

The role of the OPPS is to recognize the object in the figure extracted by the VB. The best candidate object models are sent to AM. The OPPS also has feedback connections to the VB, in order to help it with processing the figure outline.

Input

From VB:

- Set of high curvature points

From AM:

- Prior distribution across object models

From IS:

- Sampling rate

Output

To AM:

- Confidences for all object models

To VB:

- High curvature points of object model with highest confidence
- The confidence level for this object model

Representation and processing

The object properties processing system (OPPS) stores prototypical object models (i.e., objects viewed from a specific viewing angle). There are object models for whole objects and for parts of objects (see discussion of AM and the IS below). While many kinds of information about object models are stored in the analogous system in the brain, including color, texture, and shape, we have simplified the model by focusing only on shape. Each object model is two N-dimensional vectors whose values represent the distances and angles (i.e., polar coordinates) for each of N center relative high curvature points on the outline of that object model. In reflexive top-down processing, these sets of high curvature points, along with the confidence for the most likely object model, are sent to the VB for processing (for both perception and imagery).

First the OPPS converts the input from the VB into polar coordinates and orders both vectors based on the magnitude of the distance. It also scales the resulting distance vectors by the largest vector, to overcome effects of changes in size of the original image. The main processing task in the OPPS is to match these input vectors to stored representations in order to find the most likely object model. It does this by following this algorithm:

Matching to stored descriptions.

1) Each time-step sample from the VB of the distances and angles of the high curvature points will produce two 12-dimensional vectors. After the second sample, these vectors will be averaged with the previously sampled vectors to produce the sample mean (I).
2) Two standard errors of the mean (SEM) will be calculated from the standard deviations for each vector. These will be σ_d and σ_a for the SEMs of the distances and angles. These SEMs are assumed to be unequal.
3) We ultimately wish to compute the following Bayes equation for each stored object structural description, or object model (OM):

$$P(OM_i \mid I) = \frac{P(OM_i) * P(I \mid OM_i)}{\sum_{j=0}^{j=n} P(OM_j) * P(I \mid Obj_j)} \quad (1)$$

where n is the number of total stored object descriptions. $P(I \mid OM_i)$ is based on the distance between the stored object model and the mean of the input vector samples (computed using simple Euclidean distance between the stored object model vectors for distances and angles and the standard errors of the vector components, i.e., the 2 SEMs calculated above).

4) The exact equation for $P(I \mid OM_i)$ uses the multivariate elliptical form of the Gaussian function (since the input is a multidimensional vector and the SEMs are not equal). For simplicity, it is assumed that the covariance between these two sets of inputs is 0:

$$f(I,OM_i) = \frac{e^{-\left[(OM_{iD}-I_D)^2/2\sigma^2_D+(OM_{iA}-I_A)^2/2\sigma^2_A\right]}}{(2\pi\sigma_D\sigma_A)^{n/2}} \quad (2)$$

In the above equation (2), the distance between the sample input mean (I) and the object model is calculated for each component, and is divided by the SEM of that component. The subscripts used for the SEMs are also used for the components of the input and object model vectors.

The denominator of the above equation is a constant *k*:

$$k = (2\pi\sigma_D\sigma_A)^{n/2} \quad (3)$$

that does not vary across the different object models, so it will cancel out in the calculation of the Bayes equation (with one exception, see step 5 below). Removing it, produces the simplified equation 3, which is used in computations:

$$g(I,OM_i) = e^{-\left[(OM_{iD}-I_D)^2/2\sigma^2_D+(OM_{iA}-I_A)^2/2\sigma^2_A\right]} \quad (4)$$

5) To deal with situations where insufficient information is available to make a good classification, we introduced a convergence factor, OM_x, such that $P(I, OM_x)$ is some small constant (currently set to 1). While $P(OM_x, I)$ is calculated using equation 1 above and is included in the sum of all object models, it is not an object model. If $P(OM_x, I)$ is higher than the likelihood for all object models, then this means that the system has not converged on a good match and will continue calculations. Thus, this convergence factor keeps the model from settling quickly on a bad match just because it is not as bad as all the other matches.

6) $P(OM_i)$ is the prior probability, which by default will be equal for all object models, except for the convergence factor (where $P(OM_x)$ is a parameter in the model currently set to 10^{-3}). However, the underlying brain system has probably weighted these by frequency of occurrence (e.g., non-canonical views would be far less frequent). Furthermore, these prior probabilities can also be set by input from associative memory (AM) to reflect top-down processing (similar to changing β in signal detection theory).

Shape classification. Using a similar classification process, but with uniform priors (this information is assumed to be calculated on the fly without the benefit of pre-stored knowledge in AM), OPPS will classify the general 2D shape of the object to a simple geometric form (e.g., circle, square, triangle, etc.; see tasks 6 and 7 below).

Reflexive top-down processing. After a few samples (parameters) have been taken, the high curvature points of the most likely object model are sent back to the VB to replace its internal

calculations. This step is designed to speed up convergence on the proper object model. These points are first converted from their polar coordinates back to Cartesian coordinates.

Sampling rate. The information shunting system (IS) may also send feedback to the OPPS telling it to take more samples from the VB per time step. By default, the rate is one sample per time step and increasing it would theoretically require more resources, so the IS would only do so if the identification process was not converging on schedule. Changing this sampling rate should make the system more sensitive (similar to changing D' in signal detection theory).

Noise. Gaussian noise is added to the OPPS output to AM and the curve points it sends as feedback to the VB. This noise simulates analog noise in inter-processing subsystem communications.

Spatial Properties Processing System

The SPPS keeps track of the object map of the world. This is a representation of the locations of all known objects in the vicinity. The size of the object map is thus much larger than the visual field that the VB takes in.

Input

From VB:

- Center of figure location
- Spread of figure
- Location of AW
- Size of AW

From AM:

- Most likely location of center of current object
- Standard deviation of its location

From IS:

- Sampling rate

Output

To VB:

- Center of object in VB coordinates

To AM:

- Center of object in object map coordinates

Representation and processing

The SPPS keeps track of all the objects surrounding the individual. Their locations are stored in the object map. While it seems that the brain stores these locations in polar coordinates relative to the individual's location in space, and they are updated with the individual's movement, we have simplified this by using Cartesian coordinates (especially since we do not model self-motion). Furthermore we have simplified the three-dimensional aspect of these representations into only two dimensions, since that is all that is necessary for our tasks.

In the brain, the spatial properties processing system would also keep track of the VB's position relative to the object map, but we do not model this processing, since our VB model does not move while performing our tasks.

The SPPS calculates the center based on the inputs it gets from both the VB and AM. It keeps samples of the location of the center and spread points from VB while using the input from AM as the priors. This information is combined using the following equations:

Posterior Mean:

$$M = \frac{\bar{x}\frac{n}{s^2} + \mu_0 \frac{1}{\sigma_0^2}}{\frac{n}{s^2} + \frac{1}{\sigma_0^2}} \quad (5)$$

Posterior Variance:

$$V = \frac{1}{\frac{n}{s^2} + \frac{1}{\sigma_0^2}} \quad (6)$$

where x is the sample mean, s^2 is the sample variance, and n is the number of samples, while μ is the prior mean and σ is the prior variance. This process is repeated for the means of the x and y coordinates for the center and spread points. This assumes a Gaussian distribution and independence between the coordinates. Furthermore, we assume equal variance among the five points and pool our sample of their variance (for both x and y coordinates) into a single number. We furthermore use this sample estimate as an approximation of the population variance (considering the latter known instead of estimated).

