Running Head: SEMANTIC MEMORY STRUCTURE

# Semantic memory structure:

## How what we know about the world is organized in the mind

Jeremy F. Gottlieb

Carnegie Mellon University

Advisors: Dr. Kenneth Kotovsky, Dr. Herbert A. Simon

Committee: Dr. Tai-Sing Lee, Dr. David Plaut, Dr. Marlene Behrmann

# ABSTRACT

A source of significant debate in psychology is the issue of how information is stored in semantic memory. The two primary frameworks are the Unitary Content Hypothesis, which holds that information is stored based on categories, and the Multiple Semantics Hypothesis, which holds that information is stored based on sensory modality. In a series of three experiments, I attempt to shed some light on which of these two frameworks is the most probable explanation of a number of phenomena associated with semantic memory. The results indicate that the Multiple Semantics Hypothesis is the most likely explanation. A further analysis of existing computational models shows that the “graded specificity” model (Plaut, 2002) provides the most comprehensive explanation of the available data.

# CHAPTER 1: INTRODUCTION

## 1.1 Semantic memory

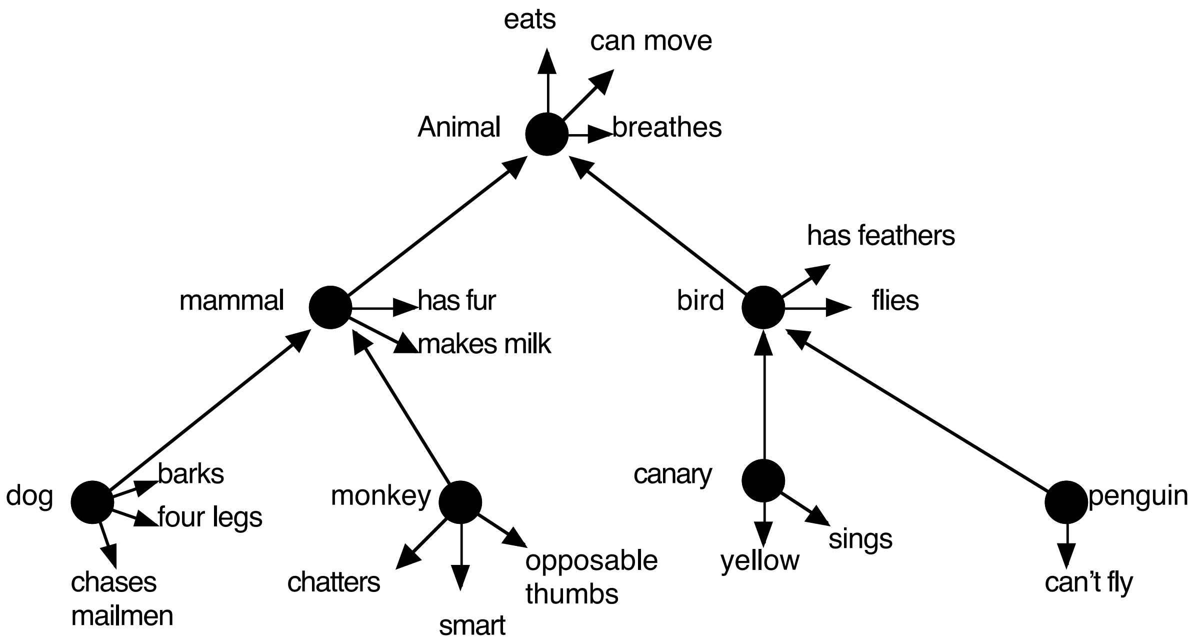
As we interact with the world, we acquire a large amount of information about our environment, particularly about the objects contained within our environment. This information comes to us from a variety of sources. Some of it we gather with our eyes, some with our ears, some with our hands, etc. This raw sensory information is then processed in a number of different brain areas to convert it into a representation that can be used by the brain as it carries out its various activities.

However, it should be clear that it is not the case that people are constantly re-extracting the same information from their environment. People are not, so to speak, reinventing the wheel over and over again. The activities of the brain that are using those representations primarily revolve around implementing the proper responses to stimuli in the environment. In order to develop those proper responses, a person must gain what would be philosophically called an “understanding” of the “meaning” of a stimulus or set of stimuli that exist in the environment at a particular time. The process of gaining this understanding involves the acquisition of the rather large amount of information referred to at the start of this section. Furthermore, this acquisition requires that information be stored in a readily accessible fashion.

Semantic memory is generically thought of as the storage system for all of the information we have about objects in the world, both informational and perceptual, such as names, visual properties like shape and color, or abstract information like “can fly” or “lives in Tahiti.” The question that this thesis is going to address is that of how this information from these disparate input modalities is organized in our semantic memory.

Everyone has the introspective experience that the information they know about the world is organized categorically. When we think about dogs, we seem to have ready access to all the various bits of information we have learned about dogs. Likewise with cars, chairs, trees, cats, or any other class of objects that you can think of. As a result, most early theories of semantic memory hypothesized that the information contained within it was also organized categorically. All the information about dogs would be stored within or in close proximity to the DOG concept node. Likewise for all other types of information.

**Figure 1.1: Collins & Quillian (1969) Semantic Network**



Perhaps the best known of these early theories is Collins and Quillian's (1969) semantic network model of semantic memory. They theorized that semantic memory was organized as a hierarchical, tree-like structure in which nodes in the network represented categories (see Figure 1.1). The relationships among categories would be largely reflected in the organization of this hierarchy. All the four-legged animals would share a common ancestor node reflecting the fact that they are all somewhat similar. Likewise for the various types of furniture.

However, at about the same time that the semantic network model was becoming popular, other researchers were asking a seemingly different question: How does mental imagery work? Most people have the experience of being able to conjure up pictures in their minds. You can close your eyes and picture the house you grew up in, your third grade teacher, your old dorm room, or even an elephant. What rapidly became the central question when trying to determine how mental imagery worked was – are the images real pictures in the mind? If they are, where do these pictures come from? Are they stored as pictures in our memory, or constructed from the abstract information contained in memory.

Paivio (1971) decided that the answer was that, in fact, people performing a mental imagery task were viewing real pictures in their minds (the Mind's Eye hypothesis). Furthermore, he theorized that these pictures were drawn straight from information stored in a visual code in semantic memory. He argued that the same also held true for auditory (such as imagining the sound of your mother's voice), taste, olfactory, and tactile information as well. Thus, in his Dual Code theory, he argued that semantic memory had at least two separate storage areas. One contained "imaginal codes" and was for perceptually based information, such as pictures of objects or places that could then be used when a person was performing imagery. The other contained "verbal codes" and was for more abstract, non-perceptually based information – primarily verbal information.

Thus was born a debate about the specific nature of how information is stored in semantic memory. Is it stored conceptually in one, abstract, all-encompassing format, or is it organized based upon what type of information (visual, verbal, etc.) is being stored? This thesis will attempt to shed some significant light on the issues raised in this debate through three means. One will be an analysis of data that can be brought to bear on this question from different disciplines that are not traditionally directly compared to each other. A second approach will be through a series of experiments designed to differentiate between the two major classes of models currently in existence. Finally, existing computational models of various semantic processes will be analyzed based upon the results of the previous two steps to determine the most likely candidates for accurately characterizing how, precisely, information is organized and stored in semantic memory.

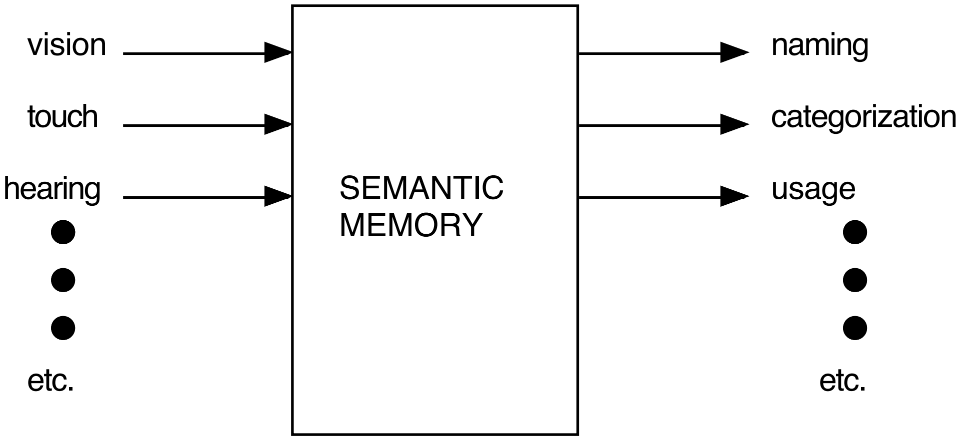
## 1.2 Competing hypotheses

Regardless of which theoretical perspective you subscribe to, everyone agrees that the overarching purpose of semantic memory is to take information from multiple input modalities – vision, speech, touch, etc. – and associate it with meaningful representations such that a wide variety of appropriate responses – naming, categorization, use, etc. - can be generated. The debate is about precisely how this mapping is accomplished. In this section, I briefly describe the two main theoretical frameworks used to account for this mapping.

### *1.2.1 Unitary Content Hypothesis*

As indicated above, there are two primary classes of theories attempting to answer the questions raised above. The first of these fall under the heading of the *unitary content* *hypothesis* (UCH). Specifically, this perspective posits that semantic memory is an abstract, amodal storage system where information is organized categorically. Every input modality can access all the information in semantic memory, and semantic memory provides information to every output modality (e.g. speaking, action, etc.) (see Figure 1.2) (Pylyshyn, 1973; Fodor, 1983; Caramazza, Hillis, Rapp, & Romani, 1990; Caramazza & Shelton, 1998; Riddoch, Humphreys, Coltheart, & Funnell, 1988).

**Figure 1.2: Unitary Content Hypothesis model of semantic memory**



More specifically, the key tenets are that information about the same category that enters the system from different input modalities will activate the same information about that category, regardless of what the modality of input was. In other words, regardless of whether you see a chair or hear the word "chair," all of the information you have about chairs will be activated and readily available. The second central tenet is – the reason why it is the case that the same information is activated, regardless of input modality, is because all of the information about a particular category is stored in the same mental (and, by extension, physical) location in the mind, and thus is closely and strongly interconnected. As a result, these theories match our introspective experience of there being no differences in our abilities to manipulate and process information across several different input and output modalities, such as naming or grasping an object we are looking at.

### *1.2.2 Multiple Semantics Hypothesis*

The counterpoint to the UCH is the *multiple semantics hypothesis* (MSH). Building upon Paivio's dual-code theory, the MSH holds that information in semantic memory is stored and processed the same way it is in the rest of the cognitive system – based on the sensory modality of the information. Under the MSH framework, semantic memory is conceived of as being divided into numerous semantic subsystems, based primarily on the sensory modality of the information stored in the subsystem (see Figure 1.3). Thus, there is a visual semantic system containing information about visual properties such as shape and color, a verbal semantic system containing information such as the names of objects, a tactile semantic system, etc. These subsystems are then linked together, allowing for information to pass between them (Shallice, 1987, 1988a, 1988b; McCarthy & Warrington, 1988; Warrington & Shallice, 1984; Beauvois, 1982).

The explanation for how we seem to understand our environment in terms of categories lies in how these subsystems are linked together. As with any type of simple neural learning (e.g., Hebbian learning), stimuli that occur together with a high degree of frequency are going to tend to be very strongly linked with one another. Thus, after you have seen a dog and heard it bark at the same time, the visual image of the dog and the sound of the dog barking will be strongly linked together in your memory. The way the semantic subsystems are linked together in the MSH framework is through just these types of associations. Information from different subsystems belonging to the same category (e.g., the image of the dog and the sound of it barking) are strongly linked together. Thus, the structure of a particular category is the result of what features of that category are linked together across semantic subsystems.

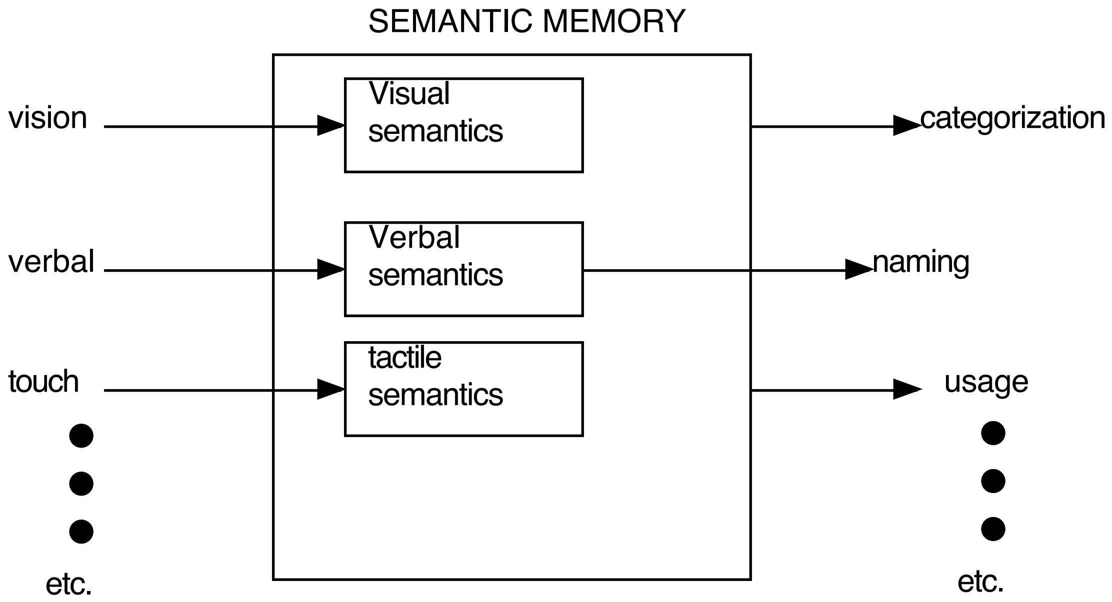


**Figure 1.3: Multiple Semantics Hypothesis model of semantic memory**

Relationships between different categories are then the result of how this information is organized within particular semantic subsystems. Thus, superordinate categories based on shared visual features (e.g., four-legged animals or purple things) ought to result from the fact that these items will have strongly linked visual semantic representations due to their high degree of visual similarity or association. The same would be true for rhyming words or taxonomic categories in verbal semantic memory.

The primary problem with the MSH theories put forth to date is that they tend to be underspecified. How many semantic subsystems are there? What, exactly, goes into visual semantics? It is a pretty safe bet that very few people have actually seen a purple elephant, but anyone can imagine what one looks like. Does everyone have an image of a purple elephant stored in visual semantics? All I have done here is verbally describe the purple elephant. Is the semantic representation you now have of the purple elephant stored in your visual semantic system, or your verbal semantic system? One can argue that we merely construct the image out of our stored visual representation of an elephant and our stored visual knowledge of the color purple. One can further argue that, even though I only verbally described the purple elephant to you, the way you understood it was to imagine it. By imagining the purple elephant, arguably you are creating a visual representation of it that can then be stored in visual semantics, where it presumably belongs. This is an especially compelling argument if one believes the functional-equivalence hypothesis of visual imagery mentioned above (Kosslyn, 1994).

Martin and his colleagues have been developing a theory that hinges upon the notion that there are as many semantic subsystems as there are perceptual and action systems. Results from brain imaging show, for example, selective activation of brain areas responsible for motion perception during the generation of action words and selective activation of areas responsible for color perception during the generation of color words (Martin et al., 1995; Martin et al., 1996; Martin & Chao, 2001). I discuss these findings in much more detail below (see section 1.3.1.5), but suffice it to say that these results do not provide compelling evidence for a modality-specific organization of semantics over one based on categories. In fact, much of the evidence being brought to bear on the question of how information is organized in semantic memory does not conclusively favor one set of theories over the other. Much of the data that exists can be explained by one of the many variations of each class of theories that have been proposed. The next section will explore the existing data that can be brought to bear on the question and how each framework, in a general way, attempts and/or fails to explain each piece of data.

## 1.3 The data

There are two major domains of work that were explored for the purposes of this review. There is a large literature in cognitive psychology about semantic processing tasks involving pictures, words, or both that are designed to investigate the structure of semantic memory. There is also research in cognitive neuropsychology studying brain damaged patients that show very selective deficits of semantic processing, such as modality-specific or category-specific deficits. The most apparent commonality between the two literatures is that they both strongly indicate that pictures and words are, at the very least, processed differently with regards to issues of access to semantic memory. Whether it be reaction time data from normal individuals or patient data, it becomes apparent that any model of semantic memory must take into account the fact that one does not find the same patterns of behavior when a participant is dealing with pictures on a semantic level as when the same participant is dealing with words.[1] Data from brain-damaged patients indicates that the same pattern holds with regards to somatosensation and audition.

---

[1] It should be noted that all of the experiments discussed are comparing pictures to written words. How this impacts conceptions of the MSH is addressed in more detail at the beginning of Chapter 2, but at this point it is simply worth noting that it can be argued that, at a semantic level, the distinction between the words being written versus oral may not be significant.

### *1.3.1 Data from cognitive neuropsychology*

#### 1.3.1.1 Split-brain patients

Split-brain patients show a couple of interesting patterns of behavior that relate to the current review. Patients are able to normally name and categorize pictures and words presented in the right visual field (left hemisphere). Yet, in general, when a picture is presented to the left visual field of a split brain patient, so that the information is being presented to his (presumably non-verbal or less verbal) right hemisphere, the patient is not only unable to name the picture, but is generally unable to report anything other than a bright blur or a vague sensation of having seen something, nor can they consciously perform any semantic tasks such as categorization with stimuli presented to the right hemisphere. Nevertheless, they do appear to have access to semantic information based on pictures presented to the right hemisphere. For example, Springer and Deutsch (1993) report on a couple of patients that display behavior commensurate with the semantic processing of an image without being able to name verbally the objects involved. For one patient, in the process of flashing pictures randomly to one hemisphere or the other, a nude picture was flashed to the right hemisphere. While the patient could not report that she had seen anything other than perhaps a brief flash of light, she blushed, giggled, and otherwise acted embarrassed. Presumably, in order to be embarrassed, the patient had to understand the stimulus enough to recognize that it was out of place in the experiment, and also a little indecent. This would seem to indicate that there was semantic understanding of the image at some level, even if verbal processing was missing.

They also report evidence from patients who were quickly presented with a picture of an object to the right hemisphere and then asked to pick out a related object with their left hand from a group of objects that they could not see. When a picture of a cigarette was flashed, participants were able to pick out a book of matches from the hidden objects, even though they could not report having seen the cigarette, nor could they name the book of matches they were holding in their left hand. The same thing could also be accomplished if they were given an object to hold in their left hand that they could not see, and then asked to use the same hand to pick out a related object.

What we can observe from the studies of split-brain patients is that semantic information seems to be able to be accessed even when name information is not available. What this implies is that, at the very least, name information and other semantic information are stored differentially. However, it may simply be the case that name information is stored in a separate lexicon external to semantic memory, while semantic memory itself still maintains a concept-based organization.

#### 1.3.1.2 Modality specific aphasias

Beauvois (1982) reports on a set of patients that she studied with a syndrome she called “optic aphasia.” These patients were characterized by the fact that they were severely impaired at naming objects that were presented to them visually, but were not impaired at naming them from any other sensory domain, such as touch. Furthermore, they were unable to point to the proper object when given its spoken name. They also exhibited semantic knowledge of the objects they were unable to name by being able to correctly mime how to use the objects and by being able to draw objects and complex scenes upon a verbal request. She also reports parallel cases of tactile and auditory aphasic individuals.

Beauvois theorized that these patterns of behavior were due to a severed connection between visual and verbal semantics. She bolstered that claim by having a patient who specifically had optic aphasia for colors perform a number of tasks. These tasks were divided into three groups:

- Verbal tasks where the stimulus, the response, and the intervening processes were strictly verbal and the patient was strongly encouraged to use only verbal strategies. An example would be asking the participant to respond with a color name linked to some abstract concept, i.e. “What color name is generally associated with envy?”
- Visual tasks were analogous to verbal ones. An example would be deciding if two colors were the same or not, or pointing out the properly colored object from a group.
- Visuo-verbal tasks were designed to require the use of both visual and verbal information. Typically the stimulus was in one modality and the response in the other modality, for example, naming a color or pointing to a color when told, “Show me what color a cherry is.”