The above process returns a posterior estimate of both the mean and variance of the center and spread. Using these distribution numbers we can complete several estimates of the location of the object and its overall aspect (see tasks 2-5 below).

Sampling Rate. IS can increase the sampling rate for the SPPS if the center is not being converged upon quickly enough. This process directly parallels the sampling rate increase that IS can initiate in the OPPS.

Noise. Gaussian noise is added to the output of the system to both AM and the VB.

Associative Memory

Associative memory (AM) stores information about objects and scenes, including their different views and details. In perception, AM takes as input likely object model candidates and calculates the likeliest object supported by such evidence. It also helps resolve processing of object models by sending new priors to the OPPS. AM also receives information about object locations from the SPPS and binds this information to the identity information that it computes from the OPPS input.

Input

From OPPS:

- Confidences for all object models

From SPPS:

- Center of currently viewed object

From IS:

- Requests for information about object details
- Priors across objects

Output

To IS:

- Most likely object and its viewing angle
- Confidence for above object
- Center of current object
- Standard deviation of its location
- Diagnostic feature of the object (i.e., a detailed part)
- Location of the diagnostic feature

To OPPS:

- Prior distribution across object models

To SPPS:

- Center of current object
- Standard deviation of its location

Representation and processing

Object Identification. While human long term memory has many functions that bind together all known information about an object, the processing subsystem in our model implements a subset of these functions. The primary function of AM in our model is to identify objects based on input from the OPPS about object models (object/view combination).

At each time slice, the input coming from the OPPS is a vector of the recognition confidences across the object. Since several object models may all represent the same object, these confidences are pooled into their respective objects to produce a vector of confidences across all objects. For example, consider that object models 1-3 are all different views of a single object. Their respective confidences are: 0.1, 0.2, and 0.5. The confidence value for the underlying object would be 0.8 and similar calculations would be conducted for the other object model confidences. These resulting object vectors are stored for each time slice. Averages and SEMs are calculated from these stored sample object vectors.

Ultimately, likelihoods of each object, $P(O_i|I)$ are calculated by the following Bayes equation:

$$P(O_i \mid I) = \frac{P(O_i) * P(I \mid O_i)}{\sum_{0}^{j=n} P(O_j) * P(I \mid O_j)} \qquad (7)$$

where n is the total number of objects. The first component $P(O_i)$ starts off being equal for all stored objects, except for the convergence factor, whose initial probability is a parameter (currently $P(O_x) = 10^{-3}$). AM can change these prior probabilities based on task expectations and input from the IS.

$P(I|O_i)$ is calculated using the multivariate circular Gaussian function:

$$f(I,O_i) = \frac{e^{-D_i^2/2\sigma_M^2}}{(2\pi\sigma_M^2)^{n/2}} \quad (8)$$

where D_i is the Euclidean distance between the average of the stored sample object vectors and the ideal input vector, with 1 in the position for object *i* and 0 for all other objects, and σ refers to the SEM of the stored sample object vectors.

As the SEM of the stored sample object vectors decreases (i.e., as more of them are collected), the confidence for a single object should eventually reach cutoff for identification (a parameter set to 0.95). While the convergence factor is higher than all other object likelihoods (it may even be higher than cutoff), it is not allowed to be selected and the system will continue processing. The most likely object, its viewing angle, confidence, and a single diagnostic feature for the object are sent to the IS.

Center. AM also keeps track of the location of the center of the object and its variance, information that it receives from the SPPS. It keeps samples of the x-y coordinates and keeps track of their variance. This information is bound to the current identity of the object so it is here in AM that the what and where information for objects is kept together.

Reflexive top-down processing. The identification and recognition processes occur simultaneously. Information about objects in AM may be useful in recognizing object models in the OPPS. To this end, there is a reflexive feedback loop from AM to the OPPS. AM uses the confidences that it calculates for each of the objects to determine new priors for the object models, which it sends to the OPPS. The confidence for each object is split evenly among its corresponding object models, thus producing the new prior probabilities for all the object models.

AM also can send back priors to change the way that the SPPS looks for objects. This is most evident when AM loads scene information into the OPPS and SPPS by manipulating their priors. However, even in the perception process, AM still sends the current stored location back to SPPS as feedback.

Object parts. Included in the object representation is a specification of a single detail (a simplification of the many parts of objects that the brain keeps track of) and its location relative to the object center. The object parts are also stored as their own objects and can also be recognized.

Information Shunting System

The Information Shunting system (IS) is the logic controller of the whole model. It monitors processing to make sure that it is progressing to convergence; the IS is also responsible for initiating and resolving strategic top-down processing if the lower-level systems cannot resolve location and identification information by themselves. Finally, it is responsible for initiating imagery in the model by telling AM to load a particular scene.

Input

From AM:

- Most likely object and its viewing angle
- Confidence for above object
- Center of current object
- Standard deviation of its location
- Diagnostic feature of the object (i.e., a detailed part)

- Location of the diagnostic feature

Output

To AS:

- Center of location to attend to
- Size of location to attend to

To AM:

- Request to load a scene

To OPPS:

- Sampling rate

To SPPS:

- Sampling rate

Representation and processing

In the simplest case, where AM and the OPPS can identify which object is being viewed and the SPPS can locate it correctly, the IS simply passes this information on to other systems. This system becomes much more important when problems arise with the perception, and particularly identification, of the correct object. This system is also crucial for initiating imagery.

Strategic top-down processing. After a certain number of time steps (minimum 3), information reaches the IS from AM. The IS monitors the confidence for the most likely object for a certain number of time steps (parameter in the system). If during this time, the object reaches criterion in AM and is selected, the IS does nothing. Otherwise, the IS checks to see whether the confidence is increasing. If this is the case, then all the IS does is send a signal to the OPPS and/or the SPPS to increase their sampling rate.

However, if there is a problem with converging on the correct object, then the IS invokes a check for the diagnostic parts of the most likely objects. This process consists of the following steps:

1) Query AM for part information for the most likely object
2) Send the likely coordinates of this part to the AS to shift attention to the part
3) Reset AM and the OPPS for recognizing and retrieving the object part
4) Update AM priors to reflect looking at parts, not whole objects
5) Monitor the success or failure of identifying the proper part, if process fails to converge in certain number of time steps then return to the SPPS the most likely object and its lower than threshold confidence (which will result in a "Don't know" response).

Reflexive top-down processing. The above process is strategy dependent and time/resource intensive. The IS also reflexively sends back information to AM. This includes the priors across objects and the criterion for selection. Priors would change based on task and/or scene knowledge. These types of knowledge are not currently modeled beyond having the IS send new object priors to AM to help it more quickly identify object parts in strategic top-down processing (step 4 above).

Attention Shifting System

The attention shifting system (AS) relays changes to the size and/or position of the attention window (AW) to the VB. It receives this input from the IS.

Input

From IS:

- Center of location to attend to
- Size of location to attend to

Output

To VB:

- Center of AW
- Size of AW

Representation and processing

The AS keeps track of the current center and size of the AW in the VB, both in absolute coordinates and in terms of internal coordinates of the VB. When directed it can change the location of the AW by sending the new center and size information to the VB for the next time step.