What Beauvois found was that this patient, who was severely impaired at naming colors that were presented visually, performed at ceiling for both of the verbal tasks and for both of the visual tasks (almost 100% correct). However, for the four visuo-verbal tasks used, the patient performed at less than 50% on all of them. Furthermore, when the patient was given the same battery of tests five years later, his results were virtually the same. Her claim is that the ceiling performance on the modally pure tasks indicates that access to semantic memory from the visual and verbal modalities is intact. It is particularly noteworthy that access to semantic memory from visual input is intact since the deficit that optic aphasics show is one of naming objects that are presented visually. Thus, if access to semantic memory from both visual and verbal inputs is preserved, then the underlying deficit that causes optic aphasia must be one of information flow between visual semantic information and verbal semantic information, as evidenced by the patient's poor performance on the visuo-verbal tasks.

On the other hand, Riddoch and Humphreys (1987) argue that the underlying deficit in optic aphasia is a deficit of semantic access. They present patient JB, who exhibits classic optic aphasic symptoms. His naming from touch and verbal definitions, and his gesturing from vision or from spoken names are all relatively preserved (75% accuracy or higher) but his naming from vision is impaired (45% correct). He was perfectly fine at copying line drawings, at matching different views of objects, at an object-decision task requiring him to judge the reality of different line drawings, at picture-word matching, and at various other tasks judging his ability to access "structural" information from both visual and verbal inputs. However, he was impaired at naming line drawings of objects. Riddoch and Humphreys argue that the categories that J.B. is most impaired at are those that are most structurally distinct, and thus less likely to be affected by noise caused by the damage between a visual structural system and semantic memory. They further presume that the preserved gesturing that J.B. shows must be because gesturing is somehow related to visual structural information, rather than semantic information. As evidence they use the fact that J.B. got a knife and a fork confused in the picture-word naming task, but still gestured appropriately for each object.

The main problem with Riddoch and Humphrey's analysis is that it is unclear what the difference is between visual *structural* information and visual *semantic* information. Most of the properties that Riddoch and Humphreys insist belong in a visual structural system (e.g. shape, color, size, etc.) are properties that proponents of the MSH tend to argue belong in visual semantics. Furthermore, it is unclear where one draws the line between a structural system that does not have semantic content, and semantic knowledge about visual features, such as that canaries are yellow.

#### 1.3.1.3 Modality-specific effects in degraded store impairments

Lauro-Grotto, Piccini, and Shallice (1997) report on a patient suffering from semantic dementia who shows a significantly impaired ability to make proper semantic judgments from verbal inputs relative to visual inputs. She was unable to provide names from both spoken definitions and from pictures, was impaired at word-picture matching, but performed almost perfectly at sorting pictures by category. One could argue that this is the result of a generic semantic deficit, except that on a later test, the same patient was at chance for sorting words by category where the words corresponded to the pictures used in the picture sorting task. Furthermore, when the patient was asked to “mime the use or give any other kind of indication she could about the object” for 60 familiar household objects presented visually and verbally, she performed at chance for the objects presented verbally but well above chance for those objects presented visually.

### 1.3.1.4 Category Specific disorders

In apparent contrast with the modality-specific deficits that were discussed in the previous section, there have also been several reports of brain-damaged patients exhibiting category-specific disorders (see Forde & Humphreys, 1999, and Caramazza & Shelton, 1998, for review). Patients suffering from these deficits are generally impaired at processing semantic information (usually in the form of recognition) about either living things *vis a vis* man-made objects (e.g., Warrington & Shallice, 1984; Gainotti and Silveri, 1996) or they are impaired at processing man-made objects compared to living things (e.g., Warrington & McCarthy, 1983, 1987).

Warrington and Shallice (1984) report four patients who showed a significant impairment of comprehension of living things and foods vs. a relatively preserved understanding of inanimate objects. Patient J.B.R. in particular demonstrated a striking dissociating between his knowledge of living things and foods, on the one hand, and inanimate objects on the other across a wide variety of tasks, including verbal description, naming, picture/word matching, and miming. The deficit is very robust. J.B.R. was able to identify only 15% of visually presented living things vs. 85% for inanimate objects when they were matched for word frequency, and 20% of foods vs. 87% of inanimate objects. When matched on visual familiarity, patient S.B.Y. was able to identify only 13% of animals and 25% of foods, but 60% of inanimate objects. A more fine-grained analysis of the precise set of categories that J.B.R. was impaired at showed that they didn't fall very strictly into the living things/inanimate objects dichotomy. Along with being impaired at foods and categories that would fall under the broader class of living things, J.B.R. was also impaired at precious stones, diseases, and, most noteworthy, musical instruments (which would clearly belong to the inanimate objects category).

Warrington and McCarthy (1983) reported a patient who showed impaired recognition of inanimate objects vs. living things (foods, animals, and flowers). Using a matching to sample task, V.E.R. was able to recognize only about 60% of inanimate objects, but over 90% of foods and flowers and 85% of animals. Warrington and McCarthy (1987) found a similar deficit in a different patient, Y.O.T., whom they examined much more thoroughly to get as precise an accounting as possible of just which categories were preserved and which were impaired. In their first two experiments they found that Y.O.T. shared the same broad deficit as V.E.R. – she was impaired at recognizing inanimate objects relative to food, animals, and flowers.[2]

They then explored in more detail precisely which subcategories of inanimate objects might be impaired more than others. The first major subdivision that they found was between “small manipulable objects” and “large man-made objects.” Namely, using a spoken word/picture-matching task, Y.O.T. was relatively impaired at recognizing small objects that one uses with their hands, for example a fork, paper clip, pencil, or belt, versus larger objects such as an airplane, bus, tank, house, or train. (83% for food, 78% for large objects, 58% for manipulable objects). They then broke it down even further and found that Y.O.T.’s best performance was for categories such as animals (88%), occupations (88%), vegetables (83%), fabrics (80%), and accommodations (77%). Her worst performance was on categories such as kitchen utensils (47%), weather (44%), office supplies (39%), furniture (22%), and body parts (19%).

---

[2] It is worth noting that Y.O.T.’s deficit goes away when she does not have to respond as quickly, which will be discussed later.

Modality-specific deficits are particularly difficult to explain under the UCH due to its assumption of a category-based organization of semantic memory. Likewise, category-specific deficits are difficult to explain under the MSH due to its underlying assumption of a modality-based organization of semantic memory. To rectify this, Warrington and Shallice (1984) proposed what has since become known as the *sensory/functional theory* (SFT). The primary assumption behind the SFT is that different categories of information rely differentially on sensory features versus functional features when they are represented semantically. In particular, the argument is that the semantic representations for living things rely more heavily on sensory properties, specifically visual properties, while the representations for man-made objects rely proportionally more on functional information.

The argument is that this is because man-made objects are built to serve clearly defined functions, and those functions are an integral part of how the object is classified. On the other hand, living things do not have clear functions, and are generally interacted with in the visual modality rather than any action-based modality. Furthermore, outside of the field of biology, living things are categorized largely based on what they look like. Dogs are all dogs because they all look like other dogs and not like cats. Man-made objects, on the other hand, are categorized based on what they do. While form and function are often going to be closely related to each other (e.g., pens and pencils and other members of the category “writing utensils” look similar), it is still the object’s function that determines what category it will be placed into. A thin wooden object with a pointed end that can put marks on paper will be called a “writing utensil,” as opposed to a similar looking thin wooden object that cannot serve this function, and thus will only be called a "stick."

The SFT then takes a standard multiple-semantics viewpoint that there is one semantic subsystem containing visual information, and another containing functional information. What appears as a category-specific deficit is simply a result of damage to one of these two subsystems, on which living things and inanimate objects differentially rely for their semantic representations. Farah and McClelland (1991) implemented exactly this system in a PDP model (discussed in more detail in section 5.2.4 below).

As an example, think about how we differentiate a tiger from a leopard from a lion. The primary features used to differentiate these three objects are the knowledge that tigers have stripes, leopards have spots, and lions have neither. For the most part, all three objects look very similar except for these salient features, and all three of them serve essentially the same function (wild animals). On the other hand, one can think about how we differentiate between a dishwasher, an oven, and a clothes dryer. Once again, the three objects look very similar – boxy objects that have buttons and knobs on the front – but the functions that they serve differ from each other to a large degree, and it is these differences in function that allow us to differentiate between these three objects. Thus, if we have damage to the visual semantic system, we should still be relatively preserved at differentiating the three man-made objects because their visual features do not serve as primary a role in discriminating them. Likewise, if we have damage to the functional semantic system, discriminating the man-made objects will be relatively more difficult than discriminating the living things.

One of the primary criticisms of the SFT is that, while there might be some intuitive sense to theorizing that living things and inanimate objects rely differentially on visual versus functional information in their semantic representations, there is no evidence to support this view. On this point, Bloom (1996) has argued that the function an object serves is the primary information used to categorize man-made objects. Bloom overstates the case by arguing that not only is function central, but the function an object's creator intended for it to serve is, in fact, the overriding datum used to categorize man-made objects. Komatsu et al. (1997) demonstrated that this claim was too strong by demonstrating two things. First, they showed that, if standard objects are given ridiculous functions, people still classify those objects according to their standard category. More interestingly, they found that if ridiculous objects were given standard functions, people generally would not classify those objects according to those functions.

However, Malt and Johnson (1992), while also showing that Bloom's extreme conception of the role of function in classifying man-made objects was too strong, demonstrated that functional information is likely to serve as a large part of the core of artifact concepts. Thus, while functional features alone cannot exhaustively account for membership in inanimate categories, there is an argument to be made that they do play a significant role in categorizing man-made objects.

Saffran, Coslett, and Keener (2003) presented participants with both pictures and words and asked them to produce "the first word that comes to mind." The items were divided up along two dimensions: animacy and manipulability. In support of the notion that man-made objects rely more on functional features for their representations, they found that participants provided more action words for inanimate objects than animate objects (24.15% to 8.55%). Not surprisingly, they also found that more action words were generated for manipulable than non-manipulable objects (22.55% to 10.15%). However, it should be noted that even non-manipulable inanimate objects caused participants to generate more action words than manipulable animate objects (14.3% to 11.1%). Also of interest is that picture stimuli caused more action words to be generated than word stimuli in every condition (19.9% to 12.9% overall). Thus, there is at least some evidence that man-made objects, and particularly manipulable objects, rely more heavily on functional features for their semantic representations than natural objects.

The second primary criticism of the SFT is related to a lack of evidence regarding behaviors that patients with category-specific deficits ought to exhibit if the SFT is correct. In particular, patients with a category-specific deficit for living things also ought to be more impaired in understanding questions about visual semantic features, particularly for living things, than they are with regards to functional features. The most compelling evidence that supports this prediction is Gainotti and Silveri (1996). They present a patient, L.A., who shows a strong category-specific deficit in processing living things, foods, and musical instruments (like J.B.R. discussed above). They gave L.A. a task where she was supposed to name the object when presented verbally with a definition of it. L.A. was presented with two sets of sentences that described the exact same sets of objects either in terms of visual/perceptual features or in terms of functional features, for both animals and man-made objects. In essence, the type of features contained in the definition had a significant influence on L.A.'s ability to name the object if it was an animal in that she was worse at naming from the definitions based on visual/perceptual features, but there was no effect on L.A.'s naming ability for the man-made objects. Gainotti and Silveri interpret this as support for the SFT.

There is, however, a broader issue that seems to be overlooked in discussions of the veracity of the SFT. Namely, on what basis do we judge an object's function when we see it? Patients who exhibit category-specific deficits generally do so not just from verbal definitions or verbal descriptions, but also from visual presentations of objects. Yet, when presented with the objects visually, they still show a relatively preserved ability to name and otherwise identify the man-made objects that they are being visually presented with. If the argument is that they are able to accurately process man-made objects based on the functional features of the objects, and that they are impaired at identifying living things because their visual semantic subsystem is damaged, on what basis are these patients making judgments about any given object's function? If they are not deriving the function of the object based on the visual information contained in the picture of the object, from what information are they in fact extracting this functional information from?

If, however, they *are* deriving the function of an object based on the visual information being presented, then clearly the only basis on which proponents of the SFT can claim that the functional information can exist in its own functional semantic subsystem is to theorize that visual information must also have a direct input into this functional semantic subsystem. Otherwise, if one does not presume that the functional semantic subsystem receives direct visual input, then one would expect that damage to the visual semantic system would also impair a patient's ability to extract the relevant function for man-made objects, and thus one would expect those patients to be equally impaired at identifying man-made objects as at natural things. Furthermore, under this scenario, the functional semantic subsystem would not be able to be an entirely separate semantic system, but rather would be one that is subservient to the visual semantic system (or at least strongly tied to those visual features that relate to function, such as size and shape). Thus, under the SFT, it clearly cannot be the case that the visual component inherent in extracting functional information from the presentation of a picture of an object is in any way dependent upon the visual semantic subsystem. In fact, in Farah & McClelland's (1991) connectionist implementation of the SFT, they included direct connections between visual input and the functional semantic subsystem to help their model better map the existing data (see section 5.2.4 below).

If we follow this through to its logical conclusion and presume that the functional semantic subsystem receives input directly from the visual modality, then this proves problematic for another syndrome that the multiple-semantics model was designed to address and that we have discussed already – namely optic aphasia. Recall that optic aphasia was explained as damage to the connections between visual semantics and verbal semantics. Now, in order to explain category specific deficits, proponents of the multiple-semantics hypothesis must acknowledge that for the SFT to be true, the functional semantic subsystem must receive direct visual input. In that case, what is there to prevent the verbal semantic subsystem from also receiving direct visual input?

Clearly it is the data that prevents this connection from existing, as then we should not see patients with optic aphasia, but this simply begs the question of whether there is any systematicity to determining what sensory modalities have direct visual inputs to what semantic subsystems. A second counter-argument would simply be to argue that optic aphasia is the result of damage to two transmission pathways – those from visual semantics to verbal semantics, and also those from visual input to verbal semantics – but then one needs to be able to explain why those two neural pathways should be close enough together that they would be damaged together in all instances of this particular syndrome.

Furthermore, if there can be a direct connection from visual input to functional semantics and verbal semantics, why can't there be one to every semantic subsystem? Now the complication is that optic aphasia needs to result from damage to the visual input pathways to every semantic subsystem *except* visual semantics, and damage between visual semantics and verbal semantics, but not from any other subsystem into verbal semantics. Thus, without a systematic way of being able to determine just what semantic subsystems receive direct visual input, it becomes difficult to reconcile the multiple-semantics explanation of optic aphasia with the multiple-semantics explanation of category-specific deficits (the SFT).

Caramazza and Shelton (1998) offer an alternative explanation for category-specific deficits based on their belief that the primary "reductionist" accounts of these deficits, primarily the SFT and the OUCH model (discussed in more detail in Chapter 5 below), are both inadequate for explaining the precise patterns of categories that can be damaged in different patients. They propose a model whereby there are essentially three different semantic subsystems based on categories rather than on modality – one for animals, one for plants (mainly represented by fruits and vegetables) and one for "other" objects. They put forth an evolutionary argument to support this in which they claim that it was evolutionarily advantageous to have a separate system for processing animals because, once upon a time, it was very important for people to be able to easily and accurately distinguish between animals that were a threat (i.e. predators) and animals that were not a threat and/or that could be eaten as food. A similar argument can be made for plants. Finally, that leaves basically everything else for which there isn't an evolutionary reason for them to be processed differently – namely, the (relatively) recently developed man-made objects.

The primary evidence that Caramazza and Shelton (1998) use to support this model is their analysis of a patient, E.W., who had a category-specific deficit for just the category of animals. She was very impaired at naming animals (16/47) but was very good at naming fruits (12/12) and vegetables (12/12), as well as categories of man-made objects such as kitchenware (16/17), clothing (27/27) and furniture (15/15). E.W. also showed a significant deficit in processing animals versus other objects for a real/unreal object decision task, showed no deficit for processing visual versus functional features of both animal and non-animal objects. Finally, she is impaired at processing auditory information about animals as well as visual information.

Thus, Caramazza and Shelton argue, E.W.'s deficit cannot be the result of damage to a visual semantic subsystem because then there should also be impairments for processing fruits and vegetables as are normally found in patients with category-specific deficits for living things. Furthermore, if the SFT were correct, then E.W. should have shown relatively greater impairments at processing visual and other sensory information versus functional and other associative information, at least for animals if not also for non-animals. Finally, they argue, the veracity of the SFT hinges crucially on whether or not categories of knowledge can be strictly differentiated based upon the information used to represent those categories. As noted above, the evidence on this point is murky at best, and Caramazza and Shelton claim that the evidence is effectively non-existent if one includes associative properties with functional properties.

Thus, category specific deficits arise because, in fact, the semantic system is organized categorically. Category-specific deficits for living things are a result of damage to the specialized systems for animals and plants. A deficit for living things is the more common deficit precisely because there are specialized semantic systems for living things that do not exist for man-made objects. Deficits for man-made objects result from damage to the more general semantic system that processes everything that does not fall into the two evolutionarily determined subsystems.

We will discuss the Caramazza and Shelton model in more detail in Chapter 5 below, but one question deserves to be raised: How would such a system account for the modality-specific effects observed in both normal people and brain-damaged patients? Specifically, how can this model account for both modality-specific deficits such as optic aphasia as well as the results discussed above showing that pictures and words are processed differently even by normal people trying to perform semantic operations? The most obvious solution is to propose the existence of a system that contains verbal information, or at least name information, that some might call a lexicon. Optic aphasia can be explained by the existence of damage in the pathways from visual input to the lexicon. The other results can be explained merely by requiring that verbal inputs access the lexicon before they can access other parts of semantic memory.

However, this is not entirely satisfactory as it hinges on the development of a separate memory system solely for verbal information. Arguably such a system exists as there is ample evidence to support the notion of some sort of lexicon or verbal memory existing in people's long-term memories. But, with regards to the Caramazza and Shelton explanation of category-specific deficits, their explanation hinges on the fact that there are long-standing evolutionary reasons for why we would have category-specific semantic subsystems to store and process information about animals and plants, but not for man-made objects. Man-made objects have not been around long enough. However, depending upon your definition of "man-made object," arguably they have been around longer than words. Thus, while it is clear that language processing is special and occupies its own special part of the cognitive system, a justification for why that should be the case evolutionarily as opposed to the lack of a specialized system for processing man-made objects would bolster this argument.

#### 1.3.1.5 Functional brain imaging

In more recent years, many researchers have attempted to explore how information is organized in semantic memory by using fMRI or PET to scan what parts of the brain are activated as people perform various semantic tasks. Martin, Wiggs, Ungerleider, and Haxby (1996) used PET to scan the brains of normal people while they performed a picture naming task where the pictures were either tools or animals. What they found was that there was category-specific activation of areas of the brain that depended on the characteristics of the object being named.

In particular, there was selective activation of early visual processing areas (specifically the left medial occipital lobe) when animals were being named. Arguably, this could simply be due to the animal pictures being more difficult to process a possibility supported by the fact that animals were named 54ms slower than tools and had a higher error rate (16,7% to 10.0%). Martin et al. claim that this cannot be the case as there was heightened activation of the left medial occipital lobe, even when compared to visually complex nonsense objects. On the other hand, the areas of the brain selectively activated when participants were naming tools were more strongly associated with an area responsible for generating action words (left middle temporal gyrus) and an area that becomes active when you ask people to imagine grasping an object (left premotor cortex).

Based on Martin et al. (1996), there is clearly a distinction between the areas of the brain that become active when naming tools versus those that become active when naming animals. Furthermore, the areas that become differentially active are related to action and motor processing when tools are being named and visual processing when animals are being named.

On the face of it, this would appear to provide clear support for the MSH, and for the sensory-functional theory in particular. However, there was no condition in which visual properties of tools or functional properties of animals were tested. Thus, the strongest conclusion that can be drawn from these data is that animals and tools are stored in different areas of the brain, and that naming them at least seems to elicit activation in areas related to vision in the case of animals and function in the case of tools. This does not, however, rule out the UCH as an explanation for this data. It could simply be that the categorical representations of animals and tools cause spreading activation to these brain areas, while it still is the case that semantic memory is organized along categorical dimensions, rather than by sensory modality.

However, another study by the same lab (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995) showed that different areas of the brain become active depending on whether participants are asked to generate color information versus when they are generate action information in their responses. Using only man-made objects, Martin et al. (1995) presented participants with line drawings and asked them to name the object, give a color associated with the object, and tell them what the object could be used for.

Using the object naming condition as a baseline, they found that generating color information selectively produced bilateral activation in the fusiform gyrus of the temporal lobe. This area is immediately anterior to the part of the brain known to be responsible for color perception. Producing action words selectively produced activating in the left posterior middle and temporal gyri, as well as in Broca's area. The areas in the temporal gyrus are immediately anterior to the area of the brain responsible for motion perception. They then did the same experiment, but presented participants with words instead of pictures. They found that the same brain areas were selectively activated for color and action words, even when the input stimulus was a word instead of a picture.

What Martin et al. (1995) provides is evidence that individual features of objects are stored not in one anatomical location based upon the object (i.e. in a categorical fashion) but rather that they are stored based on what the type of feature is, such that particular types of features (e.g., color) are stored in close proximity to those areas of the brain responsible for processing the type of information represented by that feature.

On the other hand, it is difficult to know exactly what Martin et al. (1995) are tapping into in these experiments. Specifically, it is unclear whether the activated areas that they are observing are being activated because they are being used in the generation of features of objects, or because they are the areas where the concepts representing those features are stored.

The example they give of what their task requires is that of a participant being presented with a drawing of a pencil and responding “yellow” for the color and “write” for the action. However, “yellow” and “write” are words that represent concepts in and of themselves, with semantic representations. Arguably, what Martin et al. (1995) are measuring is not activation of particular features of the concept PENCIL, but rather activation of the concepts YELLOW or WRITE depending on which response the participant is supposed to give. Regardless of whether one subscribes to the UCH or MSH, one expects that information belonging to the same category is going to be very strongly linked together. Therefore, one would expect that, in a normal person, once one piece of information about the category is activated, such as the name, that activation should spread to the other information related to the category such that both functional and color features would both be readily available. It is only once a person has to pick one of those features to focus on, and, in particular, to generate the *name* of a particular feature that the brain area associated with that feature becomes differentially activated. However, the generation of that name presumably requires accessing semantic information contained in the conceptual representation of that feature. Thus, while it may be the case that color *concepts* and action *concepts* are stored in anatomical proximity to color and action processing areas, this does not provide conclusive proof that object *features* are stored in such a fashion.

In support of this notion, Tranel, Kemmerer, Adolphs, Damasio, and Damasio (2003) demonstrated that patients with damage to areas known to process motor information (e.g., left premotor cortex, left parietal region, and in white matter under area MT, which is known to be involved in motion perception) also showed deficits in retrieval of semantic knowledge about actions, such as the relative noise levels of different actions, which ones are more tiring, which ones require specific movements, and the relative time to complete various actions.

These areas are similar to those that Martin et al. (1996) found to be active when participants were naming pictures of tools (particularly the left premotor cortex), but different from those that Martin et al. (1995) found to be active when participants were generating action features of man-made objects (specifically, the left temporal gyrus).

Thompson-Schill, Aguirre, D'Esposito, and Farah (1999) provide fMRI data that is purportedly in support of the SFT explanation of category-specific deficits. They focused on the left fusiform gyrus, which they claim is known to respond preferentially to visual semantic information. Their hypothesis was that, if the SFT was correct, then they would expect this area to be selectively activated when information about living things is being processed, regardless of whether that information is visual, due to the fact that the representations of living things rely heavily on visual information. Furthermore, they expect that the left fusiform gyrus should also be selectively activated when visual properties of man-made objects are being processed, since it presumably is a locus for visual semantics, but not when participants are processing non-visual features of man-made objects. This is essentially the result that they found.

However, Caramazza (2000) argues that the data they collected does not support their theory for a number of reasons, most damning of which is that patients with category-specific deficits rarely have brain damage to the left fusiform gyrus. Given that fact, he finds it difficult to understand how one can claim that this area is the locus for visual semantics. He argues that what Thompson-Schill et al. found was an area that plays a role in mental imagery by claiming that the task used, involving questions about complex visual and conceptual information, could potentially most easily accomplished using imagery. However, what this fails to explain is why activation of the left fusiform gyrus for processing non-visual properties of man-made objects is significantly less than for processing non-visual properties of natural objects. Caramazza even uses an example question ("Can headphones play stereo music?") to support his imagery argument that comes out of the set of questions that did *not* provide heightened activation of the left fusiform gyrus. Thus, while it certainly is not clear that Thompson-Schill et al.'s results support as strong a claim about the role of the left fusiform gyrus in semantic representations as they might like to make, it is also not as clear that their results are as easily dismissed as Caramazza would like.

Finally, Tyler, Bright, Dick, Tavares, Pilgrim, Fletcher, Greer, and Moss (2003) did a PET study looking at what areas of the brain became active when participants performed a category matching task on four different categories – animals, fruits and vegetables, vehicles, and tools. Presumably, if the SFT is correct, one would expect to see different areas become active when categorizing animals and fruits and vegetables versus vehicles and tools, as per Martin et al. (1996). Instead, what Tyler et al. found was that the same brain regions were activated regardless of the category being processed, with the exception of an area in the right occipital lobe that was selectively active for animals. Even looking specifically at left premotor areas, they were unable to replicated the Martin et al. (1996) results.

#### 1.3.1.6 Summary of cognitive neuropsychology evidence

To summarize, the cognitive neuropsychology literature provides the following phenomena that must be accounted for:

- Split-brain patients who are unable to name pictures presented to the left visual field can still show evidence of intact semantic knowledge of the pictures they were unable to name (Springer & Deutsch, 1993).
- Brain damaged patients can experience modality-specific disorders of naming while having intact semantic knowledge and naming knowledge from unimpaired modalities (Beauvois, 1982; Shallice, 1987, 1988a).
- Brain damaged patients can experience modality-specific priming of their impaired responses (Warrington & Shallice, 1979).
- Brain damaged patients can also experience category-specific disorders that impair their knowledge of either living things or man-made objects with respect to the other category (Shallice, 1987, 1988a; McCarthy & Warrington, 1988).
- Semantic dementia patients show significant modality-specific effects of the degradation of their semantic knowledge (Lauro-Grotto, Piccini, and Shallice, 1997).
- Brain-imaging data is very suggestive of a semantic system organized based on the modality of information rather than based on the category, but is still not entirely conclusive due to uncertainty about whether these studies are finding activation of concept features, or of separate modality-specific concepts (Martin et al., 1995; Martin et al., 1996; Tranel et al., 2003).

### *1.3.2 Cognitive data*

#### 1.3.2.1 Naming vs. categorization

Potter and Faulconer (1975) demonstrated that people process pictures and words differently when given different semantic tasks. They found that people were faster to name (i.e. read) words than they were to name pictures, but that participants were faster to categorize pictures than words. This has been found to be a relatively robust effect (see Guenther, Klatzky, & Putnam, 1980; Seifert, 1997; Theios & Amrhein, 1989).

Both naming and categorization are generally thought of as tasks involving semantic memory. Thus, this effect appears to provide clear evidence that pictures and words are processed differently at the semantic level. One can see that it is not an effect at the sensory level because in all of these experiments both the pictures and the words are initially being processed visually. Nor is it an effect of processing at the pre-semantic level, because then one would expect either pictures or words to be consistently slower for both semantic tasks, rather than the double dissociation pattern that is found, due to the fact that prior to the semantic level processing is modality based (Theios & Amrhein, 1989; Glaser & Glaser, 1989; Glaser, 1992).

However, there are two significant problems with this class of studies. The first problem has to do with the nature of the naming task. “Naming” a word consists merely of reading it aloud, which it can be argued is not a task involving semantic memory but merely one of mapping the letters onto the phonology required to pronounce the word (see Figure 1.5). Theios and Amrhein (1989) argue that picture naming actually requires the extra steps of extracting the meaning of the picture and then finding the name that the direct grapheme-to-phoneme translation of word reading does not. Seifert (1997, Experiment 1) provides evidence to support this notion. Participants' average latency for word naming was 385ms, as opposed to 597ms for picture naming, 666ms for picture categorization, and 727ms for word categorization. Thus, word reading was a full 200ms faster than any of the other tasks requiring the use of semantic memory. One could argue that this implies that word reading, at least in these tasks, bypasses semantic memory and uses a direct grapheme-to-phoneme mapping that is, in normal adults, highly practiced (see Figure 1.5). Fraisse (1969) even presents the peculiar example of the symbol O. When this symbol was presented to participants in an object naming task, they named it as a circle in 619ms and as a zero in 514ms. When it was presented in a reading task, though, they read it as "oh" in only 453ms.

This explanation of the latency difference between word reading and word/picture categorization and picture naming falls under the purview of the “dual-route” models of word reading (Coltheart, Curtis, Atkins,& Haller, 1993). These theories claim that there are two possible cognitive routes that contribute to word reading in normal people (see Figure 1.4). The first route primarily involves a direct grapheme-phoneme rule system to generate pronunciations for regular words and non-words. Regular words are those that have pronunciations that can be generated by rules, which are learned and stored in the grapheme-phoneme rule system. These are words like BAKE and RATE and BARE, which all conform to the rule that if a word ends vowel-consonant-e, the vowel is long and the e is silent. This route is also used to generate pronunciations for non-words that conform to these rules, such as RABE, BAZE, or FIME. The second route uses the semantic system to recall stored pronunciations of exception words, such as HAVE and PINT, that do not follow the most rules for the most common words. It is the first pathway that is most relevant to the discussion at hand. Specifically, this first route exists independent of semantics. Pronouncing regular words does not require access to semantic memory, and thus occurs more quickly than pronouncing exception words, which requires access to semantic memory.

There is a competing theory which says that word reading is the result of interactions between two pathways – a phonological pathway and a semantic pathway (Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996). They argue that semantic information is a necessity for skilled word reading, and point particularly to the role of semantics in pronouncing ambiguous homophones such as WIND and READ, where the context in which the word is presented determines the pronunciation. Figure 1.5 shows how this network is structured.

While it is beyond the purview of this thesis to explore the model in great detail, the important thing to note is that both of the models manage to account for large amounts of the data on word reading without requiring feedback from or access to the semantic system (Seidenberg & McClelland, 1989; Plaut et. al, 1996, simulations 1-3). Thus, while semantics might be necessary for some aspects of word reading such as properly pronouncing homophones, it does not appear to be necessary for reading regular words. Regardless, a case can be made that word reading of the type required by these tasks at most simply requires feedback from semantics, rather than full access, and very likely does not require semantic access at all. Thus, the initial point about these tasks not necessarily capturing semantic organization still stands.

On the other hand, experiments using a Stroop like paradigm with words and pictures presented together provide evidence that access to names is faster from words than from pictures. Rosinski, Golinkoff, & Kukish (1975) showed that if an incongruent word is superimposed on a picture, naming of the picture is slower than if a nonsense syllable is superimposed. Conversely, naming of the word is unaffected by the presence of an incongruent picture. One can argue that retrieving the name information from semantic memory is faster for words than for pictures, and thus the retrieval of the incongruent word interferes with the retrieval of the name of the picture by activating irrelevant semantic information, meaning that the semantic information related to the picture, including its name, must compete more to be noticed.

In this particular experiment, there is one primary piece of evidence that indicates that the effect is a semantic one. First, if the retrieval of the word name was not a semantic retrieval, it is unclear why there should be more interference from incongruent words than from nonsense syllables. If the words are being named solely from an automatic grapheme-to-phoneme mapping that, even when the task is to name the picture, bypasses semantic memory, then presumably the source of the interference observed should be the fact that the first set of phonemes activated for speech production (as the output of this grapheme-to-phoneme pathway) are those for the word, and the phonemes for the name of the picture must overcome this activation before they can be produced. Thus, reading a nonsense syllable that is still easily pronounceable should also have an interfering effect on the production of the picture name.

One response is to point out that Rosinski et al. (1975) did not have a condition where participants were expected to name the picture without any letter string present, so it cannot be said whether the nonsense syllable still interferes with the naming of the picture, but just to a lesser degree. If this were, in fact, the case then one could simply argue that real words generate a much higher level of activation for the production of phonemes than nonsense syllables do simply because of their higher frequency of use, and thus this higher level of activation takes longer for the picture name to overcome. Ehri (1976) has also demonstrated that picture naming in the presence of an incongruent word is delayed relative to a condition where no word is present, but there is still no direct comparison of a condition with a nonsense syllable with a condition with no word at all.

The second problem with studies comparing categorization to naming across words and pictures has to do with the nature of the categorization task. The vast majority of these studies use one of two methodologies. In one set of studies, a categorization task is used where they are measuring the time a participant takes to make a yes or no decision as to whether the presented stimulus belongs to a particular category, such as living things (Guenther et al. 1980). In the other set, participants are making forced choices between two visually distinct categories, such as vegetables and tools (Seifert, 1997). In the latter case, participants' speed at categorizing pictures could simply be due to the ease of discriminating visually between the two dissimilar categories. Vegetables and tools do not share many common visual features, and thus the task may boil down to verifying the presence or absence of one or two distinct features, rather than recognizing the picture. In the former case, it is not clear whether the processes involved in deciding whether an object belongs to a given, broadly defined category are the same processes involved in generating the name of the category that an object belongs to. One could conceive of a situation whereby participants are actually simply generating a list of the members of the target category and either storing that list in working memory or generating a retrieval structure to store the list. Then, when they are presented with an item, they simply compare it to the list and see if they find a match.

However, in one study that actually required participant to generate the name of the category (Smith & Magee, 1980), participants were still faster at categorizing pictures than words. They also found in a Stroop-like experiment that when a word was superimposed on a picture, if the task was to categorize the word, an incongruent picture would interfere with categorization. However, if the task was to categorize the picture, an incongruent word had no effect, implying that the pictures are accessing category information faster than the words.. It has also been found that participants are faster to decide whether two stimuli belong to the same category when the stimuli are pictures as opposed to words (Pellegrino, Rosinski, Chiesi, & Siegal, 1977; Rosch, 1975). This supports the results from the categorization task normally used and indicates that people do, in fact, have faster access to categorization information from pictures than they do from words.

By itself, the finding that pictures have faster access to categorization information than words does not necessarily mean much. One possibility is simply that pictures are processed faster at a perceptual level. However, there are researchers (Potter & Faulconer, 1975; Theios & Amrhein, 1989) who make the claim that sensory encoding of pictures and words takes roughly the same amount of time. Theios & Amrhein (1989) claim that one significant problem with many of the word/picture categorization and naming experiments is the fact that the words and the pictures are presented at significantly different sizes, with the pictures sometimes being up to 10 times larger than the words, making them easier to perceive and thus faster to process. They conducted an experiment similar to Posner and Mitchell (1967) in which participants were presented with a stimulus that could be one of three sizes - small, medium, and large. They are then presented with a second stimulus and must decide as quickly as possible whether the second stimulus is the same as the first one, regardless of their relative sizes. Because participants were required to respond "same" regardless of size, they could not rely upon a purely physical match to make their response.

They found that there were no differences between trials containing two pictures and trials containing two words (574 and 585ms respectively for "same" trials; 643 and 646ms respectively for "different" trials), which they then argued meant that the early visual processing of both pictures and words was the same when stimuli were equated for size, brightness, contrast, and other visual properties. Support for their argument comes from comparing their study to an earlier one that was similar in many ways, but different in one important respect. In Posner and Mitchell's (1967) experiment, participants were also told to decide whether two stimuli presented were the "same" or not, based on various criteria. Stimuli could either physically match (A and A), match based on name (A and a), or match based on category (A and E vs. A and X). The comparison of interest here is to note that participants were faster at responding to stimuli that physically matched than those that matched by name.

At first glance, it appears that Theios and Amrhein (1989) replicated this result, and extended it to pictures and words as well as letters. However, in the Theios and Amrhein experiment, even when the stimuli were not exact physical matches, the difference between them was merely one of scale. The small stimuli were merely shrunken versions of the larger stimuli. On the other hand, in the Posner and Mitchell experiment, when stimuli were being matched on name, they did not match physically, even in scale. The property that the 'A' and the 'a' share is their name, not their physical appearance. Thus, while Posner and Mitchell could correctly claim that participants were being slowed down by virtue of needing to access name information to perform the task, it is not at all obvious that that is what is taking place in the Theios and Amrhein study. It could simply be that smaller stimuli are harder to process, and in fact the only significant effect of size of stimulus that they find is that when the second stimulus is small then participants respond the slowest. One can argue that this entire task is merely an exercise in visual comparison of different sized objects. Thus, by virtue of the fact that pictures and words take an almost identical amount of time to process in this fashion, an argument can be made that they take equivalent amounts of time to undergo early visual processing, even for tasks that eventually involve semantic memory.

Thus, if we are willing to accept for the moment that early visual processing takes roughly the same amount of time whether you are processing pictures or words, then the locus for the faster categorization of pictures has to be in one of two places. One possibility is that picture information is transmitted from working memory to semantic memory faster than verbal information. The other possibility is that, once semantic memory is accessed, pictures have faster access to categorization information than words do.

#### 1.3.2.2 Semantic relatedness effects

Guenther and Klatzky (1977) found that pictures cause a greater “negative relatedness” effect than words. Negative relatedness describes a situation where a participant is slower to reject an item as not belonging to the target category based on the degree to which the item is semantically related to the target category. They claimed that this meant that pictures and words could not be accessing exactly the same semantic information, as in the unitary content hypotheses, and theorized that negative relatedness occurs because the item to be rejected shares many features with the target category, as with a car being presented for the target category TRUCK. If pictures and words caused exactly the same information to be activated in semantic memory, as the unitary content hypotheses assume, then a picture and its corresponding word should both cause roughly the same negative relatedness effect.

They explain this difference by arguing that the locus for the effect is not in semantic memory, but rather in visual short-term memory by claiming that whereas words only activate semantic information, pictures also activate structural information (information about visual features), which is stored in visual short-term memory, and causes the greater negative relatedness effect. This, of course, relies upon the assumption that semantically related objects are also visually similar, which is certainly true for some categories (e.g. dogs, trees) but not others (e.g. furniture, tools). Put another way, as dissimilar as a collie and a beagle might appear, they are still arguably more visually similar than a chair and a bookcase or a hammer and a saw. Thus, it is possible that for some categories, such as living things, part of the semantic relatedness effect is due to structural information, but it is not likely that this is true for all categories.

Guenther, Klatzky, and Putnam (1980) indirectly demonstrate this in their Experiment 3 (see Table 1.1) where the semantic relatedness effect is disproportionately large for the category of living things in the picture-picture condition, but not for any other condition. In fact, if you ignore this disproportionate effect, the word-word condition is the one with the largest semantic-relatedness effect. If one presumes that the cross-modal (PW and WP) conditions are slower precisely because they are cross-modal (discussed in more detail below), then the implication is that if you take out the interference from structural information, as is potentially the case for the non-living things in the PP condition, then pictures are easier to distinguish semantically when making category judgments than words are.

Guenther et al. (1980) also found a pair of conflicting results. After replicating the negative relatedness result, they ran a pair of experiments in which participants were required to make judgements about whether objects presented as either words or pictures were either living things or non-living things. One trial was the presentation of two items sequentially. In the first of these experiments (Guenther et al. 1980, experiment 2) participants were presented with all four possible modality conditions: picture-picture, picture-word, word-picture, and word-word. They used four subordinate categories of living things (insects, mammals, flowers, and vegetables) and four subordinate categories of non-living things (clothes, furniture, tools, and vehicles). Using the case when the two items belonged to different superordinate categories (i.e. one living and one non-living thing) as their baseline, their data shows that there was essentially no priming effect at all for items belonging to the same superordinate category (i.e. both living or non-living things), but that there was a significant priming effect when both items belong to the same subordinate category. Furthermore, while this priming effect was greater in the two like-modal conditions, it still existed in the two cross-modal conditions, and was not significantly different for living vs. non-living things (data processed from Table 1 of Guenther et al., 1980).

**Table 1.1: Data from Guenther et al. (1980)**

| *Experiment 2: Priming effect* (ms) | Living things | Non-living things | Overall |
|---|---|---|---|
| PP | -109 | -31 | -78 |
| PW | -39 | -14 | -26 |
| WP | -40 | -35 | -37 |
| WW | -91 | -79 | -86 |
| | | | |
| *Experiment 3* | | | |
| PW | -93 | -74 | -84 |
| WP | -37 | -40 | -39 |

However, in the second of these experiments, Guenther et al. (1980, experiment 3) presented participants with only the cross-modal conditions. In this experiment, their table shows that there is no difference in the priming effect in the word-picture condition, but the picture-word condition shows a priming effect three times as large as what was seen in their first experiment (-84ms vs. –26ms). Once again, there was no difference for living vs. non-living things. Thus, while cross-modal priming does appear to exist, and appears to be equivalent for living and non-living things, it is unclear whether it is equivalent for cases whether or not the word or the picture is presented first.