The AS also keeps track of the current location of the visual field (i.e., the VB) in the object map, and thus works on the same absolute coordinate system as the SPPS, not just the limited coordinates of the VB. Though not implemented in our model, the AS could initiate eye movements, including those that move beyond the current fixation of the VB.

Noise. Gaussian noise is added to the output of the AS with regards to the location and size of the AW.

3.2 Significant changes to technical approach.

N/A

3.3 Progress against planned objectives.

We have achieved most of our goals. The remaining major goal (if more funding were secured) is to expand the model in order to process a greater range of stimuli and to accomplish a greater range of tasks.

3.4 Technical accomplishments.

The model was designed to perform five perception and imagery based tasks. All perception and imagery tasks are analogous. We used a set of photographs of six real objects to create our stimuli: three toy cars and three toy animals from the Amsterdam library of object images (Geusebroek, Burghouts, & Smeulders, 2005). Each object had three photographs taken from three different views (we arbitrarily picked viewing angles of 0, 75, and 150 degrees), making for a total of 18 images (768x576 pixels, color PNG files). Each of these images was then placed in four locations of a larger image (1200x900 pixels) over a black background. This produced a total of 72 scenes used in our experiments. What follows is a detailed description of the tasks that the model can perform.

1. *Identification* (perception only; object task): this is the most basic task where the model must send back the most likely object in AM when it reaches a certain confidence level or after a certain number of time steps, whichever comes first. This task is similar to having subjects name the picture out loud (though we do not model the word production aspect of this process). In the case of perception the experiment sends the test scenes one by one to the VB of the model which starts a cascade of bottom-up processing, while in the case of imagery the scene number is sent to the IS system of the model, which then loads the scene from AM in a cascade of feedback processing (see model description above). It is clear that this analogous imagery task is trivially simple, since the system knows already which objects are in the requested scene, since it has “memorized” the scenes. However, we include this task to see how it is affected by added noise to the system, such as in the case of brain damage.

2. Low-resolution aspect (perception and imagery; spatial task): in this task the model decides whether the overall shape of the object is wider than it is tall, or vice-versa, or about the same. The calculation is performed in the SPPS and is fairly easy to do in the case of perception, but more difficult in the case of imagery.
3. High-resolution aspect (perception and imagery; object task): similarly to the previous task, the model has to decide whether the diagnostic detail is wider than it is tall, vice-versa, or about the same. This task requires a strategic intervention from the IS to refocus the AW on the diagnostic detail, before the SPPS can compute the necessary width to height comparison. In perception, this process requires the additional step of identifying the object as a whole before AM can retrieve the appropriate diagnostic detail and its location.
4. Low-resolution object location (perception and imagery; spatial task): in this case the model has to identify what quadrant the object occupies most (e.g., upper-left, bottom-right, etc.). This task does not require a high-resolution (i.e., low standard error) knowledge of the location of the center.
5. High-resolution object location (perception and imagery; spatial task): similarly to the previous task, the model now has to decide where the center of the object is more exactly. This task is similar to having participants point to the center of an object (both in imagery and perception).
6. Low-resolution shape matching (perception and imagery; object task): this task, performed by OPPS, is similar to object recognition, but instead of finding the matching object model, the OPPS compares the overall shape of the object to a set of polygons. This is similar to having the model answer such questions as "what is the shape of a ball?"
7. High-resolution shape matching (perception and imagery; object task): this is the same task as above with the exception that an object part instead of a whole object is matched to a set of polygons.

We ran three preliminary experiments looking at 1) how noise at each level of the system affects performance on the different tasks, 2) whether the top-down feedback connections enhance processing at earlier stages in perception, and 3) how different damage settings affected performance on the different tasks. In the first preliminary experiment we looked at each individual noise level in order to determine minimal (normal participants), moderate, and severe levels of noise. The dependent measures were both accuracy across scenes and average number of time slices for correct results. In the second preliminary experiment we looked at whether our model shows positive effects of strategic and reflexive top-down processing by individually severing the top-down connections and looking at performance across scenes on just the perception tasks. In our third preliminary experiment we looked at all pairwise combinations of damage settings in the model across tasks and their effects on the different tasks in our battery. Performance was measured as both accuracy and average time per correct answer. We were looking to find different patterns of dissociation based on the location of the damage (increased noise). We are beginning an article for publication. Since we have been able to conduct only preliminary experiments, we are seeking more funding in order to run more extensive and rigorous tests of the model, and to expand this initial implementation.

4. Results and Discussion II: Computational Modeling of Early and Intermediate Vision

4.1 BD model specifications.

Finally, we have been collaborating with Prof. Bruce Draper and his lab, trying to develop a model of early and intermediate vision that would, essentially, provide a more detailed implementation of

some input components postulated by IMPER. The following section reports progress on this aspect of our overall project. This work ended up also exploring some alternative ways to implement aspects of the model discussed earlier, and it would be of great interest to compare directly the efficacy of the different approaches.

As noted earlier, the gross regional functional anatomy of the human visual system is well-known. The early vision system includes the retina, the dorsal lateral geniculate nucleus of the thalamus (LGNd), the superior colliculus of the midbrain, and cortical regions V1 through V4. Beyond early vision the system splits into the ventral and dorsal streams. The ventral stream includes the lateral occipital complex (LOC) and posterior inferotemporal cortex (pIT). It processes object properties for tasks such as object recognition and landmark-based navigation. The dorsal stream includes region V3a, the mediotemporal cortex (MT), and structures in the posterior parietal cortex. It processes spatial and movement properties for tasks such as tracking, ego-motion estimation, and hand-eye coordination. The two streams converge on associative memories in the anterior inferior temporal cortex (aIT), the angular gyrus and area 19. The associative memories in turn communicate with the dorsolateral prefrontal cortex, which closes the loop by providing feedback to LGNd and superior colliculus through pathways that include the frontal eye field. For accessible overviews of the anatomy of human vision, see Milner & Goodale (1995), Kosslyn (1994), or Palmer (1999).

Regional functional anatomy does not by itself define a software architecture, however. Architectures specify both components and interfaces. One of the goals of this project was to define an architecture with interfaces based on a literature review of behavioral studies, lesion studies, brain imaging techniques and electro-physical recordings, and to use this architecture to explore computational models of top-down vision. The architecture is limited to the task of object recognition. It does not consider visual tasks such as tracking or ego-motion estimation that are computed in the dorsal visual stream, allowing us to concentrate on the early vision system, the ventral stream, and associative memories. The result is a model of the inferotemporal cortex and its relation to associative memory, as well as refined models of early vision and the lateral occipital complex.

A Biomimetic Software Architecture

The software architecture below formalizes the major components of the human visual system and adds well-defined interfaces. Readers may already be familiar with rough functional characterizations of many of the modules. In particular, Figure 4 shows four modules outlined in black: the early visual system (attention and retinotopic processing), lateral occipital complex (feature extraction), posterior inferotemporal cortex (object recognition) and associative memories (object identification). As discussed below, the most distinctive part of this architecture lies in the definition of the object recognition module, which models pIT, and its interface to the associative memories. Object recognition is defined as an unsupervised clustering task, not a supervised (or even unsupervised) labeling problem. Labeling and other forms of cross-modal associations are modeled in the associative memories, which operate over clusters, not samples.