In other words, it is unclear from these experiments whether pictures prime words better than words prime pictures. This matters because if pictures have faster access to category information than words, one would expect them to create a larger priming effect than words would in the cross-modal conditions (PW > WP). If, however, there is no difference in the priming caused by either pictures or words on the other modality of stimulus, then the implication becomes that perhaps the superiority that pictures have for accessing category information is a general superiority for accessing semantic information that may have nothing to do with the nature of semantic memory and simply be a result of differences in perceptual processing.

It is unclear precisely what these conflicting results have to say on this particular issue. One possible argument is that, unlike in their first experiment, in the second experiment, participants knew what modality to expect the second stimulus in because it had to be the opposite of the modality of the first stimulus. Thus, participants could prepare for the second stimulus in the interval (400ms) between when they responded to the first stimulus and when the second one was presented by converting the first stimulus into the appropriate modality. If this is the case, then we would expect to see a similar increase in priming for the word-picture condition that we saw for the picture-word condition. However, one could argue that the interval is long enough for participants to generate the name of the picture they just saw, but not long enough to generate a mental image of the word they just saw. This is potentially supported by the fact that the priming effect of the picture-word condition in the second experiment (-84ms) is almost equivalent to the priming effect from the word-word condition of the first experiment (-86ms), and, in fact, when broken up by category, the priming effects for both living and non-living things in the picture-word condition of the second experiment are equivalent to that of the word-word condition of the first experiment.

Given the radical difference in the priming effect in the picture-word condition between the two experiments, participants in the second experiment are clearly performing the task differently than participants in the first experiment, and, again, this may be a result of participants naming or generating the category name for the picture they are presented with so as to be prepared for the second stimulus. On the other hand, the priming effect of the word-picture condition is equivalent in both experiments. This could bolster the contention that it is harder to convert word stimuli into picture representations for later use.

Another possible explanation is that whenever participants perform tasks like this one, they always convert the stimuli to a verbal representation. In the picture-word condition of Experiment 3, they have time to generate this verbal representation before the word is presented. In the word-picture condition, they still have to generate the verbal representation of the picture while they are being timed, just like in Experiment 2, and so that condition should not be expected to differ between the two experiments. However, there is no compelling reason to believe that this is, in fact, what participants do. If they always converted pictorial stimuli to verbal representations, then we would expect participants to react slower to pictures than to words, regardless of the semantic task involved (i.e. naming, categorization, etc.). Clearly this is not the case since across multiple experiments participants always categorize pictures faster than words.

#### 1.3.2.3 Memory effects

Several studies have looked at an effect known as the picture superiority effect. In general, it has been shown that pictures show both better recognition (Shepard, 1967) and recall (Bousfield, Esterson, & Whitmarsh, 1957) than words. The effect is not small. Shepard (1967) found that participants recognized roughly 87% of several hundred pictures. Pictures also do not show such phenomena as serial position effects and list primacy or recency effects. Stenberg, Radeborg, and Hedman (1995) extended this work to include cross-modality matching as well. In a series of experiments, they presented participants with sets of pictures and words to study. The participants were then shown another series that contained both the item they studied and its opposite modality counterpart. For example, if they originally studied a picture of a dog in the first series, they would be shown the picture of the dog and the word "dog" in the second. They found that items studied as pictures were more rapidly recognized than items studied as words, regardless of recognition modality, and that participants were more likely to remember the same modality version of the item they studied.

These results would seem to indicate that pictures are activating different information from words on at least some level. It cannot be a purely perceptual effect because the recognition advantage of pictures also carried over to recognizing the word that corresponds to the picture, but not vice versa. One possible interpretation of this is to assume that people automatically name pictures that they see, in much the same manner as it is argued that word reading is automatic. Thus, one could make a levels-of-processing argument and claim that pictures are generally subject to a deeper level of processing because in order to access their names, we not only have to perform the perceptual processing, but we have to access the pictures' semantic representations to retrieve the name information. On the other hand, to remember words we only need to process the name information, which could simply be the result of the grapheme-to-phoneme mapping mentioned above.

#### 1.3.2.4 Summary of cognitive evidence

To summarize, the evidence from the cognitive literature breaks down into the following points that any model must account for:

- People have faster recall access to object names from words than they do from pictures (for example, Potter & Faulconer, 1975; Guenther et al., 1980; Smith & Magee, 1980), but this may simply be due to a faster grapheme-to-phoneme access that words have that bypasses semantic memory.
- People have faster access to categorization information from pictures than they do from words (see above).
- Pictures cause a greater negative relatedness effect in categorization judgments than words do (Guenther & Klatzky, 1977).
- Pictures are both better recognized (Shepard, 1967) and recalled (Bousfield, Esterson, & Whitmarsh, 1957) than words.
- Cross-modal priming is equivalent regardless of whether the superordinate category that the objects belong to is living or non-living things. It is unclear whether it is equivalent regardless of the modality of the first stimulus. (Guenther et al., 1980).
- Pictures prime words better than words prime pictures in a recognition task (Stenberg at al., 1995).

### *1.4 Summary*

As can be seen, there are problems with some of the cognitive studies that present firm conclusions from being drawn regarding the organization of information in semantic memory. In particular, the robust finding that words are named faster than pictures while pictures are categorized faster than words is suspect due to the uncertainty about just how much semantic information is required for word reading. Furthermore, even the categorization tasks do not necessarily provide accurate reflections of how semantic memory is organized.

As a result, a new series of experiments was devised that attempts to shed more light on this issue. These experiments differed from the studies reviewed above in one key fashion. Participants were not told what categories to classify the stimuli into, but rather were required to respond based on whatever categories they felt were appropriate. Thus, since they did not know what the stimuli would be ahead of time, they could not construct anticipatory retrieval structures, and would thus be required to use semantic memory to perform the task.

# CHAPTER 2: EXPERIMENT 1 – NATURAL CATEGORIZATION

## 2.1 Background and Predictions

The primary purpose of Experiment 1 was to give a better overview of the time-course for making judgments about category membership for both words and pictures by addressing a number of issues that arose with regards to the categorization experiments cited in Chapter 1. Foremost among these is the fact that those experiments did not require participants to retrieve the category of a given object on their own. Rather, these experiments came in two flavors. One merely required the participants to either respond “yes” or “no” that a given object belonged to a given target category, which the participants knew ahead of time. The other simply asked participants to make a forced choice between two pre-specified, broad categories, such as living or non-living things.

It is an open question whether or not these tasks require the participant to fully access semantic representations of the objects presented to them, particularly those objects that do not match the target category. One could make the argument that, for example, if a person is asked to determine whether an object belongs to the target category "tools" or not that they create a retrieval structure (Ericsson & Kitsch, 1995) containing information about tools, and then simply compare the stimulus object to the objects contained in this working memory structure. This would mean that participants could complete these tasks without using the underlying structure of semantic memory, but rather by using whatever retrieval structure they generated.

Put another way, when participants are told that the target category is tools, they can create a retrieval structure containing the information above tools needed to perform the task, such as names or examples. When given a task requiring them to simply decide whether a presented object belongs to the target category, rather than having to look up the necessary information in semantic memory again, they can simply reference this retrieval structure. If the object is present in the retrieval structure, they will answer "yes."

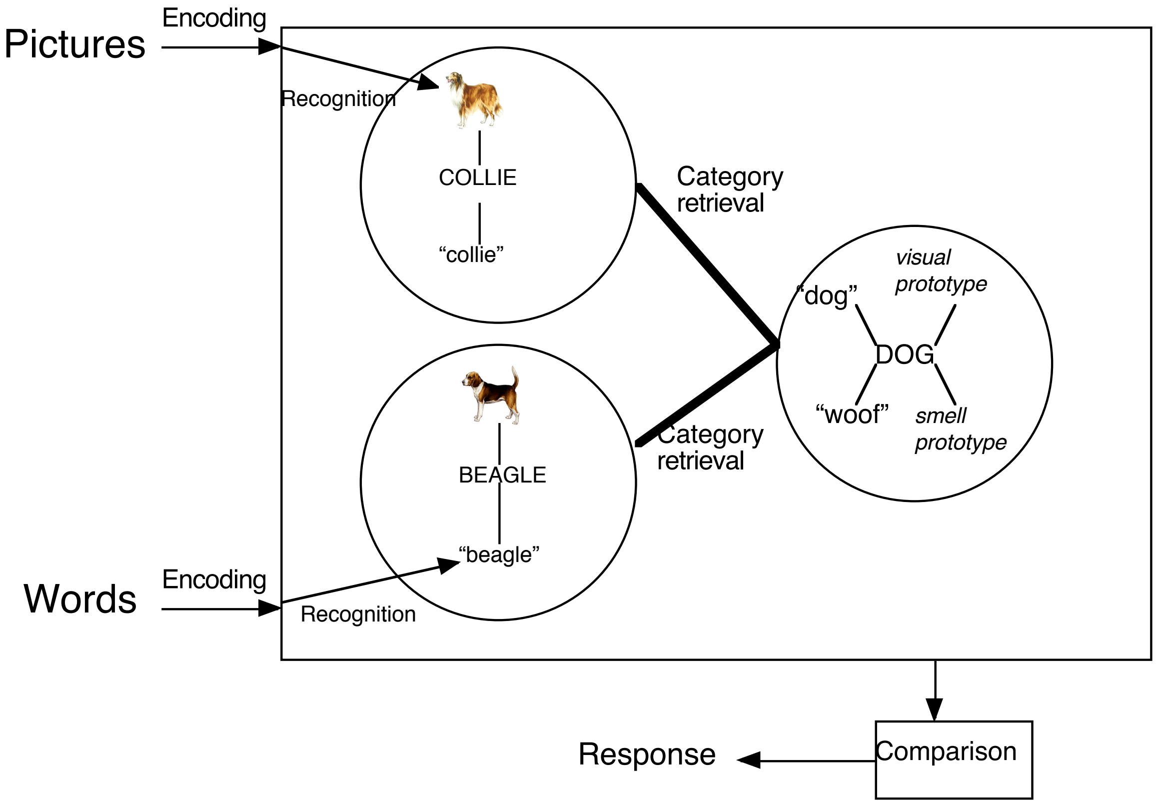
The primary point is that there is no guarantee that participants are accessing the underlying semantic representations used to create the retrieval structure on every trial, and thus there is no guarantee that the results of these experiments reflect the structure of semantic memory as opposed to the structure of the retrieval structures participants are able to generate. As a result, Experiment 1 was designed to force participants to make their own judgments about category membership, rather than having one imposed upon them.

In Experiment 1, participants were forced to look up category information in semantic memory and were denied the ability to create short-cut strategies or mental structures that might bypass semantic memory. Participants were simultaneously presented with two stimuli, either two pictures, two words, or a picture and a word and asked to determine whether or not both stimuli "depict or describe objects belonging to the same category." Broadly, this task requires four stages:

- Perceptual encoding and recognition - converting the input stimulus into a usable representation, taking the encoded representation as input, and finding the appropriate semantic representation for the input
- Category retrieval – Retrieving the superordinate category of the input
- Comparison – Comparing the two retrieved categories to each other
- Response preparation and generation – responding "yes, they belong to the same category" or "no, they don't"

How, exactly, these processes operates depends on whether you subscribe to the UCH or the MSH (this will be explained in more detail below). There is no difference between the two models with regards to steps 1, 3, and 4, thus the crux of the difference between the two models rests in the recognition and category retrieval steps – i.e., the ones that require semantic access.

**Figure 2.1: UCH model of Experiment 1**



To envision how the two models behave differently with regards to step 2, let us first analyze how these steps would occur under the UCH. In Figure 2.1 we use the example of having a picture of a collie and the word "beagle" as input. Under the UCH, the information coming into the system from either the picture or the word input will eventually activate the COLLIE or BEAGLE node in semantic memory, respectively, with the picture obviously coming through the visual input stream and the word coming through the verbal input stream.

At this point, according to the UCH, there can no longer be any differences in how the two stimuli are processed. They would both traverse a semantic link to the same category node, DOG, to retrieve the category of the input. After the categories of both stimuli have been retrieved, they can be compared and a response generated, in this case "yes."

It is a relatively robust finding that pictures are categorized faster than words (see section 1.2.2.1 above). According to the UCH, this can be the result of a difference in processing in one of two stages – the perceptual encoding and recognition stage, or the category retrieval. Ideally, we'd like to be able to presume that whatever differences might exist in this stage are relatively consistent. If encoding and recognizing the word "beagle" is x milliseconds slower than encoding and recognizing a picture of a collie, we would like to be able to presume that, on average, it always takes approximately x milliseconds longer for a word to be encoded and recognized than a picture.

Unfortunately, it is also well known that variables such as familiarity and word frequency can have an effect on how long it takes for an object to be recognized, and typicality can have an effect on how long it takes for an object to be categorized (though this is, in general, going to be correlated with familiarity and word frequency). Thus, in Experiment 1, we take three steps to attempt to control these effects – two in the design of the experiment and one in the data analysis. When choosing the words to use in the experiment, as much as possible they were selected to be high frequency words based on

Battig & Montague (1969) and Uyeda & Mandler (1981) (See Appendix A.1).[3] Secondly, items were randomly paired together with the only caveats being that roughly half the trials should be matching trials, a picture could not be paired with its naming word, no item was repeated within a trial block, and if an item was used in a trial block, its opposite modality partner was not used in that block.[4] Thus, I would claim that any differences that might result from visual familiarity or word frequency would tend to average out over the course of the 200 trials an average participant completed. As a further check, as will be seen below, when the results were analyzed, the effect of visual familiarity was regressed out of the data.

The upshot of all of this is that we feel comfortable making the assumption that the difference in the amount of time it might take to encode and recognize a word versus a picture is within a relatively constrained range, and thus, in the average case, should be reliably consistent. Therefore, under the UCH, the pattern of data that we would expect to see from this particular experiment is one where the reaction times are distributed linearly across the three conditions. To clarify, let us parameterize the four stages involved in the task. The encoding and recognition stage will be represented as ER, with words being $ER_w$ and pictures being $ER_p$. The category retrieval stage is CR, the comparison stage is C, and the response stage is R. Thus, the equations for performing the task in all three conditions would be:

Picture-picture: $(2*ER_p) + (2*CR) + C + R$
Picture-word: $(ER_p+ER_w) + (2*CR) + C + R$

---

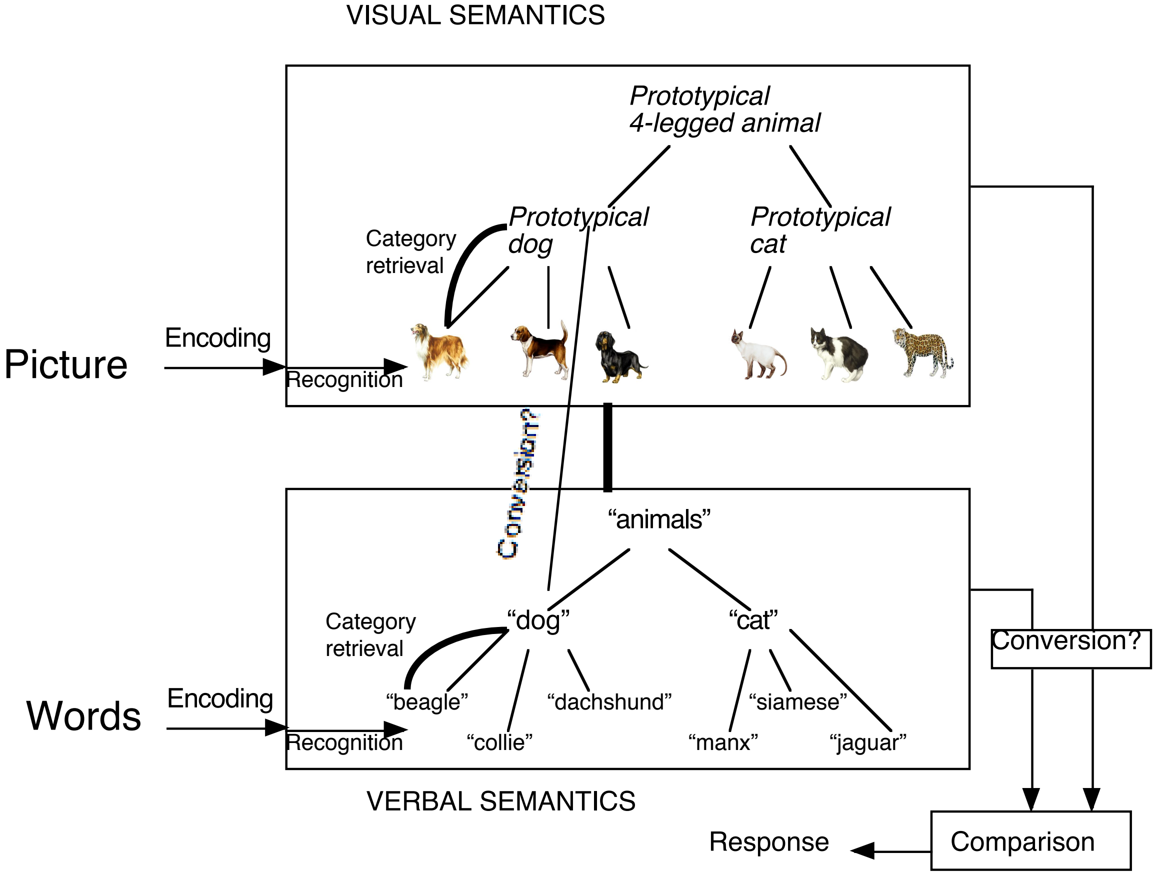
[3] In a few cases, it was impossible to select stimuli from these lists, either because the category used was not contained in the paper or because not enough items from the category were listed in the paper. Participants had the option of skipping any trial in which they did not recognize an item, and those trials were not included in the analysis. See Appendix A.1 for more details.

[4] In other words, if the word "beagle" was used in a block, that word would not be repeated for the rest of that block, nor would the picture of the beagle appear anywhere in that block.

Word-word: $(2{*}ER_w) + (2{*}CR) + C + R$

Therefore, the expected distribution of reaction times across the three conditions should average out to a linear distribution with a slope of $ER_w - ER_p$.

**Figure 2.2: MSH model of Experiment 1**



Now let us model the task based upon the MSH (see Fig. 2.2). The same four basic stages still compose the task, but now there are many potential sources of difference between pictures and words. Since the MSH makes no different claims about perceptual encoding and recognition than the UCH, it should be clear that the same potential differences exist in those stages in the MSH model as in the UCH model, and thus the same controls are still valid. It should also be clear that the comparison and response stages are the same under the MSH as the UCH.

The major difference between the two models with regards to this task is in the category retrieval stage. Under the MSH, it is unclear that, regardless of whether the input is a word or a picture, the category information retrieved is the same. Since picture inputs will directly access the visual semantic system while word inputs will directly access the verbal semantic system, it is not clear that the categories retrieved from the respective systems will necessarily be compatible for comparison. If one type of category is retrieved from the visual semantic system and another type from the verbal semantic system, and those two types cannot be directly compared with each other, then some sort of translation step will be necessary. This step could take place before category retrieval if participants attach a verbal label to pictures or generate mental images of the word stimuli. It could also occur before the comparison stage if, say, the visual category retrieved is converted into its verbal equivalent or vice versa.

This, of course, begs the question of just what qualifies as a "category" in a semantic subsystem under the MSH. While there are a number of different ways of approaching this issue, it is used here primarily as a convenient shorthand to characterize the relationships of a set of highly similar representations. If one imagines the information stored in visual semantics as being akin to a set of mental pictures, then the most obvious way to index these pictures is based on their visual features. If this is organized hierarchically, then highly similar representations will share the same superordinate node in the hierarchy. It is this superordinate node that I am referring to as a category node in visual semantics. For example, one can imagine that in a system indexed by visual features, the visual representations of a beagle and a collie will share the same superordinate node due to their high degree of visual similarity.

As a result, there are three sets of possible equations describing the total reaction time of participants doing the current task under the MSH – one with no conversion, one where the conversion takes place before category retrieval and involves always either attaching a verbal label to the pictures or imagining the words, and one where the conversion takes place as needed before the comparison stage[5]:

*No Conversion*:
Picture-picture: $(2^*ER_p) + (2^*CR) + C + R$
Picture-word: $(ER_p+ER_w) + (2^*CR) + C + R$
Word-word: $(2^*ER_w) + (2^*CR) + C + R$

*Before category retrieval*: (assume verbal labels attached to pictures)
Picture-picture: $(2^*ER_p) + (2^*Con) + (2^*CR) + C + R$
Picture-word: $(ER_p+ER_w) + Con + (2^*CR) + C + R$
Word-word: $(2^*ER_w) + (2^*CR) + C + R$

*Before comparison*:
Picture-picture: $(2^*ER_p) + (2^*CR) + C + R$
Picture-word: $(ER_p+ER_w) + Con + (2^*CR) + C + R$
Word-word: $(2^*ER_w) + (2^*CR) + C + R$

As a result, there are two possible predictions for participants' reaction times under the MSH. If there is no conversion, or if the conversion takes place before category retrieval, then we would also expect to see a linear distribution of reaction times across the three conditions with a slope of either $(ER_w - ER_p)$ or $(ER_w - ER_p) \pm Con$. However, if the conversion takes place only as needed before the comparison stage, then we would expect a non-linear distribution of reaction times across the three conditions, with the picture-word condition being slower than would be expected if the reaction times were distributed linearly.

---

[5] There is, of course, the possibility that any conversion that might take place before category retrieval could also only take place as needed. However, for this task, this is functionally equivalent to it taking place as needed before the comparison stage, so I will not be spending time discussing it.

Inherent in the generation of these equations is the assumption that the processing involved with this task can be thought of as stage-like, with each stage relatively independent of the others. This is an assumption that is not universally accepted, particularly by those who subscribe to the connectionist viewpoint of cognition. However, it is my contention that the stage-processing assumption is not completely necessary to justify the predictions made above.

Within a connectionist characterization of the task, when a picture or word is input into semantic memory, it causes a particular pattern of activation to be generated. Under the UCH, this pattern of activation would represent all of the semantic information related to the picture, regardless of modality. Thus, when the second picture or word is processed, if it belongs to the same category, it should cause a highly similar pattern of activation to be generated, regardless of modality, leading to a linear distribution of reaction times across the three modality conditions.

Under the MSH, words and pictures would cause different patterns of activation to be generated within different semantic subsystems. In the case where both stimuli are the same input modality, then each one should cause a highly similar pattern of activation to be generated within the proper semantic subsystem. Thus, two pictures belonging to the same category would cause highly similar patterns of activation to be generated in visual semantic memory. The same would also be true when both stimuli are words. Thus, the second stimulus is going to benefit from a priming effect due to residual activation that may be left over from the processing of the first stimulus. However, in the cross-modal condition, the two stimuli would cause different patterns of activation to be generated in different semantic subsystems, thus eliminating the priming effect. What this means is that participants will be slower in the picture-word condition than they would be if that condition benefited from the same priming effect, meaning that the picture-word condition would be slower than it would otherwise be if it was part of a linear distribution of reaction times.

Another assumption implicit in the predictions generated above is that words and pictures access semantic memory from different input pathways. However, given that the words are being presented visually, one can ask how it is that they access semantic memory differently from other visual stimuli, such as pictures. For the purposes of generating the above predictions, I operated from the assumption that words are automatically converted into an auditory image (i.e., sounded out in your head) before they are processed further. This is a theory known as the strong phonological theory of word reading which, while not universally accepted, has some support (see, for example, Frost, 1998).

I would argue, however, that making this assumption does not have a significant impact on what the predictions for Experiment 1 would look like. What the assumption of strong phonology does is it says that the semantic information activated by reading a word is the same semantic information activated by hearing that word. Effectively, this makes all language use access a verbal semantic sub-system that can be views as being part of a broader auditory semantic system. Nevertheless, if we presume, as we did above, that the semantic subsystems are organized such that similar representations are stored topographically near each other, then it can be argued that the semantic representations of auditory verbal information are much more likely to be stored close to each other than to the auditory non-verbal representations stored in auditory semantics.

From there, it is not a tremendous leap to argue that, since we arguably have far more experience with auditory verbal information than with other types of auditory information, this segregation would be reinforced to the point that verbal representations would effectively exist in their own subsystem, independent of other auditory semantic representations.

Based on this argument, if we take it as a given that, instead of a phonological image being created, words can directly access semantic memory from their orthography (i.e., through the visual input stream), then essentially what we are saying is that the semantic information activated when you read a word is not necessarily the same as the semantic information activated when you hear a word. This would be because the verbal semantic subsystem that is linked to reading would be part of the broader visual semantic system. However, it could be argued that just as auditory verbal information becomes segregated from the rest of auditory semantics, so would visual verbal information become segregated from the rest of visual semantics. Therefore, the conception of semantic memory used to generated the MSH predictions for Experiment 1 would be effectively correct regardless of whether the strong phonological theory of reading is, in fact, the correct one.

## 2.2 Methods

### *2.2.1 Participants*

The participants for Experiment 1 were 30 Carnegie Mellon University undergraduates who were compensated with course credit for their participation. Participants had the option of pressing the space bar if they did not recognize a picture or a word. Two participants who did this on over 30% of the trials involving words were excluded from the analyses.

### *2.2.2 Stimuli*

The picture stimuli were 138 colored pictures of living things that fell into nine different categories – dogs, cats, horses, fish, trees, birds, flowers, fruits, and vegetables. The word stimuli used in this experiment were the names of the objects used as the picture stimuli. [A full list of the stimuli used and, where possible, their typicality ratings can be found in Appendix A.1.] Stimuli were presented on a Macintosh computer using the PsyScope experiment software (Cohen, MacWhinney, Flatt, & Provost, 1993).

Participants were seated a normal distance from the computer so that the screen was at eye level, roughly two feet away. The pictures were roughly 1.5 inches square, and the words were presented so that they were roughly as wide as the pictures on average. There was on average a one inch separation between the two stimuli.

### *2.2.3 Procedure*

On any given trial, participants were presented with two stimuli simultaneously. The stimuli were either two pictures, two words, or one picture and one word, and there was an equal probability in each trial which type of stimulus pair they would see. For the picture-word trials, there was an equal probability that the stimuli would appear with the picture on the left and the word on the right or vice versa.

Participants were instructed to decide as quickly as possible whether the two stimuli “depict or describe” objects that belong to the same category. They were not told what the possible categories were, nor were they told the criteria that they were supposed to use to categorize the objects. They were given several practice trials to familiarize themselves with the task and with the types of objects they were going to be presented with. They were then given three blocks of 69 trials. Within each block, no picture or word was repeated, and once a picture was used, the word corresponding to its name could not be used in that block and once a word was used the picture it named could not be used in that block. The stimuli were present on the screen until the participant responded, at which point there was a 1s delay until the presentation of the next pair of stimuli.

### *2.2.4 Summary*

To summarize, the basic methodology was as follows. Participants walked in and were instructed that they were going to be presented with pairs of stimuli and their task was to judge as quickly as possible whether the stimuli depicted or described objects that belonged to the same category. The pairs of stimuli were either two pictures, two words, or one picture and one word. They were then presented with a series of practice trials to familiarize themselves with the task, which was followed by three blocks of 69 trials each in which no stimulus or its opposite-modality counterpart was repeated within a block.

## 2.3 Results

As mentioned above, two participants did not respond to over 30% of the trials. Two other participants had error rates over 50%. Their data were omitted from these analyses. Further, trials in which the participants answered incorrectly (i.e. saying that two objects did not belong to same category when they did and vice versa) were excluded. These accounted for 9.3% of trials. Finally, trials of over 3000ms were considered to be outliers and were excluded. These accounted for another 2.3% of all trials. All told, 11.6% of trials were excluded from the following analyses.

**Figure 2.3: Mean reaction time by condition by response**

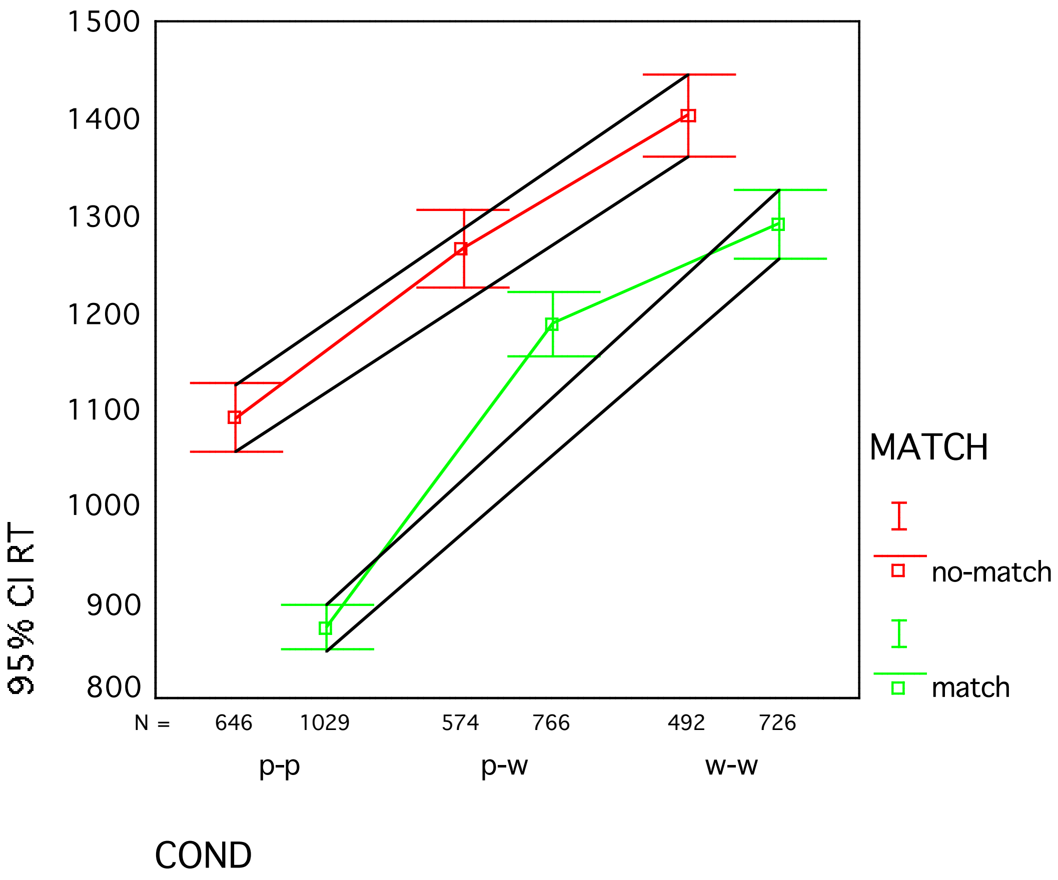


**Table 2.1: Reaction times (ms) for Experiment 1**

| | Match | | No-match | |
|---|---|---|---|---|
| | Mean | s.d. | mean | s.d. |
| Picture-picture | 874.09 | 392.78 | 1090.12 | 450.22 |
| Picture-word | 1187.26 | 483.89 | 1265.32 | 481.51 |
| Word-word | 1290.69 | 482.36 | 1403.02 | 489.13 |

There were significant main effects of modality condition (F(2, 54)=160.92, p < .001) and participant response (F(1, 27)=34.72, p < .001). There was also a significant interaction effect of these two variables (F(2, 54)=5.60, $p < .05$). Table 2.1 and Figure 2.3 show the mean reaction times.

To confirm that these effects were due to the variables of interest and not to confounding variables, individual linear regression tests were done to test for effects of possible confounds. The possible confounds tested were visual similarity of the two stimuli, visual complexity and familiarity of each stimulus, and the number of syllables in the names of each stimulus. The results of these analyses are collected in Table 2.2. The only significant effects were: visual similarity, $t = 4.79$, $p < .001$; syllables in the name of the right-side stimulus, $t = 3.22$, $p < .001$; and visual complexity of the right-side stimulus, $B = -13.23$, $p < .05$.

**Table 2.2: Effects of possible confounds**

| Variable | Coefficient ($B$) | $t$ | $p$ |
|---|---|---|---|
| Visual similarity | 49.95 | 4.79 | < .001* |
| Left-side syllables | 11.38 | 1.27 | .205 |
| Right-side syllables | 29.55 | 3.22 | <.001* |
| Left-side visual complexity | -5.83 | -0.92 | .355 |
| Left-side visual familiarity | -10.98 | -1.63 | .104 |
| Right-side visual complexity | -13.33 | -2.04 | <.05 |
| Right-side visual familiarity | 1.11 | .16 | .870 |

To provide final confirmation that the effects were due to the variables of experimental interest, all of the possible confounds were put into one linear regression and the unstandardized residuals were computed. These residuals were then used as the dependent variable of a two-way ANOVA with stimulus condition (picture-picture, picture-word, or word-word) and stimulus match (same category or different category) as the independent variables. In this ANOVA, there were main effects of both stimulus condition, F(2, 4227)=227.66, $p < .001$, and stimulus match, F(1, 4227)=58.25, $p < .001$. There was also a significant interaction of condition with match, F(2, 4227)=8.43, $p < .001$. As Figure 2.4 and Table 2.2 show, these residuals have virtually the exact same distribution as the reaction time data analyzed above.

**Figure 2.4: Mean residuals by condition by response**

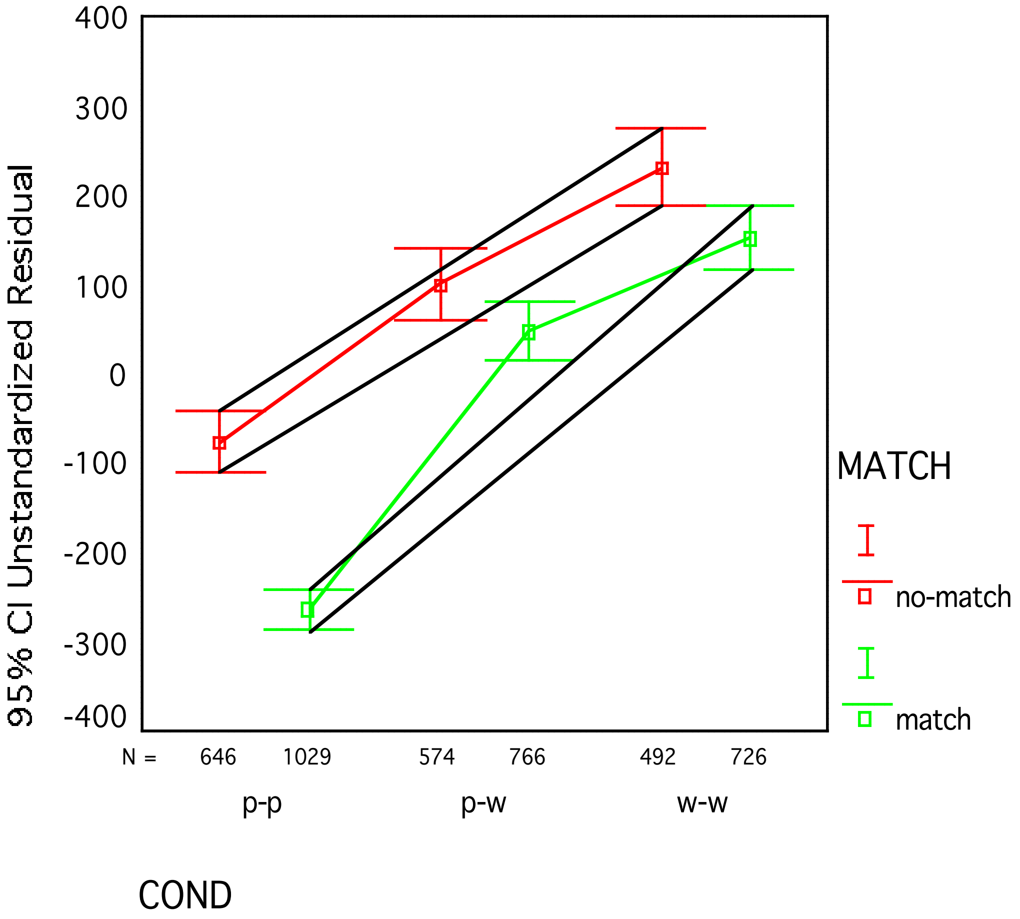


**Table 2.3: Mean residuals for Experiment 1**

| | Match | No-match |
|---|---|---|
| Picture-picture | -263.0 | -76.00 |
| Picture-word | 47.92 | 100.9 |
| Word-word | 154.6 | 231.0 |

Finally, we are interested in knowing whether or not the reaction times are distributed linearly across the three conditions. To test this, I performed a contrast-coded regression on the reaction times. What this test does is create a new variable in which every like-modality trial (i.e. picture-picture and word-word trials) is assigned a value of .5 and every cross-modality trial is assigned a value of 1. This variable is then used as the independent variable in a regression with reaction time as the dependent variable, thereby allowing a statistical comparison of where the picture-word reaction times actually falls with where it would fall if the distribution of reaction times across the three conditions was linear. For no responses, this test was not significant, $\underline{F}(1) = 1.05$, $\underline{p} > .05$. However, for yes responses, this test was significant, $\underline{F}(1) = 30.78$, $\underline{p} < .001$.

As can be seen from Table 2.1, the difference between the picture-picture condition and picture-word condition when the stimulus categories do not match is 176.9 and between the picture-word and word-word conditions is 130.1. On the other hand, when the categories do match, the difference between the picture-picture and picture-word conditions is 310.92, but between the picture-word and word-word conditions is only 106.68.

## 2.4 Discussion

Before we can draw any significant conclusions from Experiment 1, there is one result that needs to be explained – the difference in the reaction time patterns between the match and no-match conditions. The fact that the no-match condition is slower than the match condition across the board is not unexpected. Posner and Mitchell (1967) did an experiment where participants were presented with two letters and had to respond yes if both letters had the same name (i.e., were the same letter regardless of case) and no otherwise. They found that the fastest condition was when the same physical letter was presented both times (e.g., A – A). A little slower was when the letters were different cases but had the same name (A – a). Finally, the slowest condition was with different letters (e.g., A – B). The fact that the no-match condition is slower essentially extends this result from letters to objects.

The more interesting difference between the match and no-match conditions is why the distribution of the no-match condition is linear, whereas the distribution of the match condition deviates significantly from linearity. Thus far, we only have two explanations for a lack of linearity – a conversion step before category retrieval or before the comparison step. Both of these are processes that should take place regardless of whether the eventual answer is yes or no. Thus, we have a seeming contradiction in the two pieces of data.

If we examine the equations for the MSH prediction and apply them to the yes condition reaction times, we find (assuming conversion before comparison):

Picture-picture: $(2*ER_p) + (2*CR) + C + R = 874\text{ms}$
Picture-word: $(ER_p+ER_w) + Con + (2*CR) + C + R = 1187\text{ms}$
Word-word: $(2*ER_w) + (2*CR) + C + R = 1290\text{ms}$

Based upon the picture-picture and word-word conditions, we can get an estimate of 1082ms for the picture-word condition if the data were distributed linearly. This yields an estimate of 105ms for the conversion parameter. Thus, since the same parameter should presumably be present in the no condition as well, we would expect the following equations:

Picture-picture: $(2*ER_p) + (2*CR) + C + R = 1090\text{ms}$
Picture-word: $(ER_p+ER_w) + Con + (2*CR) + C + R = 1351\text{ms}$
Word-word: $(2*ER_w) + (2*CR) + C + R = 1403\text{ms}$

Instead, the picture-word condition only takes 1265ms, which is roughly what the estimate would be if the reaction times were linear (1246.5ms). There is, however, another possibility that we failed to consider when creating our predictions in the introduction.

The most likely explanation hinges upon a key difference in the semantic representations being accessed in the match and no-match conditions under the MSH framework. Look again at Figure 2.2. When a participant is presented with two pictures from the same category – say, a beagle and a collie – they will process one picture, which will then activate the DOG category in visual semantic memory. When the other picture is processed, in order to perform the task correctly, it will activate the exact same DOG node in visual semantics, which would produce the standard semantic priming effect. Likewise when two words from the same category are presented. However, when the trial involves a word and a picture, they are activating *different* category nodes – one in visual semantics and one in verbal semantics – and so the picture-word condition does not have a priming effect. Thus, some or all of the non-linearity in the match condition could be the result of the fact that the like-modality trials benefit from a semantic priming effect, whereas the cross-modality trials do not. For the non-match trials, none of them benefit from a semantic priming effect, and thus the distribution is more linear.