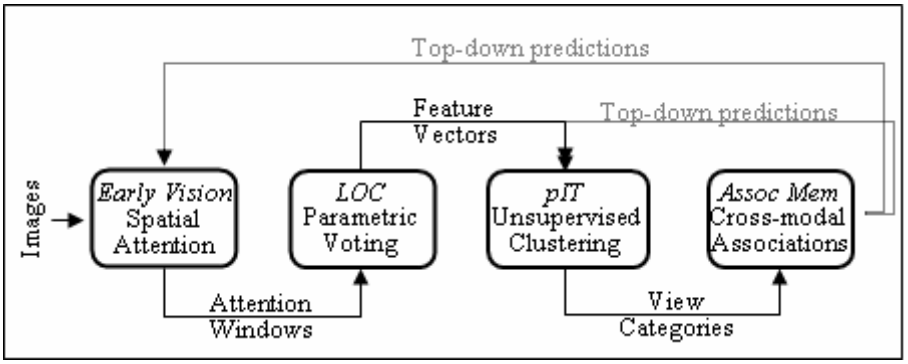

Figure 4*. The biomimetic architecture. LOC refers to the lateral occipital complex, pIT to the posterior inferotemporal cortex, and Assoc Mem to associative memories. Arrows in gray (representing top-down processing) were implemented for the first time as part of this work.*

Figure 4 shows arrows in light gray which are part of *top-down* object recognition and which were added to the model as part of this work. Some of these top-down connections pass through the dorsolateral prefrontal cortex and frontal eye field. Without these connections, the architecture models object recognition in the absence of context, when in fact, most recognition relies on predictions from the ongoing context.

What follows is a description of the basis in the literature for the computational model we worked with. Note that, in order to build a working computational model, we have to simplify the architecture, leaving out some components (e.g., almost the complete dorsal stream), and making assumptions about others that are consistent with the biological literature but not necessarily dictated by it.

Early vision

Architectural description. The early vision system is modeled as a spatial selective attention function. It consumes raw images and top-down predictions, and produces image windows defined in terms of image positions and scales. The function should optimize stability in the sense that if the same object appears in two images at different positions and scales but from the same 3D viewpoint (and under similar illumination), the system should center attention windows at the same positions and relative sizes on the object.

Biological justification. The human early vision system is perhaps the most thoroughly studied system in neuroanatomy. After decades of study, we have detailed models of ganglion cell responses in the retina and the parvocellular, magnocellular and interlaminar layers of LGNd. Types of known orientation-selective cells in V1 include simple cells, complex cells, end-stopped cells, and grating cells, to name just a few. Other cells are sensitive to colors, disparities, or motions.

For all the discussion of edge sensitivity and feature maps, however, the products of early vision are spatial attention windows. The early vision system is retinotopic, which is to say that every cell has a fixed receptive field in the retina (although they also receive efferent inputs), and neighboring cells generally have neighboring receptive fields. Features in the early vision system are therefore kept in a 2D spatial format. Feature maps in the early vision system also cover the entire retinal image, creating essentially a series of image buffers. Moreover, the early vision system is almost the only part of the brain with this organization. As a result, it is a valuable resource: mental imagery recruits image buffers top-

down to reconstitute images from memory (Kosslyn, et al. 1999), and tactile input triggers V1 when subjects read Braille (Burton, et al. 2001).

Why would the brain compute any feature across the entire retinal image, when downstream processing is restricted to spatial windows selected by early vision? After all, it requires far fewer neurons to compute features over attention windows in LOC and/or MT than to compute them across the entire image. If we assume that human vision is efficient, the only features computed in the early vision system should be those used to select attention windows. This is why we model early vision as a spatial attention engine. There is one caveat: some dorsal pathway tasks such as ego-motion estimation rely on extra-attentional features which must be computed across the field of view. Motion features are also important for spatial attention, however, so the general rule still holds: all features computed in early vision are needed for attention.

We should be careful to distinguish among types of attention, particularly overt from covert attention, and spatial attention from feature-based or object-based attention. Overt attention refers to the movement of the eyes and head to fixate attention on a particular point in 3D space. Overt attention appears to be driven largely top-down, by a pathway through the frontal eye fields to the superior colliculus. (Superior colliculus also integrates bottom-up data from early vision, particularly motion data.) Overt attention occurs before a retinal image is acquired, however. We model early vision as selecting covert attention windows in the retinal image. These windows may or may not be foveal, and we do not know the average dwell time of covert attention or whether it is sequential or coarsely parallel.

We also distinguish spatial attention from feature-based or object-based attention. There is evidence that downstream visual processing may select data at the level of features or objects as well as spatial windows. This is not incompatible with spatial attention, but is not yet included in our model.

Direct evidence that spatial attention selects windows in terms of position and scale comes from Grill-Specter, et al. (1999), who used repetition suppression effects in fMRI to show that the input to LOC from the early vision system was unchanged when the stimulus was translated or scaled within a factor of 2. Oddly, the same study showed that human spatial attention does not impart rotational invariance, despite evidence from computational systems such as scale invariant feature transform, SIFT (Lowe 2004) that attention windows can compensate for image rotations as well.

Implementation of early vision. We implemented early vision as finding local maxima in multi-scale difference of Gaussians, DoG responses. This approach was first proposed by Koch and Ullman (1985), and has been refined over the years to form the basis of both NVS (Itti and Koch, 2000) and SIFT (Lowe 2004). Our implementation is based on neuromorphic vision system, NVS, but was modified to select scales as well as positions and to be less sensitive to image translations and scales (Draper and Lionelle 2005).

Whether DoG responses are a good biological model of bottom-up spatial attention in humans is debatable. Parkhurst, Law and Niebur (2002) and Ouerhani et al (2004) show better than random correspondence between DoG responses and human eye tracking data. Eye tracking, however, measures overt rather than covert attention, and Privetera and Stark (2000) show that almost any high-frequency feature has a better-than-random correspondence to eye tracking data. Recently, Kadir and Brady (2001) have proposed another bottom-up salience function based on local entropy.

Feature extraction in LOC

Architectural description. The lateral occipital complex is modeled as a feature extraction mechanism that converts spatial attention windows into feature vectors. The feature vectors are sparse and high-dimensional, and should capture the local geometric structure and to a lesser extent the color information in attention windows. The goal is to project the contents of attention windows into a high-dimensional feature space such that structurally similar windows will cluster.

Biological justification. The term lateral occipital complex denotes a large cortical region that spatially connects parts of the early vision system to the inferotemporal cortex. Although it has been studied for years, its exact boundaries in people and monkeys remains open to debate, as does the question of whether it is a single functional unit, two units, or possibly more. A general discussion of LOC can be found in Grill-Spector, Kourtzi and Kanwisher (2001).

Although the anatomy of LOC is unclear, its significance is not. A subject with bilateral lesions to LOC immediately developed visual form agnosia, a condition which left her unable to recognize even the simplest objects and shapes (James, et al. 2003). By measuring repetition suppression in fMRI, Kourtzi and Kanwisher showed that parts of LOC respond identically to an image of an object or its edge image (Kourtzi and Kanwisher 2000), even if its profile is interrupted (Kourtzi and Kanwisher 2001). Using a similar technique, Lerner et al (2002) showed that LOC responses are able to "fill in" gaps created by projecting bars over images.