To put it another way, the implicit assumption in the introduction was that the match and no match conditions should behave roughly the same, and thus be subject to essentially the same equations. What Experiment 1 has demonstrated is that this is not necessarily the case, and a more accurate way of characterizing the task would be to generate two sets of prediction equations. In these equations, the P parameter is the priming effect:

MSH – no match condition
p-p: $(2*ER_p) + (2*CR) + C + R_n$
p-w: $(ER_p+ER_w) + (2*CR) + C + R_n$
w-w: $(2*ER_w) + (2*CR) + C + R_n$

MSH – match condition
p-p: $(2*ER_p - P) + (2*CR) + C + R_y$
p-w: $(ER_p+ER_w) + (2*CR) + C + R_y$
w-w: $(2*ER_w - P) + (2*CR) + C + R_y$

Thus, what Experiment 1 has found is not a conversion parameter but a priming effect of 105ms that is only present in the conditions where participants are being presented with the same category of stimulus in the same modality.

The UCH has a more difficult time explaining this discrepancy between the distributions of the match and no-match trials. Since the underlying feature of the UCH framework is that inputs of any modality activate the same category information in semantic memory, then presumably any semantic priming effect observed in the same-modality trials in the match condition should also be present in the cross-modal trials since the picture and the word are accessing the same category node in semantic memory. Clearly, this is not what is happening, thus this discrepancy is troublesome for the UCH.

What happens to the conversion step that we theorized would lead to the non-linear distribution in the first place? The answer to this is unclear. Looking at Table 2.3, we can see that the non-match condition does show a slight, but statistically insignificant, deviation from linearity (pw–pp = 175.2ms; ww–pw = 137.7ms) that may hint at the presence of a conversion step, but that may also just be a statistical artifact. Thus, it is unclear whether a conversion step is necessary or whether semantic priming can explain the entire non-linearity witnessed. Regardless, though, both notions can only be supported by the MSH framework.

The evidence from Experiment 1 seems to provide clear support for the MSH framework. There is a steadily increasing reaction time across the three conditions, with a significant non-linearity where the picture-word condition is slower than it should be if there was a linear distribution. Both likely explanations of this non-linearity – either the need for a conversion step between visual and verbal semantic representations or a semantic priming effect in the like-modal match trials – support the MSH model of the task used.

However, there is one other possibility that could explain these data yet still provide support for the UCH. It is possible that the nonlinearity is caused by a task-switching effect. This effect would be the result having to mentally "switch gears," so to speak, from processing pictures to processing words (or vice versa). Thus, the effect would only be present in the picture-word case, causing the non-linearity of the reaction time distribution found above. Experiment 2 was designed to address whether or not such a task-switching effect could be found.[6]

---

[6] If there is a task-switching effect, presumably it should be present in both the match and no-match conditions. While there is no statistically significant deviation from linearity in the no-match condition, there is a slight deviation that may or may not be due to a task-switching effect. Thus, experiment 2 was still performed to provide extra certainty that the effects observed were not due to task-switching.

## CHAPTER 3: Experiment 2 - Object/lexical decision task

### 3.1 Background and Predictions

Experiment 2 was designed to address one possible confound that was discovered in the design of Experiment 1. Specifically, the non-linear pattern of reaction times observed could be explained by the existence of an as-needed conversion from one category type to another (e.g., visual to verbal or vice versa) in the picture-word condition, as we have suggested. Alternatively, it could be explained by the existence of a task switching effect that imposes a processing cost only in the picture-word condition and is independent of how information is being processed in semantic memory.

One can think of this task switching effect as being akin to a priming effect. The non-linear distribution of data in Experiment 1 would result from the fact that when a participant has a same-modality trial, either picture-picture or word-word, they obviously will process each stimulus one at a time. Once one modality is used as input, the participant will be "primed" to process information in that modality, either through activation of that modality's input pathways or attentional priming or some other mechanical process. Thus, when it comes time to process the second stimulus, that processing will go somewhat quicker because the input pathway has been primed. This would make the same-modality trials somewhat faster than they would be without the priming effect. However, such an effect would not exist for the picture-word trials because of the two different input pathways being used to process both stimuli.

Put another way, it could be the case that, rather than the equations presented in Section 2.4, are more appropriate set of equations would be:

Picture-picture: $(2*ER_p) + (2*CR) + C + R$
Picture-word: $TS+ (ER_p+ER_w) + (2*CR) + C + R$
Word-word: $(2*ER_w) + (2*CR) + C + R$

On the other hand, we would also expect to see this task switching effect in the no condition, where it does not seem to appear. Thus, it is unclear that this will end up accounting for the effect, but Experiment 2 can at least rule it out as a possible explanation.

Experiment 2 uses a lexical/object decision task where participants are asked to judge whether or not the stimuli presented are real objects or real words. As with Experiment 1, two stimuli are presented simultaneously – two pictures, two letter strings, or one of each – and participants are asked to judge whether some minimal number of these stimuli (either one or two, depending on the condition) are real or not. This task is structured exactly the same as the task in Experiment 1, but the underlying processing involved eliminates the category retrieval stage and replaces the comparison stage with a simple tally of how many real items there are. Thus, we have a task with the following three stages:

- Perceptual encoding and recognition
- Tally
- Response

In Chapter 2, we theorized that any non-linearity we might observe would need to be the result of either a conversion between category types or the result of a task switching effect. Since Experiment 2 eliminates the category retrieval and comparison stages, and the tally stage does not require any conversion, then any deviation from linearity across the three conditions in this experiment would be due to a task switching effect.

## 3.2 Methods

### *3.2.1 Participants*

The participants for Experiment 2 were 25 Carnegie Mellon University undergraduates who were compensated with course credit for their participation. 5 participants with error rates over 20% were excluded from the analyses.

### *3.2.2 Stimuli*

The stimuli were the same pictures and words used in Experiment 1 (see Chapter 2), and modified versions that were divided into three groups:

real stimuli – pictures of real objects; real words

plausibly unreal stimuli – pictures of objects with analogous parts interchanged (e.g., a beagle with a horse's head); pronounceable letter strings (e.g. "mave")

unreal stimuli – pictures of objects with non-analogous parts interchanged (e.g., the bottom of a beagle with the top of a tree); random letter strings (e.g., "metj").

Participants were seated a normal distance from the computer so that the screen was at eye level, roughly two feet away. The pictures were roughly 1.5 inches square, and the words were presented so that they were roughly as wide as the pictures on average. There was on average a one inch separation between the two stimuli.

### *3.2.3 Procedure*

On any given trial, participants were presented with two stimuli simultaneously. The stimuli were either two pictures, two letter strings, or one picture and one letter string, and there was an equal probability in each trial which of the four possible pairings of stimuli they would see (picture-picture, picture-word, word-picture, word-word).

Half of the participants were instructed to decide as quickly either whether both items they saw were real (2-yes) while the other half were instructed to decide whether at least one item was real (1-yes). This was to control for the fact that in the 2-yes condition, participants could also just respond "no" as soon as they saw an unreal item, meaning that rather than always having to look at both stimuli and make a decision they could just be making their decision as soon as they saw a stimulus that determined the answer, i.e. an unreal stimulus. Thus, the 1-yes condition was devised to balance this by having a set of trials where the deciding stimulus was the real one.

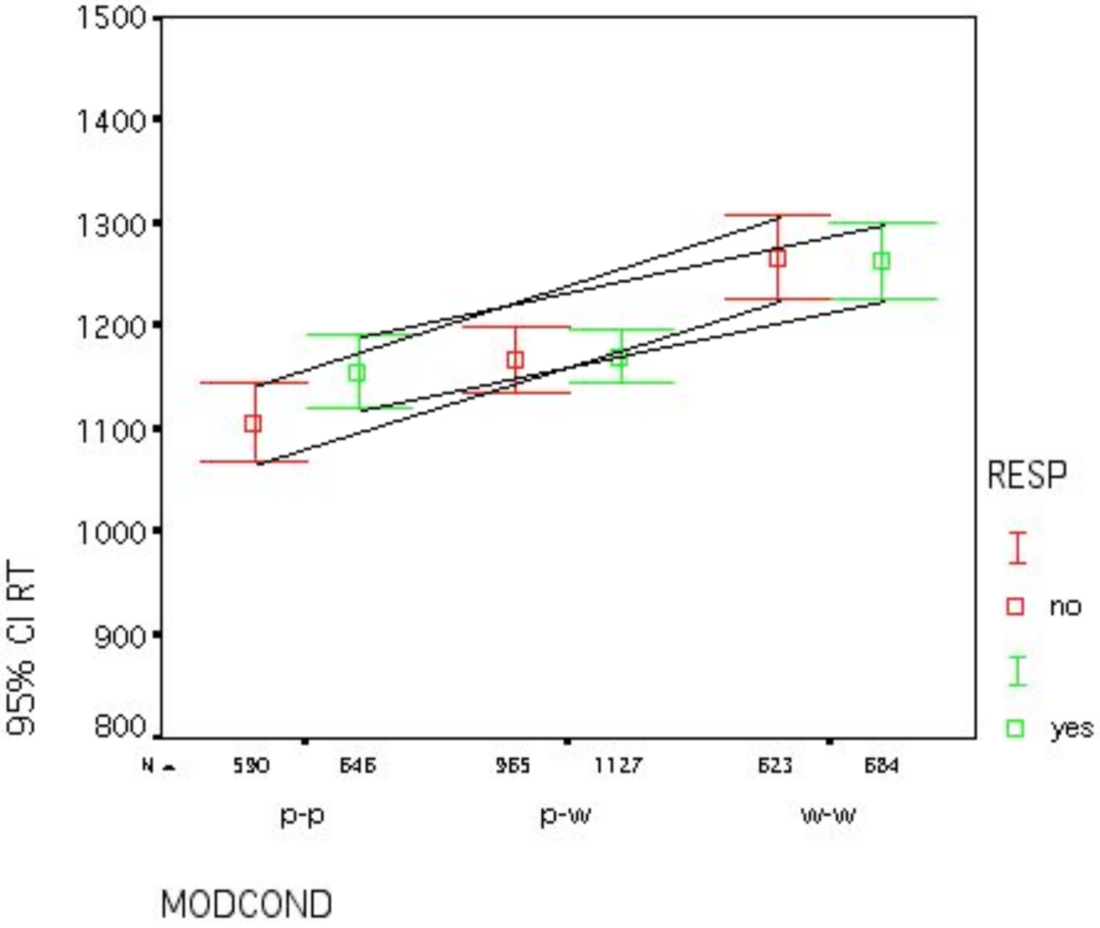
Participants were then given several practice trials to familiarize themselves with the task and with the types of objects they were going to be presented with. They were then given three blocks of 69 trials. Within each block, no picture or letter string was repeated, and once a picture was used, the word corresponding to its name could not be used in that block and once a word was used the picture it named could not be used in that block. The stimuli were present on the screen until the participant responded, at which point there was a 1s delay until the presentation of the next pair of stimuli.

### *3.2.4 Summary*

To summarize, the basic methodology was as follows. Participants walked in and were instructed that they were going to be presented with pairs of stimuli and their task was to judge as quickly as possible whether the stimuli were either both real, or whether at least one of the two stimuli was real. The pairs of stimuli were either two pictures, two letter strings, or one picture and one letter string. They were then presented with a series of practice trials to familiarize themselves with the task, which was followed by three blocks of 69 trials each in which no stimulus or its opposite-modality counterpart was repeated within a block.

## 3.3 Results

**Figure 3.1: Mean reaction time by condition**



As mentioned above, five participants with error rates over 20% were excluded from the analyses. Trials that participants answered incorrectly were also excluded from these analyses. These accounted for 8.8% of all trials. Trials longer than 3000ms were also excluded as outliers. These accounted for another 2.8% of trials. All told, 11.6% of trials were excluded from these analyses.

A two-way ANOVA of the reaction times with modality condition and response as the two factors was performed. There was a significant effect of modality condition, F(2, 4629) = 27.06, p < .001. However, unlike in Experiment 1, there was not a significant effect of response, F(1, 4629) = 1.31, p > .10. Figure 3.1 and Table3.1 show the mean reaction times by condition and response.

**Table 3.1: Mean reaction times (ms) for Experiment 2**

| | yes | | no | |
|---|---|---|---|---|
| | Mean | s.d | Mean | s.d |
| p-p | 1156.78 | 466.40 | 1106.47 | 463.63 |
| p-w | 1171.30 | 448.48 | 1167.95 | 503.89 |
| w-w | 1263.84 | 485.74 | 1267.65 | 521.90 |

As further confirmation that the effects found were based on the variables of experimental interest, the data for Experiment 2 were analyzed in a similar style to those of Experiment 1. Individual regressions were performed on the possible confounds to determine whether their effects were significant or not. Then, a single regression containing all the possible confounds was performed and the residuals calculated so that an ANOVA could be performed on the residuals for the variables of experimental interest. Finally, the data were tested for linearity. Both potential confounds were found to be significant: Yes condition – t = 64.325, p < .001; reality condition – t = 64.582, p < .001.

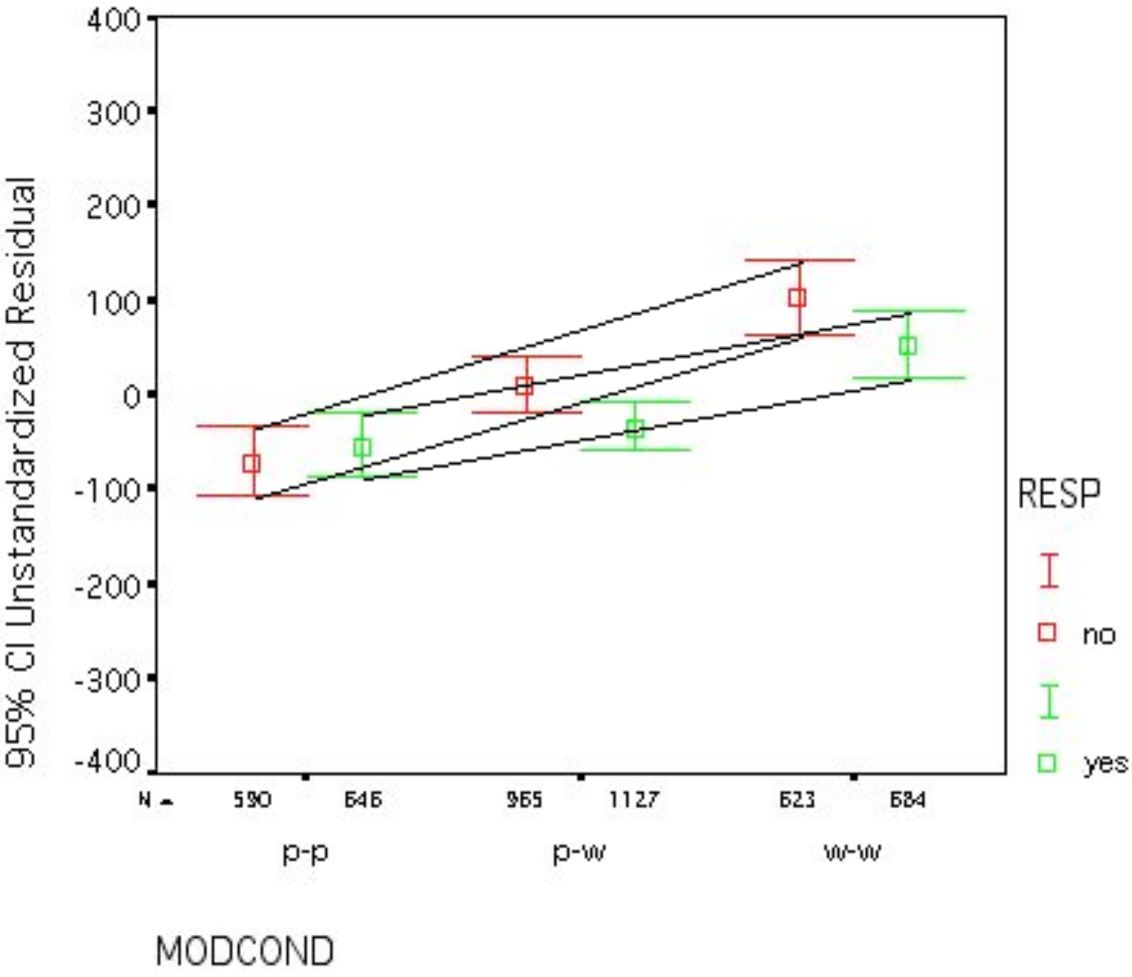
An ANOVA on the residuals was performed with modality condition (picture-picture, picture-word, or word-word) and response (yes or no) as the independent variables. A main effect of modality condition was found, F(2, 4629) = 30.18, p < .001.

However, unlike in Experiment 1, no main effect of response was found, nor was there an interaction effect between modality and response. Figure 3.2 and Table 3.2 contain the mean residuals.

**Table 3.2: Mean residiuals for Experiment 2**

| | yes | no |
|---|---|---|
| p-p | -52.45 | -69.82 |
| p-w | -32.96 | 11.55 |
| w-w | 53.51 | 103.48 |

**Figure 3.2: Mean residuals by condition by response**



Finally, the same linearity analysis that was used in Experiment 1 was performed on the reaction times for Experiment 2. There was no significant deviation from linearity in either the yes condition ($\underline{F}(1) = .057$, $\underline{p} > .10$) or the no condition ($\underline{F}(1) = 3.06$, $\underline{p} > .05$).

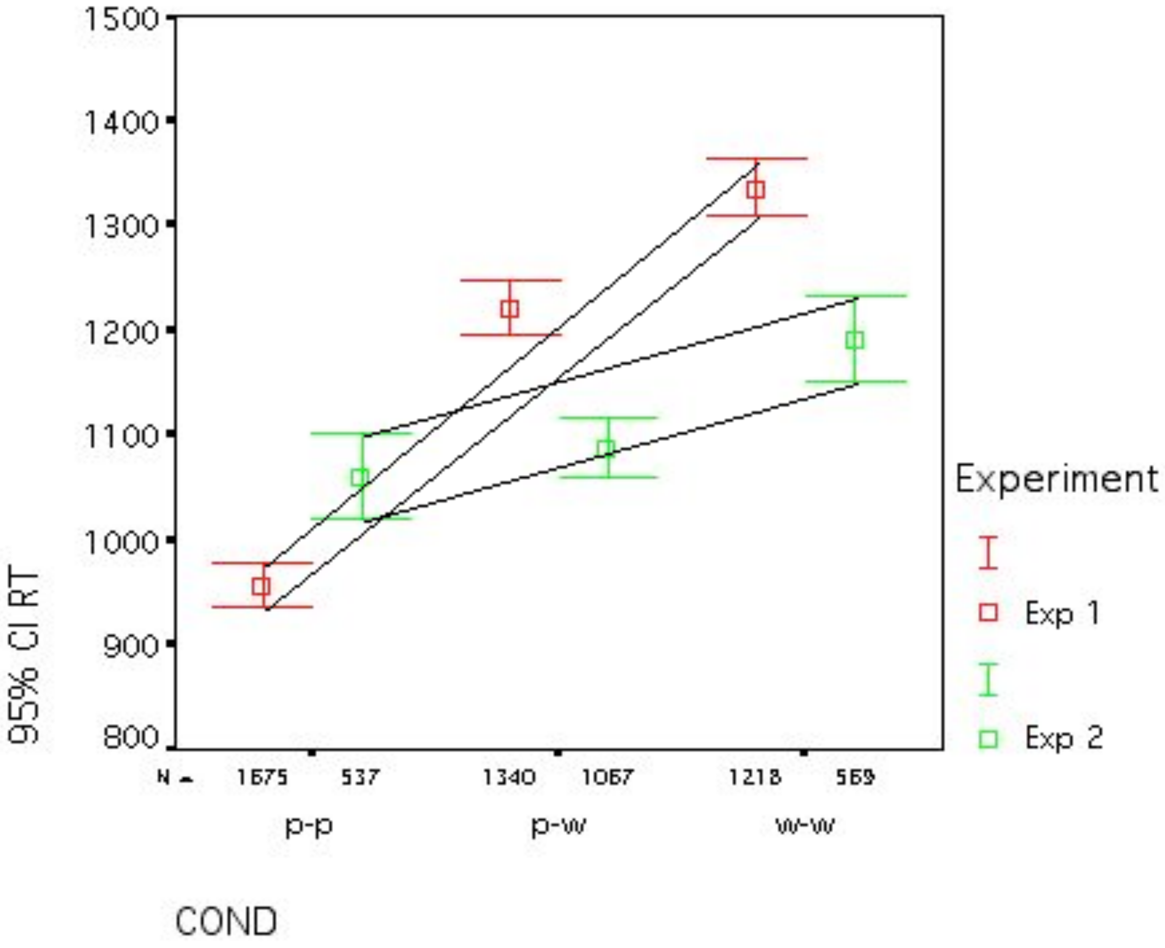
To further check whether this apparent lack of deviation from linearity means anything, an analysis was performed that compared the "yes" (same category) responses from Experiment 1 to the "yes" responses from the two-yes condition of Experiment 2 (i.e. both items are real). These two conditions were compared with modality condition and experiment as the independent variables. If there is a significant interaction between modality condition and experiment, this implies is that there is a significant difference in the distribution of reaction times across the modality conditions between the two experiments, which would confirm the lack of linearity. This analysis was possible because the real stimuli used in Experiment 2 were a subset of the stimuli used in Experiment 1.

There was, as expected, a significant effect of modality condition, $\underline{F}(2, 3361) = 69.72$, $\underline{p} < .001$. There was not a significant effect of experiment, $\underline{F}(1, 3361) = 1.92$, $\underline{p} > .10$. However, there was a significant interaction between the modality condition and the experiment, $\underline{F}(2, 3361) = 25.80$, $\underline{p} < .001$. Figure 3.3 and Table 3.3 show the means compared for this analysis.

**Table 3.3: Means for yes conditions – Exp. 1 vs. Exp. 2**

| | Exp. 1 | Exp. 2 |
|---|---|---|
| p-p | 957.40 | 1060.83 |
| p-w | 1220.70 | 1088.25 |
| w-w | 1336.06 | 1192.81 |

**Figure 3.3: Experiment 1 vs. Experiment 2 – yes conditions**



## 3.4 Discussion

The results of Experiment 2 demonstrate that the non-linearity observed in the results of Experiment 1 is unlikely to be the result of a task-switching effect. The fact that the results in Experiment 2 are effectively linear (and the slight non-linearity that does exist is in the wrong direction) shows that the non-linearity observed in Experiment 1 is not the result of an effect in either the perceptual encoding and recognition stage or in the response stage. Thus, the non-linear distribution observed in Experiment 1 is not the result of task switching effect or of priming of the input pathways in the same-modality trials.

At this point, one potential argument that could be made is that the results observed in Experiment 1 are the result of a semantic priming effect, and not necessarily a result *per se* of how semantic memory is structured, and thus the UCH could still explain the results. However, this explanation is less than compelling for a number of reasons. First among them is the fact that, according to the UCH, information from different modalities about the same category should always activate the same category node in semantic memory. It should not matter whether you are reading the word “beagle” or seeing a picture of one – the same set of information should be activated, and thus readily available. If this were the case, one would expect the semantic priming effect, an effect presumably present in the category retrieval stage, to be present in the cross-modality trials just as much as it is in the same-modality trials as the picture and the word would still be activating the same information. Thus, if there is a semantic priming effect that is only present in the same-modality trials and not in the cross-modality trials, this necessarily entails that the different modalities are activating different sets of information. This, of course, is precisely what the MSH is arguing for and the UCH is arguing against.

# CHAPTER 4: Experiment 3 - Functional categorization

## 4.1 Background and Predictions

Experiments 1 and 2 provide evidence for the MSH model of how information is structured in semantic memory. In Experiment 1, we found a pattern of reaction times that was consistent with either the existence of a task-switching effect or the MSH model. In Experiment 2, we found evidence that there was no task switching effect in the results of Experiment 1. In an effort to provide yet more support for the MSH, a third experiment was devised.

However, we wanted to devise an experiment that would provide a strong test of one of the key components of the MSH model. Specifically, in order to explain the existence of patients that suffer from category-specific semantic deficits, proponents of the MSH have devised what they call the “sensory-functional hypothesis.” Patients who suffer from category specific deficits are those who are relatively impaired at semantically processing living things *vis a vis* man-made objects or vice versa due to brain damage.

This is a syndrome that is clearly difficult to explain under the MSH framework. The solution that was devised was to theorize about just what it is that separates living things from man-made objects. The primary notion is that there are significant differences in precisely what types of features make up the bulk of the representations of the different classes of objects. Natural objects are primarily interacted with in a visual fashion – we observe trees and flowers and wild animals and what-not. Thus, the bulk of the features used to represent these objects in semantic memory are visual, which in the

MSH framework means that most of the features in their categorical representations are stored in visual semantic memory.

Man-made objects, on the other hand, are interacted with in a primarily functional fashion. We use man-made objects, we build them to serve a particular function, and we classify them based on the functions that they serve. Thus, the representations for man-made objects contain proportionately more functional features as opposed to physical features, which would be stored in a semantic subsystem other than visual semantics (it is left unspecified whether there exists a separate functional semantic system or whether this information would be stored as part of the verbal semantic system or some other system). Category specific deficits, in this hypothesis, are simply the result of damage either to the visual semantic system or whatever system contains functional information.[7]

Experiment 3 was designed to attempt to test whether or not functional semantic information is stored differentially from at least visual semantic information, and possibly verbal semantic information. The task was designed to be an exact analogue of the task used in Experiment 1 in that, once again, participants are presented with two stimuli simultaneously – two pictures, two words, or one of each. In this experiment, however, participants are asked to judge whether the stimuli depict or describe objects belonging to the same *functional* category – i.e., do the two objects serve the same broad function. An example would be "chair" and "bench" both belonging to the category SITTING.

The task used in this experiment has almost the exact same set of stages as the task in experiment 1, with one exception – the existence of a function retrieval stage. Thus, the five stages we hypothesize exist are:

---

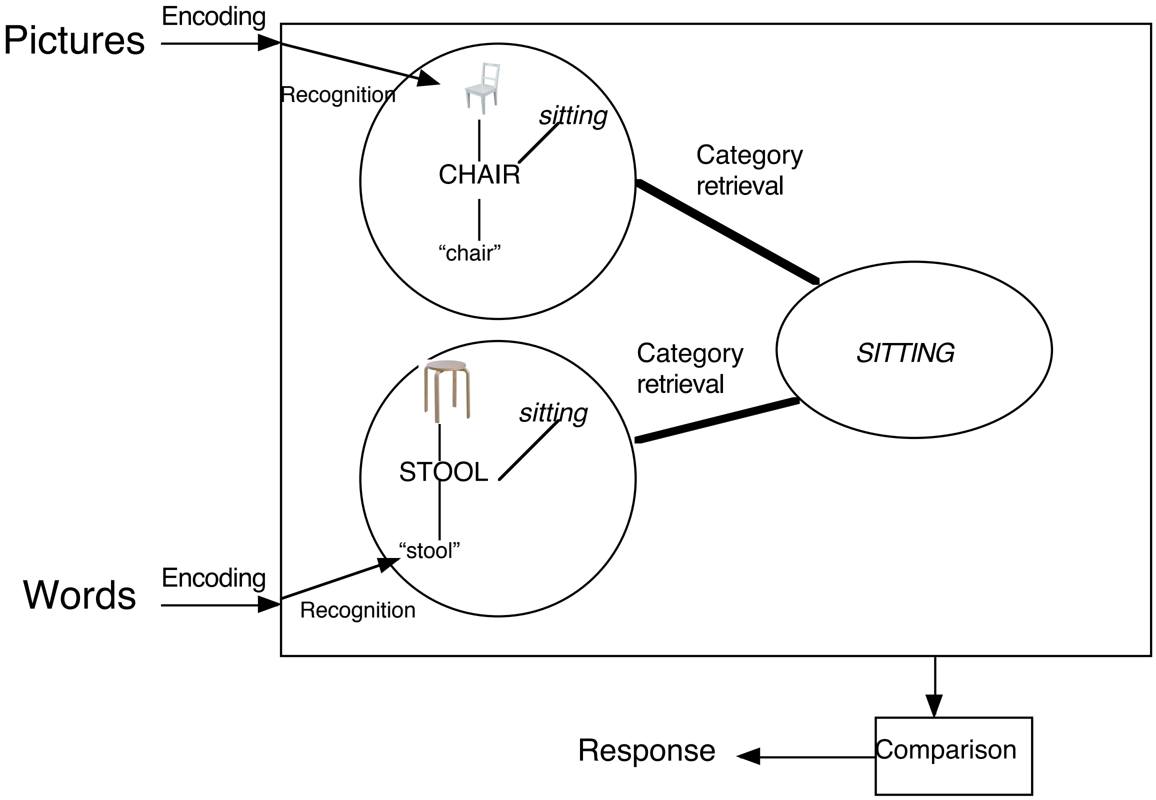
[7] The issues surrounding category specific deficits have been covered in much more detail in section 1.2.2.4 above, but for the purposes of designing Experiment 3, the veracity of the hypothesis was assumed.

- Perceptual encoding and recognition.
- Function retrieval
- Functional category retrieval
- Comparison
- Response

As with Experiment 1, one can anticipate that variables such as category typicality can be expected to have an impact on the category retrieval stage. These were controlled for in this experiment by doing another simple experiment to determine what stimuli would be used in Experiment 3. Participants were randomly selected to receive one of two surveys. One survey gave the participants a list of objects and asked them to list the first three functions that came to mind when they thought of that object (object->function). The second survey gave participants a list of functions and asked them to list the first seven objects that came to mind when they thought of that function (function->object). Items were also given a rating based on where in a participant's list they appeared (i.e., and item listed first by a participant was given a higher score than one listed third).

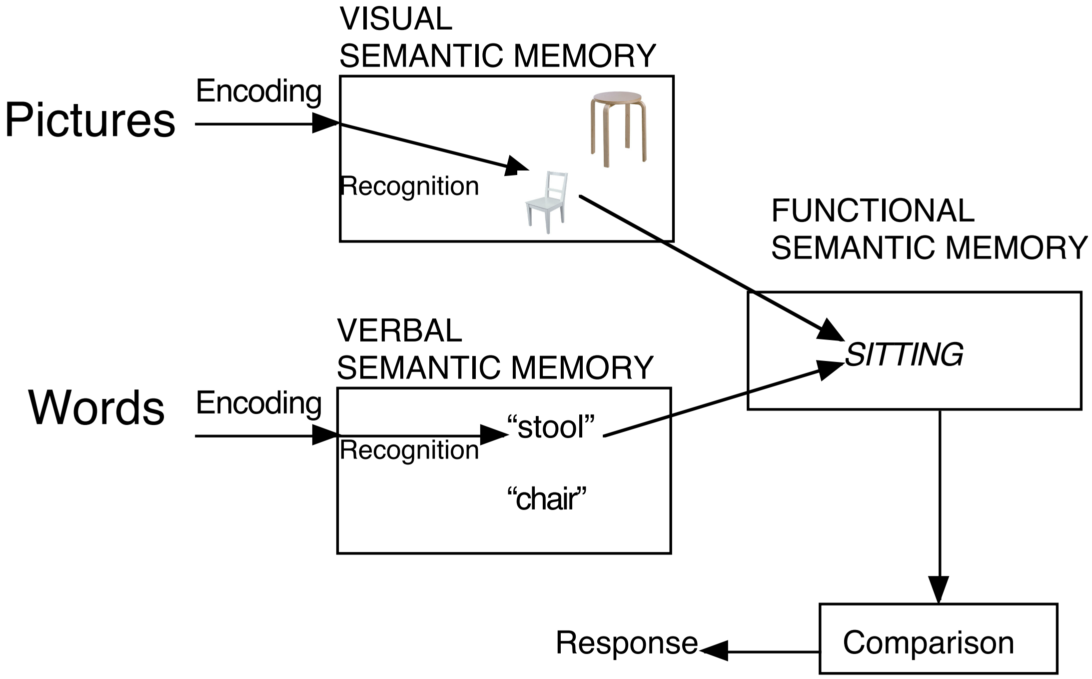
An item was only used as a stimulus if it was both listed under a function on the function->object survey and if that function was associated with it on the object->function survey.  Where possible, this list was double-checked against Battig & Montague (1969) and Uyeda & Mandler (1980). For example, the five items used in the category of "sheltering" based on the method above were all in the top 14 most typical items on Battig & Montague's (1969) list of "a type of human dwelling," and were the most visually distinct items in the top 14. The nine items used in the category "transporting" were all in the top 10 of Battig & Montague's list of "a type of vehicle." More details can be found in Appendix A.2.

**Figure 4.1: UCH model of Experiment 3**



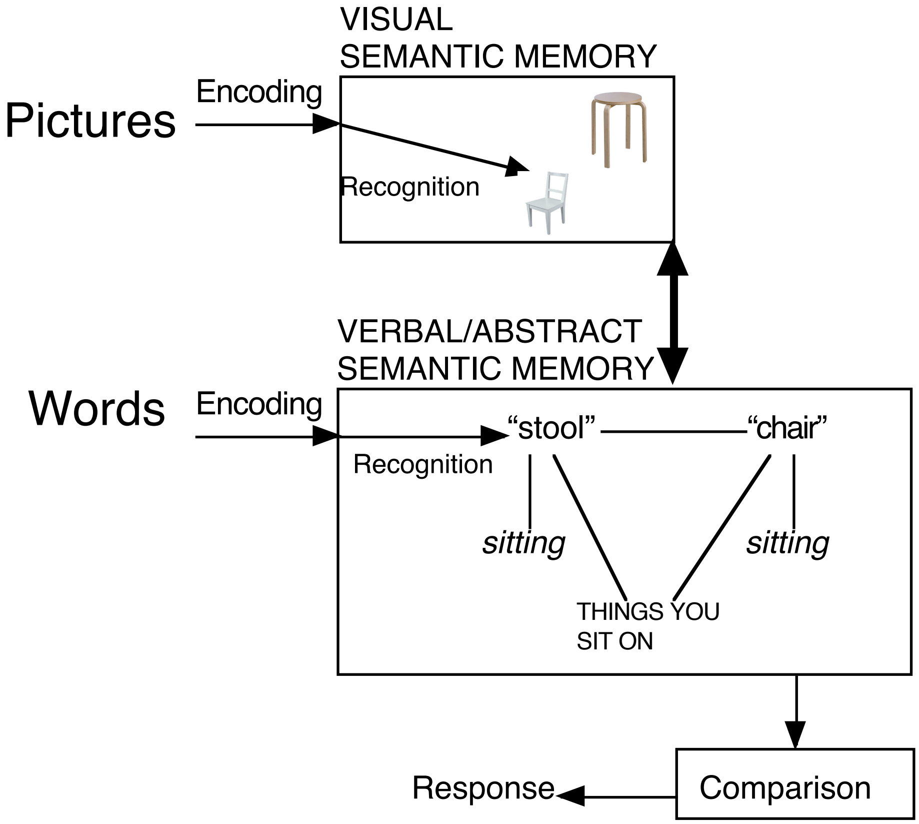
Therefore, as a result, the UCH would once again predict that there should be a linear distribution of reaction times across the three conditions. If we examine figure 4.1, we can see that the picture of the chair eventually will arrive at the CHAIR node, whereupon the function of a chair will be retrieved and then its broader functional category retrieved. The same process happens with the word "stool," except that it uses the verbal input stream instead of the visual input stream. Once both functional categories have been retrieved, they can be compared and a response generated. Since the functional categories for both items are of the same type, no conversion step is necessary. Thus, we would expect there to be a linear distribution of the data across the three modality conditions.

**Figure 4.2: MSH model of Experiment 3 – separate functional subsystem**



As would probably be anticipated, things can be expected to behave differently according to the MSH. There are two possibilities for how the MSH would explain behavior on this task. One possibility involves the existence of a separate functional semantic subsystem, similar to the already discussed verbal and visual semantic subsystems (see Figure 4.2). According to this characterization of the task, the picture of the chair would come in via the visual input stream, flow into the visual semantic representation of the chair, and then into the functional semantic representation of what function the chair serves. A similar process would take place with the verbal input of "stool." Since information from both the visual and verbal input streams would need to access the functional semantic representation to derive the proper response to a trial, there is no need for a conversion step between representations like the one we hypothesized would need to exist in the MSH model of the task from Experiment 1. Thus, what we would expect to see is a linear distribution of reaction times across the three conditions.

**Figure 4.3: MSH model of Experiment 3 – verbal/abstract semantic system**



Another possibility is that functional information is stored along with verbal information in a verbal/abstract semantic subsystem (see Figure 4.3). In this case, the picture of the chair would access the visual semantic representation of the chair, and then flow into the verbal/abstract semantic memory system where the functional information could be accessed. Since the functional information would always need to be accessed through the verbal/abstract semantic system, there is once again no need for a translation step, and so we would again expect the distribution of reaction times to be linear.

How, then, can we differentiate between the various models if they all predict a linear distribution of reaction times? One way is to look at what we should expect the reaction times to be in comparison to those in Experiment 1.[8] This is a tricky proposition, given that the tasks involve two different types of items and two different types of semantic information being used to classify them. However, since these differences are consistent across the entire task, I believe that some general statements can be made. If it is the case that man-made objects are simply processed differently from natural kinds, then we would expect the reaction times in Experiment 3 to consistently be either faster or slower than those in Experiment 1. The same holds true if it is simply more difficult, across the board, to access functional information. This is, essentially, what the UCH framework would predict – a consistent difference in reaction times across all three conditions.

Under the MSH, we have two possibilities. In the case where there is a separate functional semantic system, we would once again expect the reaction times to be slower than those in Experiment 1, to reflect the extra time needed to access functional semantics. In the case where functional information is stored in a verbal/abstract semantic system, we would expect the picture-picture condition to be slower, to reflect the time necessary to access verbal semantics, but the picture-word and word-word conditions should be about the same as they were in Experiment 1.

---

[8] Obviously, it is tricky, at best, to compare results from two different experiments using two different sets of stimuli. However, the overall distribution of reaction times was similar enough between the two experiments that I decided it was worth presenting the comparison and leading it up to the reader to judge how much stock they place in the conclusions.

## 4.2 Methods

### *4.2.1 Participants*

The participants for Experiment 3 were 16 Carnegie Mellon University undergraduates who were compensated with course credit for their participation.

### *4.2.2 Stimuli*

The picture stimuli were 63 colored pictures of man-made objects. The word stimuli used in this experiment were the names of the functions that objects used as the picture stimuli were most frequently said to fulfill in Experiment 3 (see Chapter 5). A full list of the stimuli used can be found in Appendix A.3. Stimuli were presented on a Macintosh computer using the PsyScope experiment software (Cohen, MacWhinney, Flatt, & Provost, 1993).

Participants were seated a normal distance from the computer so that the screen was at eye level, roughly two feet away. The pictures were roughly 1.5 inches square, and the words were presented so that they were roughly as wide as the pictures on average. There was on average a one inch separation between the two stimuli.

### *4.2.3 Procedure*

On any given trial, participants were presented with two stimuli simultaneously. The stimuli were either two pictures, two words, or one picture and one word, and there was an equal probability in each trial which type of stimulus pair they would see. For the picture-word trials, there was an equal probability that the stimuli would appear with the picture on the left and the word on the right or vice versa.

Participants were instructed to decide as quickly as possible whether the two stimuli "depict or describe" objects that belong to the same *functional* category. They were not told what the possible categories were, nor were they told the criteria that they were supposed to use to categorize the objects. They were given several practice trials to familiarize them with the task and with the types of objects they were going to be presented with. They were then given three blocks of 31 trials. Within each block, no picture or word was repeated, and once a picture was used, the word corresponding to its name could not be used in that block and once a word was used the picture it named could not be used in that block. The stimuli were present on the screen until the participant responded, at which point there was a 1s delay until the presentation of the next pair of stimuli.