These studies provide converging evidence for a view of LOC as computing structural features of attention windows, even in the face of geometrically structured noise. More recently, Kourtzi et al (2005) have shown that LOC is involved with learning shape descriptions for later use, and that it becomes even more active if the shapes being learned are partially disguised by complex backgrounds, possibly because it has to work harder. A study by Altmann, Deubelius, and Kourtzi (2004) suggests that LOC combines edge information with motion and disparity data and/or top-down predictions.

Confusing this picture somewhat is a study that suggests that at least part of LOC also responds to colors (Hadjikhani et al, 1998), although this may depend partly on the disputed boundaries of LOC. A study by Delorme et al (2000) suggests that feature vectors may include both structural and color information, but that the two are kept separate and that some subjects take advantage of color features while others do not. Also, the size of LOC and the fact that it becomes only diffusely active in fMRI studies of object recognition suggests that the feature vectors are high-dimensional but sparse.

Implementation. We implement LOC as a collection of parametric voting spaces, in the style of a Hough transform. The studies above suggest that LOC aggregates structural information, and behavioral studies by Biederman (1987) suggest that collinearity, co-termination, symmetry, anti-symmetry and constant curvature are particularly important structural features. We therefore created parametric representations of collinearity (defined over edges), axes of symmetry and anti-symmetry (defined over edge pairs), and of centers of curvature and termination (also defined over edge pairs). Edges and edge pairs from attention windows vote in these spaces, and the vote tallies form feature vectors. A single color histogram is used as a color feature vector. The final feature space representation is the concatenation of its structural and color feature vectors.

Although not used in this implementation, it should be noted that SIFT (Lowe 2004) uses parametric voting spaces as feature vectors, and that generalized Hough transforms (Ballard 1981) could be used to detect specific structures top-down.

Object recognition in inferotemporal cortex

Architectural description. The inferotemporal cortex is modeled as unsupervised clustering. It consumes feature vectors and produces view categories, which are groups of feature vectors that are similar in structure and color. View categories do not correspond to semantic object labels; semantic object classes may be divided across many view categories. Black cats, for example, do not look like calico cats, and the front view of cat doesn't look like its side view. View categories are viewpoint and illumination dependent, and semantic object classes may be further divided because of differences among instances (e.g. black cats vs. calico). Also, view categories typically correspond to parts of objects, since attention windows do not presuppose image segmentation.

Biological justification. The psychological literature makes a distinction between unimodal recognition and multi-modal identification. As defined by Kosslyn (1994), recognition occurs when input

matches a perceptual memory, creating a feeling of familiarity. Identification, on the other hand, occurs when input accesses representations in multi-modal memory. Thus we might visually recognize an object as being familiar before we identify it as a cat, at which point we know what it looks like, sounds like, feels like, etc.

Recognition and identification can become disassociated in patients with brain damage. Farah (2004) summarizes a collection of patients with associative visual agnosia. These patients cannot recognize objects, even though they can accurately copy drawings and describe the features of an object, suggesting that the early vision system and lateral occipital cortex are intact. These patients also show no deficits in identifying objects by other modalities; their ability to identify objects from language, sound and touch is unimpaired. They therefore demonstrate behaviors that are consistent with damage to a visual recognition module while the multi-modal identification module remains intact.

The opposite scenario is seen in patients with semantic dementia (2004). These patients retain basic recognition abilities in all of their senses, but loose the ability to form cross-modal associations, for example to associate visual percepts with auditory percepts or abstract concepts. The simultaneous loss of identification abilities across senses is consistent with a damaged identification system with intact sensory recognition modules. There are also cases of selective semantic dementia, in which patients are unable to identify specific classes of objects, for example living things. This is probably the result of damage to part but not all of the identification system.

Evidence that the inferotemporal cortex learns highly specific view categories comes from several sources. An fMRI study by Haxby et al (1999) suggests that IT responds differently to views of standard and inverted faces, while a study by Troje and Kersten (1999) goes further: people are expert at recognizing other people's faces head-on or in profile, but are only expert at recognizing themselves head-on, because that is how they see themselves in mirrors. Behavioral studies of face recognition suggest that we are faster and more accurate at recognizing faces illuminated from above than below (Bruce and Young 1998). Single-cell recordings from the inferotemporal cortices of monkeys suggest different responses to images of faces based on expression (Sugase, et al. 1999). Perhaps most tellingly, Tsunoda et al (2001) combined fMRI and single-cell recordings in macaques to probe IT responses to stimulus changes, for example removing part of a target or removing its color. Every significant change resulted in different cellular-level responses in IT. Tanaka et al (2003) showed that changes in orientation triggered different cells in macaque IT.

The evidence for highly-specific and appearance-based view categories combined with the separation of recognition from identification suggests that IT should be modeled as unsupervised clustering, while associative memories combine collections of category views with training signals to create cross-model object categories. This contradicts some other recent biologically-inspired models (e.g. Serre, et al. 2005), which learn to map from stimuli to labels at the level of the lateral occipital complex.

Implementation. We implement IT as a single layer of neurons trained by repetition suppression. The idea is that every neuron individually learns to divide feature space in two without dividing any densely populated portions of feature space (i.e. clusters). As a group, the neurons produce a binary code that identifies a view category. An alternative biologically-inspired unsupervised clustering model of IT has been proposed by Rodriguez, et al. (2004) and Granger (2006).

4.2 Significant changes to technical approach.

N/A

4.3 Progress against planned objectives.

One of the key objectives for this part of the work was "an analysis of the problems to be solved in each of the major processing phases in the brain during visual object identification … focusing on specifying the type of information that is sent". The architecture of the model we applied is based on the

literature of human vision, and implies both a set of processing phases for object recognition and constraints on the information sent between them. This architecture was described above. What follows is a summary of some of the major points learned with regard to this objective:

1. The selective attention system needs to identify focus of attention (FOA) windows. In practice, these attention windows need to specify both the location and the scale of area of interest, and the scale should be accurate to within a third of an octave. Beyond that, matching is rarely successful. Perhaps most importantly, "microsaccades" – i.e. small adjustments to the location and scale of an attention window – are highly useful. We discovered that when an FOA is matched to a category in the ventral stream, the prototype of the category can be used to refine the location and scale of the attention window. This in turn allows a better window to be extracted and categorized, confirming (or occasionally refuting) the original hypothesis.
2. Top-down predictions to the early vision system are, not surprisingly, very powerful. If the system knows what it is looking for, it can alter its salience function to detect that object. This both increases the detection rate and minimizes the errors in location and scale.
3. Structural features (meaning, in our context, shape based features based on non-accidental properties) and color features (implemented in our model with color histograms) performed about equally well in the feature extraction component of our model. Combining them performed only marginally better than either feature type alone, however, suggesting that top-down "signal" features (which we have not yet experimented with) may be very important.
4. The first stages of matching determine familiarity, not object identity. Image patches (i.e. the contents of attention windows) are grouped by their features into categories of image patches that look alike. Matching at this scale produces a sense of familiarity; failing to match may trigger novelty responses in the amygdala and perirhinal cortex. It does not, however, rise to the level of "object recognition". Instead, it allows the system to recognize that it has seen these surroundings before and that they are familiar. Grouping them into complex "objects" requires further stages.
5. Anomaly detection, which we operationalize as the detection of familiar objects in unfamiliar combinations, can be implemented successfully at the current level.
6. We tested three different implementations of grouping algorithms for familiarity. We tested the thalamocortical grouping algorithm of Granger against a neural-net repetition suppression algorithm and the traditional K-Means grouping algorithm. Somewhat to our surprise, Granger's algorithm significantly outperformed the other two. (We can provide more detailed results if desired.) We have therefore adopted Granger's algorithm as our standard grouping algorithm for this stage of processing.
7. Grouping algorithms by their nature have to model all of feature space. We altered our model of memory to learn high-dimensional manifold models of each specific category (a.k.a. aspect). This allowed us to detect and correct grouping errors. It also produced a better top-down signal to be fed back to the categorizer and to the selective attention mechanism.

4.4 Technical accomplishments.

As reported at the Bio Inspired Cognitive Architectures program (BICA) meeting in San Francisco, we applied the system to a sequence of 591 images of a toy artillery piece on a turntable; one of the images is shown in Figure 5. The system selected approximately 10 attention windows per image, converted the attention windows to parametric feature vectors and then clustered the resulting feature vectors into view categories. The average image windows for the eight most commonly occurring view categories are shown in Figure 6.

***Figure 5**. One of 591 images of a toy artillery piece on a turntable. The average rotation between images is a little less than 1.5°.*

In all eight cases, we can easily identify what part of the target or background the view category represents, and in all cases the categories are "pure" in the sense that every feature vector assigned to a category comes from the same target or background location. Different views of an object part generate different categories; for example, there are two view categories for wheels: one for nearly parallel projections, and another for wheels at more oblique angles (although the later was not one of eight shown in Figure 6).

Not all of the view categories in Figure 6 are equally meaningful. The first category, in fact, corresponds to the end of the shelf in the background behind the target. This was the most common category, because it never changed viewpoint and was visible in almost all the images. We need the semantic reasoning capabilities of the dorsolateral prefrontal cortex to infer that this category is not interesting, and top-down control to suppress it from being attended to in the future.

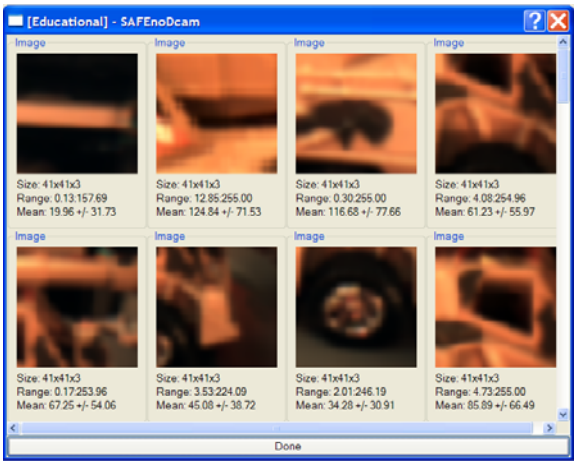

***Figure 6**. The average image windows for the eight most commonly occurring view categories.*

Also, although view categories correspond to particular points and viewpoints, not all images in which a specific view is visible get included in a category. For example, there are more side-views of wheels than were found and assigned to the 7th category in Figure 6. Often this occurs because the wheel was not attended to; sometimes it was assigned to its own singleton view category. These are errors, but we believe that top-down reasoning will correct most of them. For example, contexts that imply wheels

will generate top-down predictions that increase the frequency with which predicted view categories are found.

5. Conclusions

In this project we developed a taxonomy of top-down processes in vision, showed how it was embodied in our theory, and developed computational models that incorporate these distinctions. The first distinction we drew is between strategic versus reflexive top-down processes. Strategic top-down processes engage the information shunting subsystem in the frontal lobes and result in the modulation of processing in subsystems implemented in the ventral and dorsal streams. Examples of activities that engage this type of top-down processes are voluntary visual attention, working memory, and retrieval of visual information from long-term memory. In contrast, reflexive top-down processing is automatically engaged by bottom-up processing, and does not recruit the information shunting subsystem. An example of this type of top-down process is the modal completion of contours.

In addition, we also proposed that each of these two general types of top-down processing has different modes of operation. One mode consists of modulating the interpretation of outputs from processes; such modulation affects the sensitivity of processing (which corresponds to changing d', in signal detection parlance) or affects the decision criterion (which corresponds to β, in signal detection theory). Another mode consists of supplementing information that is present in a subsystem, such as via vector completion, which thereby can complete fragmentary patterns with stored information.

Although these distinctions are consistent with the available findings, not much extant evidence directly bears on them. One reason for the dearth of evidence lies in many technical limitations, such as the difficulty of establishing the precise time course of neural information processing in humans. Another, perhaps more interesting reason, is that researchers heretofore have not been thinking about top-down processing from the present perspective. Only after researchers begin to consider distinctions of the sorts we have proposed are they likely to turn their attention to collecting relevant data.

6. Deliverables

- Physical (printed) binder (*Biologically-Inspired Cognitive Architecture of High Level Vision: Supporting References*) containing references and abstracts of pertinent literature.
- Electronic database of the references in the binder (EndNote 9 [see http://www.endnote.com/] libraries for each of the modules indexed in the binder, plus a "fundamental papers" library); this is on a CD labeled "*ENLibs_BICA_Refs_Vision*".
- Annotated IMPER model software code (see Readme file on *Kosslyn_IMPER* CD for more information).

7. References

Altmann, C.F., A. Deubelius, and Z. Kourtzi, *Shape Saliency Modulates Contextual Processin in the Human Lateral Occipital Complex.* Journal of Cognitive Neuroscience, 2004. **16**(5): p. 794-804.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230*(4724), 456-458.

Ballard, D., *Generalizing the Hough Transform to Detect Arbitrary Shapes.* Pattern Recognition, 1981. **13**(2): p. 11-122.

Barone, P., Batardiere, A., Knoblauch, K., & Kennedy, H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J Neurosci, 20*(9), 3263-3281.

Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., & Haxby, J. V. (2001). A parametric fMRI study of overt and covert shifts of visuospatial attention. *Neuroimage, 14*(2), 310-321.

Biederman, I., *Recognition-by-Components: A Theory of Human Image Understanding.* Psychological Review, 1987. **94**(2): p. 115-147.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nat Neurosci, 2*(4), 370-374.

Bruce, V. and A. Young, *In the Eye of the Beholder: The Science of Face Perception*. 1998, New York: Oxford University Press. 280.

Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proc Biol Sci, 265*(1400), 1037-1044.

Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2005). *Latency of attentional modulation in ventral visual cortex.* Paper presented at the Society for Neuroscience, Washington DC.

Burton, H., et al., *Adaptive Changes in Early and Late Blind: A fMRI Study of Braille Reading.* Journal of Neurophysiology, 2001. **87**: p. 589-607.

Cave, K. R., & Kosslyn, S. M. (1989). Varieties of size-specific visual selection. *J Exp Psychol Gen, 118*(2), 148-164.

Chan, D., Fox, N. C., Scahill, R. I., Crum, W. R., Whitwell, J. L., Leschziner, G., et al. (2001). Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann Neurol, 49*(4), 433-442.

Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cogn Affect Behav Neurosci, 4*(2), 117-126.

Corbetta, M. (1993). Positron emission tomography as a tool to study human vision and attention. *Proc Natl Acad Sci U S A, 90*(23), 10901-10903.

Corbetta, M., & Shulman, G. L. (1998). Human cortical mechanisms of visual attention during orienting and search. *Philos Trans R Soc Lond B Biol Sci, 353*(1373), 1353-1362.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55-67.

Damasio, A. R. (1985). The frontal lobes. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology*. New York: Oxford University Press.

Delorme, A., G. Richard, and M. Fabre-Thorpe, *Ultra-Rapid Categorization of natural scenes does not rely on colour cues: A study in monkeys and humans* Vision Research, 2000. **40**: p. 2187-220.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (pp. 267-299). Amsterdam: Elsevier.

Draper, B.A. and A. Lionelle, *Evaluation of Selective Attention under Similarity Transformations.* Image Understanding 2005. **100**: p. 152-171.
Farah, M.J., *Visual Agnosia*. 2nd ed. 2004, Cambridge, MA: MIT Press. 192.
Feldman, J. & Singh, M. (2005). Theoretical note: Information along contours and object boundaries. *Psychological Review, 112*, 243–252.
Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex, 1*(1), 1-47.
Fox, P. T., Mintun, M. A., Raichle, M. E., Miezin, F. M., Allman, J. M., & Van Essen, D. C. (1986). Mapping human visual cortex with positron emission tomography. *Nature, 323*(6091), 806-809.
Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360*(6402), 343-346.
Fukushima, K. (1988). Neocognitron: a hierarchical neural network capable of visual pattern recognition. *Neural Networks, 1*, 119-130.
Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Res, 330*(2), 299-307.
Ganis, G., Schendan, H. E., & Kosslyn, S. M. (submitted). Neuroimaging evidence for object model verification theory: Role of prefrontal control in visual object categorization.
Ghosh, A. & Petkov, N. (2006). Effect of high curvature point deletion on the performance of two contour based shape recognition algorithms. *International Journal of Pattern Recognition and Artificial Intelligence, 20*, 913–924.
Girard, P., Hupe, J. M., & Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J Neurophysiol, 85*(3), 1328-1331.
Granger, R., Engines of the brain: The computational instruction set of human cognition. AI Magazine, 2006. 27(2): p. 15-32.
Gregory, R. L. (1970). *The intelligent eye*. London: Weidenfeld and Nicholson.
Grill-Spector, K., et al., *Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex.* Neuron, 1999. **24**: p. 187-203.
Grill-Spector, K., Z. Kourtzi, and N. Kanwisher, *The lateral occipital complex and its role in object recognition.* Vision Research, 2001. **41**: p. 1409-1422.
Gross, C. G., & Mishkin, M. (1977). The neural basis of stimulus equivalence across retinal translation. In S. Harnard, R. Doty, J. Jaynes, L. Goldstein & Krauthamer (Eds.), *Lateralization in the visual system* (pp. 109-122). New York: Academic Press.
Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol Rev, 92*(2), 173-211.
Hadjikhani, N., et al., *Retinotopy and color sensitivity in human visual cortical area V8.* Nature Neuroscience, 1998. **1**(3): p. 235-241.
Halgren, E., Mendola, J., Chong, C. D., & Dale, A. M. (2003). Cortical activation to illusory shapes as measured with magnetoencephalography. *Neuroimage, 18*(4), 1001-1009.
Haxby, J.V., et al., *The Effect of Face Inversion on Activity in Human Neural Systems for Face and Object Recognition.* Neuron, 1999. **22**: p. 189-199.
Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., et al. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc Natl Acad Sci U S A, 88*(5), 1621-1625.
Heeger, D. J. (1999). Linking visual perception with human brain activity. *Curr Opin Neurobiol, 9*(4), 474-479.
Hilgetag, C. C., O'Neill, M. A., & Young, M. P. (1996). Indeterminate organization of the visual system. *Science, 271*(5250), 776-777.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent colective computationsl abilities. *Proc Natl Acad Sci U S A, 79*, 2554-2588.
Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat Neurosci, 3*(3), 284-291.
Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol, 160*, 106-154.
Hubel, D. H., & Wiesel, T. N. (1965). Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. *J Neurophysiol, 28*, 229-289.
Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol, 195*(1), 215-243.
Hubel, D. H., & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *J Comp Neurol, 158*(3), 295-305.
Huxlin, K. R., Saunders, R. C., Marchionini, D., Pham, H. A., & Merigan, W. H. (2000). Perceptual deficits after lesions of inferotemporal cortex in macaques. *Cereb Cortex, 10*(7), 671-683.
Itti, L. and C. Koch, *A Saliency-based Search Mechanisms for Overt and Covert Shifts of Visual Attention.* Vision Research, 2000. **40**(10-12): p. 1489-1506.
James, T.W., et al., *Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study.* Brain, 2003. **126**: p. 2463-2475.
Kadir, T. and M. Brady, *Scale, Saliency and Image Description.* International Journal of Computer Vision, 2001. **45**(2): p. 83-105.
Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751-761.
Koch, C. and S. Ullman, *Shifts is selective visual attention: Towards the underlying neural circuitry.* Human Neurobiology, 1985. **4**: p. 219-227.
Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature, 399*(6732), 148-151.
Kosslyn, S. M. (1994). *Image and Brain*. Cambridge, MA: MIT Press.
Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., Ganis, G., Sukel, K. E., and Alpert, N. M. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science, 284*, 167-170.
Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1995). Identifying objects at different levels of hierarchy: A positron emission tomography study. *Human Brain Mapping, 3*, 107-132.
Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: a positron emission tomography study. *Neuroimage, 6*(4), 320-334.
Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York, NY: Oxford University Press.
Kosslyn, S. M., Thompson, W. L., Gitelman, D. R., & Alpert, N. M. (1998). Neural systems that encode categorical vs. coordinate spatial relations: PET investigations. *Psychobiology, 26*, 333-347.
Kourtzi, Z. and N. Kanwisher, *Cortical Regions Involved in Perceiving Object Shape.* The Journal of Neuroscience, 2000. **20**(9): p. 3310-3318.
Kourtzi, Z. and N. Kanwisher, *Representation of Perceived Object Shape by the Human Lateral Occipital Complex.* Science, 2001. **293**: p. 1506-1509.
Kourtzi, Z., et al., *Distributed Nueral Plasticity for Shape Learning in the Human Visual Cortex.* PLos Biology, 2005. **3**(7): p. 1317-1327.
LaBerge, D., & Buchsbaum, M. S. (1990). Positron emission tomographic measurements of pulvinar activity during an attention task. *J Neurosci, 10*(2), 613-619.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci, 23*(11), 571-579.
Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis, 20*(7), 1434-1448.
Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proc Natl Acad Sci U S A, 98*(4), 1907-1911.
Lerner, Y., T. Hendler, and R. Malach, *Object-completion Effects in the Human Lateral Occipital Complex.* Cerebral Cortex, 2002. **12**: p. 163-177.
Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Comput, 10*(4), 903-940.
Liu, A. K., Belliveau, J. W., & Dale, A. M. (1998). Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proc Natl Acad Sci U S A, 95*(15), 8945-8950.
Lowe, D.G., *Distinctive Image Features from Scale-Invariant Keypoints.* International Journal of Computer Vision, 2004. **60**(2): p. 91-110.
Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol, 77*(1), 24-42.
Luria, A. R. (1980). *Higher cortical functions in man*. New York: Basic Books.
McAuliffe, S. P., & Knowlton, B. J. (2000). Long-term retinotopic priming in object identification. *Percept Psychophys, 62*(5), 953-959.
McDermott, K. B., & Roediger, H. L., 3rd. (1994). Effects of imagery on perceptual implicit memory tests. *J Exp Psychol Learn Mem Cogn, 20*(6), 1379-1390.
McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron, 42*(4), 677-686.
Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb Cortex, 14*(11), 1256-1265.
Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Ann Neurol, 10*(4), 309-325.
Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann Neurol, 28*(5), 597-613.
Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci, 24*, 167-202.
Milner, A.D. and M.A. Goodale, *The Visual Brain in Action*. Oxford Psychology Series. 1995, Oxford: Oxford University Press. 248.
Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414-417.
Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331*(6151), 68-70.
Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature, 421*(6921), 370-373.
Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern, 66*(3), 241-251.
Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
Neisser, U. (1976). *Cognition and reality*. San Francisco: W.H. Freeman.
O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav Brain Sci, 24*(5), 939-973; discussion 973-1031.
Ouerhani, N., et al., *Empirical Validation of the Saliency-based Model of Visual Attention.* Electronic Letters on Computer Vision and Image Analysis, 2004. **3**(1): p. 13-24.

Palmer, S.E., *Vision Science: Photons to Phenomenology*. 1999, Cambridge, MA: MIT Press. 810.
Parkhurst, D., K. Law, and E. Neibur, *Modeling the role of salience in the allocation of overt visual attention.* Vision Research, 2002. **42**(1): p. 107-123.
Payne, B. R., Lomber, S. G., Villa, A. E., & Bullier, J. (1996). Reversible deactivation of cerebral network components. *Trends Neurosci, 19*(12), 535-542.
Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philos Trans R Soc Lond B Biol Sci, 360*(1456), 781-795.
Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu Rev Neurosci, 13*, 25-42.
Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *J Exp Psychol, 109*(2), 160-174.
Privitera, C.M. and L.W. Stark, *Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000. **22**(9): p. 970-982.
Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*(5313), 821-824.
Ress, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci, 3*(9), 940-945.
Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci, 19*(5), 1736-1753.
Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci, 2*(11), 1019-1025.
Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research, 179*, 3-20.
Rodriguez, A., J. Whitson, and R. Granger, Derivation and Analysis of Basic Computational Operations of Thalamocortical Circuits. Journal of Cognitive Neuroscience, 2004. 16(5): p. 856-877.
Rueckl, J. G., Cave, K. R., & Kosslyn, S. M. (1989). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience, 1*, 171-186.
Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiol Rev, 75*(1), 107-154.
Sandell, J. H., & Schiller, P. H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J Neurophysiol, 48*(1), 38-48.
Schacter, D. L. (1996). *Searching for memory*. New York: Harper Collins.
Seghier, M. L., & Vuilleumier, P. (2006). Functional neuroimaging findings on the human perception of illusory contours. *Neurosci Biobehav Rev, 30*(5), 595-612.
Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*(5212), 889-893.
Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science, 294*(5545), 1350-1354.
Sereno, M. I., & Tootell, R. B. (2005). From monkeys to humans: what do we now know about brain homologies? *Curr Opin Neurobiol, 15*(2), 135-144.
Serre, T., L. Wolf, and T. Poggio. *Object Recognition with Features Inspired by Visual Cortex*. in *IEEE Conference on Computer Vision and Pattern Recognition*. 2005. San Diego, CA: IEEE CS Press.
Squire, L. R. (1987). *Memory and brain*. Oxford: Oxford University Press.
Sugase, Y., et al., *Global and fine information coded by single neurons in the temporal visual cortex.* Nature, 1999. **400**: p. 869-873.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu Rev Neurosci, 19*, 109-139.

Tanaka, K., *Columns for Complex Visual Objects Features in the Inferotemporal Cortex: Clustering of Cells with Similar but Slightly Different Stimulus Selectivities.* Cerebral Cortex, 2003. **13**: p. 90-99.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol, 66*(1), 170-189.

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature, 401*(6754), 699-703.

Tootell, R. B., Hadjikhani, N. K., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., et al. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proc Natl Acad Sci U S A, 95*(3), 811-817.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognit Psychol, 12*(1), 97-136.

Troje, N.F. and D. Kersten, *Viewpoint dependent recognition of familiar faces.* Perception, 1999. **28**: p. 483-487.

Tsunoda, K., et al., *Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns.* Nature Neuroscience, 2001. **4**(8): p. 832-838.

Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition, 32*(3), 193-254.

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

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge: MIT Press.

Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B., Bakircioglu, M., et al. (2001). Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Res, 41*(10-11), 1359-1378.

VanRullen, R., Delorme, A., & Thorpe, S. J. (2001). Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing, 38*, 1003-1009.

Vezoli, J., Falchier, A., Jouve, B., Knoblauch, K., Young, M., & Kennedy, H. (2004). Quantitative analysis of connectivity in the visual cortex: extracting function from structure. *Neuroscientist, 10*(5), 476-482.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*(4654), 1260-1262.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Prog Neurobiol, 51*(2), 167-194.

Wilson, F. A., Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science, 260*(5116), 1955-1958.

Appendix

Publications

Ganis G, Schendan HE, Kosslyn SM. 2007. Neuroimaging evidence for object model verification theory: Role of prefrontal control in visual object categorization. *Neuroimage, 34*(1):384-98. [visual system, object recognition, top-down processing] [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=p]

Ganis, G, Kosslyn, SM. in press. Multiple mechanisms of top-down processing in vision. In S. Funahashi (ed.), *Representation and Brain*. Tokyo: Springer Verlag.

Ganis G, Thompson WL, Kosslyn SM. in press. Visual mental imagery: more than just "seeing with the mind's eye". In JR Brockmole (ed.), *Current Issues in Visual Memory*. New York: Psychology Press.

Meetings and Presentations

Name and purpose: DARPA BICA Technical Exchange Meeting (mandatory meeting of DARPA BICA investigators to discuss projects)
Dates: 8/5 – 8/9/2006
Location: San Francisco, CA
Attendees: Giorgio Ganis, Bruce Draper
Presentations: Report on progress of FA8750-05-2-0270.

Name and purpose: International Conference on Vision Systems (ICVS is a conference dedicated to making end-to-end vision systems work in the real world)
Dates: 3/21 – 3/24/2007
Location: Bielefeld, Germany
Attendees: Bruce Draper
Presentations: A Biomimetic Architecture

Name and purpose: Eleventh International Conference on Cognitive and Neural Systems (ICCNS is a conference organized by the Boston University Center for Adaptive Systems and is aimed at researchers of computational neuroscience, cognitive science and artificial intelligence).
Dates: 16/5 – 19/5/2007
Location: Boston, MA
Attendees: Giorgio Ganis
Presentations: FMRI evidence for object model verification theory: Prefrontal cortex and object categorization.

List of Main Symbols, Abbreviations, and Acronyms

AM: associative memory
AS: attention shifting system
AW: attention window
DoG: difference of Gaussians
fMRI: functional magnetic resonance imaging
FOA: focus of attention
IMPER: IMagery and PERception model
IS: information shunting system
LGNd: dorsal lateral geniculate nucleus
LOC: lateral occipital complex
MT: mediotemporal cortex
NVS: neuromorphic vision system
OPPS: object properties processing system
pIT: posterior inferotemporal cortex
SIFT: scale invariant feature transform
SPPS: spatial properties processing system
VB: visual buffer