### *4.2.4 Summary*

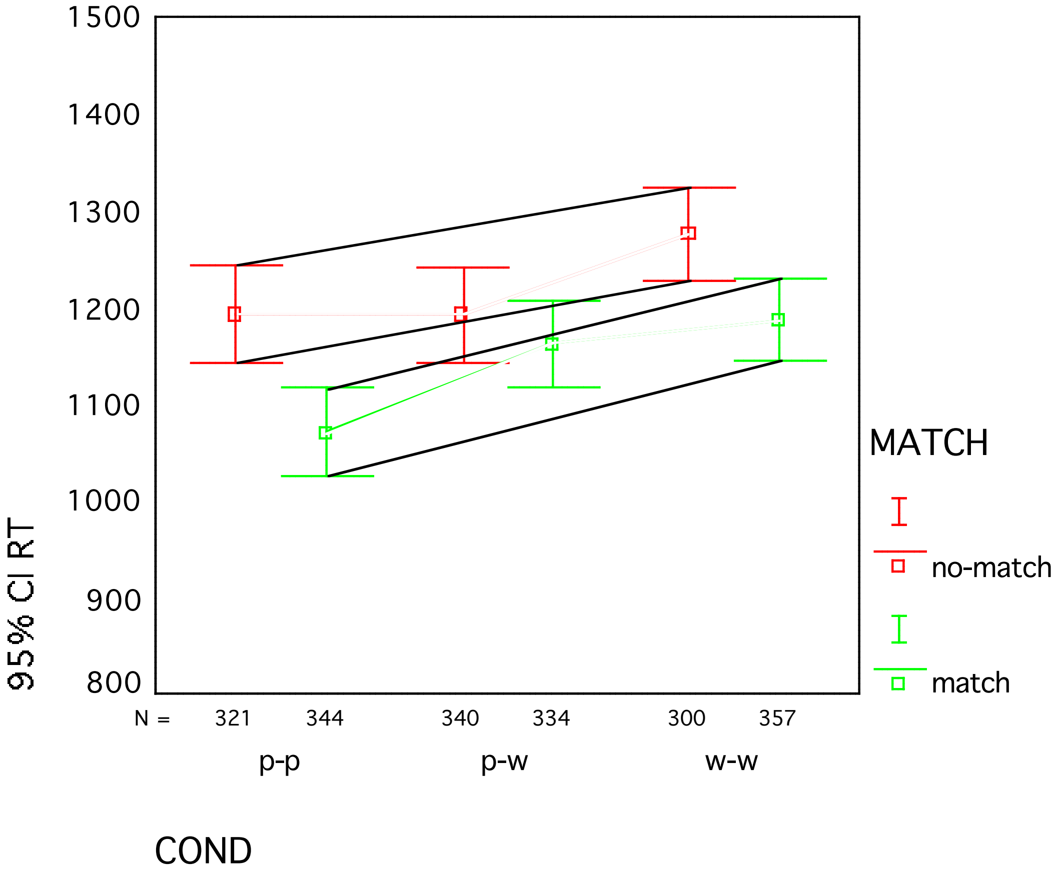
To summarize, the basic methodology was as follows. Participants walked in and were instructed that they were going to be presented with pairs of stimuli and their task was to judge as quickly as possible whether the stimuli depicted or described objects that belonged to the same category. The pairs of stimuli were either two pictures, two words, or one picture and one word. They were then presented with a series of practice trials to familiarize themselves with the task, which was followed by three blocks of 31 trials each in which no stimulus or its opposite-modality counterpart was repeated within a block.

## 4.3 Results

Trials on which participants made errors (i.e., saying two items were from different functional categories when they were from the same category) were excluded from these analyses. These accounted for 8.8% of all trials. Trials of longer than 3000ms were excluded as outliers. These accounted for another 1.6% of trials. All told, 10.4% of the trials were excluded from the following analyses.

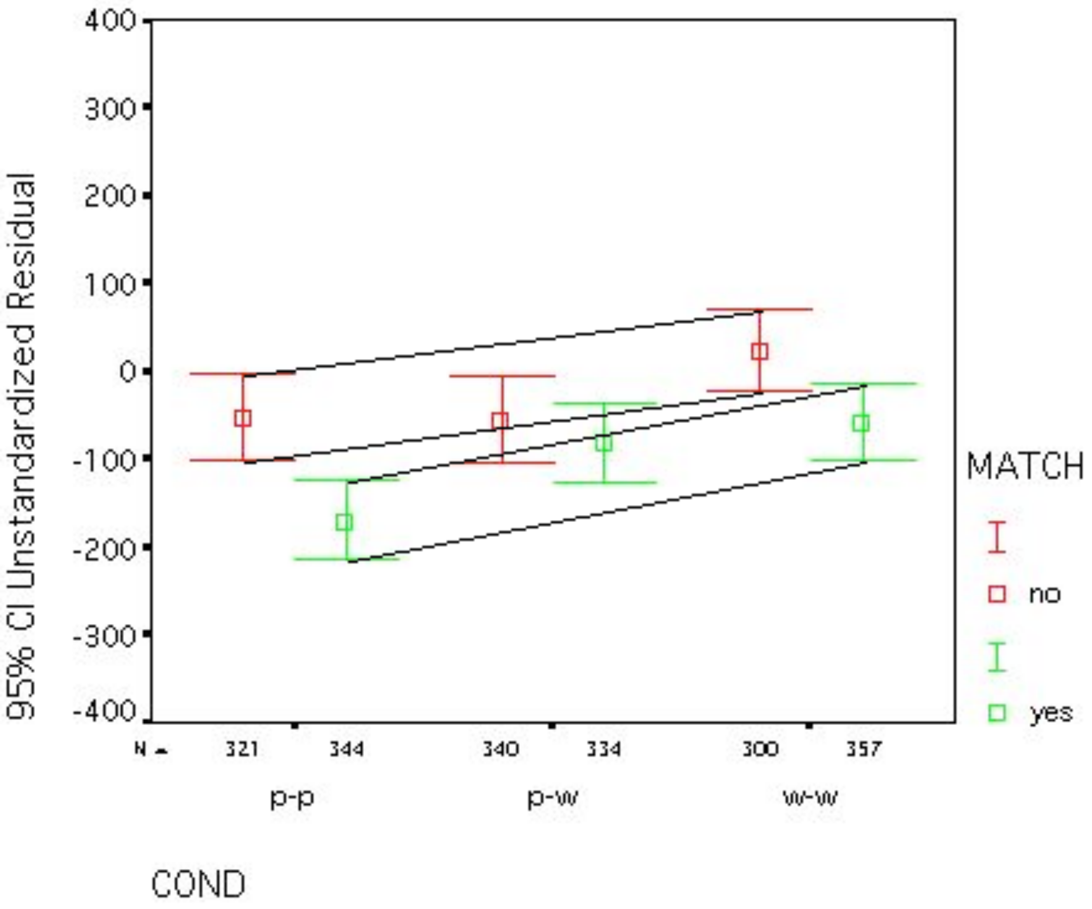
A two-way ANOVA was done on the reaction time data gathered with modality condition and response being the factors. There were significant main effects of both factors – modality condition: F(2, 22) = 9.86, $p < .001$; response: F(1, 11) = 6.99, $p < .05$. Figure 4.4 and Table 4.1 show the mean reaction times for each condition.

**Figure 4.4: Mean reaction times by condition for Experiment 3**



Once again, to confirm that the effects found were due to the experimental variables, individual linear regressions were done to test the effects of the possible confounds. Experiment 3 had six possible confounds – word length of the two stimuli, visual familiarity of the two stimuli, and visual complexity of the two stimuli. The only variable that had a significant effect was visual familiarity of the left-hand stimulus. The results of these analyses are collected in table 4.2.

**Figure 4.5: Residuals vs. condition vs. match**



All of the possible confounds were then put into a regression from which the residuals were calculated. A two-way ANOVA was then done on the residuals with modality condition (picture-picture, picture-word, and word-word) and response as the two factors. Both factors were found to be significant. For modality condition, F(2, 1990) = 7.44, p < .01; for response, F(1, 1990) = 15.59, p < .001. There was also a significant interaction of condition with match, F(2, 1990) = , p < .001. The mean residuals for each condition can be found in Table 4.3 and Figure 4.5.

**Table 4.2: Regression analyses of possible confounds**

| Variable | Coefficient (B) | t | p |
|---|---|---|---|
| Left side word length | 18.78 | 1.63 | .103 |
| Right side word length | 2.19 | 0.19 | .852 |
| Left side visual Complexity | -4.06 | -0.62 | .536 |
| Left side visual familiarity | -21.93 | -2.18 | >.05* |
| Right side visual complexity | -3.04 | -0.45 | .650 |
| Right side visual familiarity | 3.99 | 0.39 | .700 |

**Table 4.3: Mean residuals for Experiment 3**

| | Match | No-match |
|---|---|---|
| Picture-picture | -97.81 | 18.58 |
| Picture-word | -10.84 | 6.44 |
| Word-word | 5.10 | 90.99 |

Finally, we once again performed a linearity contrast on the reaction times. The no condition did not deviate significantly from linearity, F(1) = 2.64, p > .10. However, the yes condition did deviate slightly significantly from linearity, F(1) = 4.71, p < .05.

**Table 4.4: Mean reaction times (ms) of conditions in experiments 1 and 3**

| | | Experiment 1 | Experiment 3 |
|---|---|---|---|
| Match | Picture-picture | 874.09 | 1071.04 |
| | Picture-word | 1187.26 | 1161.91 |
| | Word-word | 1290.69 | 1186.97 |
| No-match | Picture-picture | 1090.12 | 1192.53 |
| | Picture-word | 1265.32 | 1192.93 |
| | Word-word | 1403.02 | 1275.72 |

The final question is how do the reaction times from Experiment 3 compare with those of Experiment 1. Table 4.4 and Figures 4.4 and 2.3 show how the means for the two experiments compare. The primary thing to note is that the primary difference in the reaction times is that the distribution of reaction times in Experiment 3 is flatter than the distribution in Experiment 1. The difference between the picture-picture and word-word means in Experiment 1 is 416.6ms in the match condition and 313.1 in the no-match condition, but in Experiment 3 it is only 115.93ms in the match condition and 83.19ms in the non-match condition. The primary source of the difference is that the picture-picture condition is slower in Experiment 3, but the word-word condition is faster in Experiment 3. It is also the case that if you compare Table 4.3 to Table 2.3, you can see that even the residuals for the two experiments follow the same pattern.

## 4.4 Discussion

The results from Experiment 3 do not seem to clearly indicate just how, exactly, functional information is stored in the semantic system. On the one hand, it is clear that visual input has slower access to functional information than it does to categorical information. Likewise, it would appear that verbal input has faster access to functional information than to categorical information. While the pattern of reaction times is linear, this particular pattern does not fit any of the predicted patterns discussed in the introduction to this chapter. Thus, it is worth exploring possible explanations for the pattern of reaction times witnessed in this experiment.

The one thing that I believe can be said for certain is that this pattern of results does not fit within the UCH framework. One can explain why pictures have slower access to functional information by arguing that the visual input only has direct access to, say, the prototype of the category, and thus an extra step is needed to retrieve the function of the category. However, if this were the case, the same argument would need to be made for verbal input as well. Presumably, verbal input would have direct access to name information, and an extra step would be required to retrieve functional information, meaning that the word-word trials should also be slower. However, this is not the case.

A more likely explanation is a variation on the MSH model presented above whereby functional information is stored with verbal information in a verbal/abstract semantic subsystem. In this scenario, it is clearly the case that pictures would have slower access to functional information than to categorical information as they would need to take an extra step to retrieve that functional information from the verbal/abstract subsystem.

The fact that verbal input seems to have faster access to functional information than to categorical information can be explained by the fact that the functional information is possibly more strongly and/or directly connected to a particular category node in verbal/abstract semantics than superordinate category information. This could potentially be the case for the man-made objects used in Experiment 3, as it would fit with the sensory-functional hypothesis discussed in section 1.2.2.4 above.

The results from these three experiments, taken as a whole, seem to provide clear evidence for the MSH model of how information is organized in semantic memory. In Experiment 1 we observed a pattern of results that could only be explained by either an MSH explanation of semantic memory structure or a task-switching effect. The results of Experiment 2 gave evidence that the effect witnessed in Experiment 1 were not the result of a task switching effect. Finally, Experiment 3 provided evidence that visual inputs have slower access to functional information than categorical information, while verbal inputs have faster access to functional information than categorical information.

Taken together the results from these three experiments are best explained by some hybrid of the models in Figures 2.2 and 4.3. In this model, different types of categorical information would be stored within each semantic subsystem. Visual categorical information would be stored in visual semantics, taxonomic categories in verbal semantics, etc. When trying to determine what category an object belongs to, we use whatever information is most readily available. If the object is a pictures, then we would use visual category information to classify them. Likewise for a word we would use verbal category information. Thus, when presented with two objects in the same modality (i.e. both pictures or both words), they will both access the same category node in either visual or verbal semantics as appropriate.

However, when presented with one item in each modality, the process of retrieving the category for each will require that different category nodes be accessed. Thus, unlike the same-modality condition, there will be no semantic priming effect, and thus the picture-word condition is slower than it would be if there was a strict linear relationship across the three trial types. Furthermore, we can be reasonably certain that this cost is not simply due to some sort of task-switching effect, but is an actual reflection of how categorical information is organized in semantic memory.

Functional information, on the other hand, most likely resides in verbal semantics and is tied more directly to the name of the object. Thus, while looking up the superordinate category of a word takes a certain amount of time, looking up the function of that same word takes less time because the function is a feature of the object itself, and thus is part of the object's representation. The superordinate category relationship, on the other hand, is encapsulated by a link between the word's concept and the superordinate concept, *a la* Collins and Quillian (1969).

This provides the most parsimonious explanation for both the results of Experiment 1, particularly for the fact that the match condition shows a deviation from linearity while the no-match condition does not, and the results of Experiment 3. In particular, it explains why, rather than seeing a consistent difference across conditions when comparing Experiment 3 to Experiment 1, we see a dissociation where visual information is processed fast in Experiment 1, but verbal information is processed faster in Experiment 3.

## CHAPTER 5: MODELS OF SEMANTIC MEMORY

Several different models of semantic memory structure and function have been proposed. The vast majority of them are variations of the two major frameworks that we have been focusing on thus far – the unitary content hypothesis and the multiple semantics hypothesis. This chapter will review several of these models. Thus far, we have seen several reasons to be skeptical of the UCH approach to understanding semantic memory. Thus, we will start by examining those models.

### 5.1 Unitary content hypothesis models

#### *5.1.1 Ancient History*

The original semantic memory models were designed primarily to explore how categories are structured – i.e., how are subordinate and superordinate categories connected to each other, do concepts inherit features from their superordinates, etc. (e.g., Collins and Quillian, 1969; Pylyshyn, 1973; Fodor, 1983). Thus, these models tended to take the form of relatively strict versions of the UCH – semantic memory exists as an abstract, amodal store where information is organized by category. As a result, the primary debate centered around how those categories were stored (Komatsu, 1992), which is a discussion I do not intend to recap here.

There were two lines of evidence that forced significant revisions to be made to the standard UCH models. The first involved the research discussed in Chapter 1 that showed that, at least with regards to how they accessed different semantic information such as names and categories, pictures and words were processed differently. The second was the reports of patients with brain damage that were exhibiting modality-specific semantic deficits, such as optic aphasia, which demonstrated that different input modalities, such as vision and touch, could be denied access to selective parts of the semantic system. For example, patients suffering from optic aphasia are unable to access name information when an object is presented to them visually, but when it is presented tactilely they have relatively preserved access to name information. The challenge for supports of the UCH was thus to generate models explaining these discrepancies.

### *5.1.2 O.U.C.H.*

To accomplish this goal, Caramazza et al. (1990) have proposed the Organized Unitary Content Hypothesis (or OUCH) to attempt to explain the results from optic aphasic patients. Caramazza et al. base their argument on two observations. The first is that one of the primary issues with regards to explaining optic aphasic patients is that different input modalities seem to have *privileged access* to different types of information in semantic memory. For example, visual input would have privileged access to visually related semantic information, such as shape and color, while verbal input has privileged access to naming information. The multiple semantics hypothesis instantiates this privileged access by positing that information within a particular semantic subsystem can only be directly accessed by inputs from the appropriate modality. Thus, the visual semantic subsystem can only be directly access from visual inputs. The question is whether one needs to theorize that there exist modality-specific subsystems in order to get privileged accesibility.

Caramazza et al. claim that the answer is no. They hypothesize that what is perceived as privileged access of, for example, visual inputs to visual semantic information is a result of where the visual semantic information comes from. Words bear no systematic relationship to the objects that they name. As he puts it, it is a “historical accident” that any word is associated with any particular object. Visual properties, on the other hand, bear a very systematic relationship to the objects that they are associated with because they are tied to specific perceptual experiences with the object itself. The fact that an apple is called “apple” is not related to the properties of the apple itself, such as the fact that it is round or that it is red or that it has a somewhat tangy taste. Thus, Caramazza et al. argue that not only does the wholistic perceptual representation of an object allow us to access that object’s semantic representation, but our perceptual representations of the features of the object can also be used to index the semantic representation of an object in a fashion that the features of a word (i.e. the phonemes) cannot. Therefore, simply by virtue of having a more direct link to the real world, perceptually oriented semantic information will have a stronger connection to perceptual inputs than naming information will.

The second observation that Caramazza et al. make is to note that the second primary assumption that the multiple semantics hypothesis is trying to satisfy is the notion that there appear to be *privileged relationships* between disparate pieces of the semantic system. In other words, though naming information cannot be accessed from visual input in patients with optic aphasia, miming and categorization information can, implying a privileged relationship (i.e. stronger link) between visual inputs and miming or categorization information.

Carmazza et al. argue that in order to explain privileged relationships one need merely assume that “semantic representations are internally structured.” The primary implication of this is simply the notion that some parts of an object's semantic representation may be more strongly associated with other parts, again potentially due to the systematicity of their relationship. Thus, visual semantic information may be more strongly associated with information on how an object is used than it is to naming information because function is systematically related to form, but name information is not.

Based on these two principles, privileged access and privileged relationships, Caramazza argues that generic damage to the semantic system may result in the pattern of deficits observed in patients suffering from optic aphasia precisely because these organizing principles lead to generally stronger links between visual perceptual input and semantic information that is systematically related to visual perceptual information, such as gesturing. Because these links are stronger, they are more robust and thus more likely to survive diffuse damage to the semantic system. Name information, on the other hand, is not systematically related to visual perceptual input, and thus, the connections between visual input and name information would be more tenuous, and therefore more likely to disappear or be rendered unusable by damage to the semantic system.

Caramazza refers to this as the Organized Unitary Content Hypothesis (or OUCH) to denote that, while it is a unitary theory of semantics, it assumes that there is some organization within the semantic representations that can account specifically for the pattern of deficits observed in patients with optic aphasia. At a higher level than the object level, however, it still maintains the core assumptions that representations are amodal and abstract, and that they are organized by category. What this means is that, while the OUCH model can potentially explain how one gets optic aphasia from diffuse damage to the semantic system, it is not at all clear that this is the type of damage all of these patients have. The patient that Beauvois (1982) reported had a localized left temporal lesion. Any sort of damage localized within the semantic system described by the OUCH model should yield category-based deficits, not modality-based deficits.

Essentially what the OUCH model is doing is creating a UCH model in which different input modalities are treated differently at the semantic level. Privileged access is relatively straightforward to understand. It can be thought of simply as very strong connections between a particular modality of input and information related to that modality, such as visual input and color information, and weaker connections between a modality and information not related to that modality, such as vision and name information. Optic aphasia could then simply be the result of noise in the visual inputs. Since noise will affect weaker connections first, the ability to retrieve name information from visual inputs will be significantly impaired.

The notion of a privileged relationship is much less specified. Caramazza et al.'s characterization of privileged relationships makes it sound like they also can be characterized as direct connections between a particular modality of input and a particular type of information, such as visual input and miming information. This is because they describe privileged relationships as strong links between inputs and particular types of information.

If this were the case, though, then why would we need to bother differentiating between privileged access and privileged relationships? On possibility is to argue that there are multiple layers to the semantic system such that, for example, visual input is more strongly connected to sensory-based information such as color and shape. This information is then more strongly connected to the next layer of abstraction, which may contain information such as miming, than to other information less systematically related to vision, such as name information. However, in this scenario, there is no compelling reason to believe that visual inputs should have direct, albeit weaker, connections to name information, and thus the explanation of optic aphasia a noise across the visual inputs becomes significantly less compelling. Furthermore, a “layered” semantic system starts to look and sound suspiciously like an MSH model rather than a UCH model of semantic memory. Certainly, in order to explain the data, Caramazza et al. have developed a UCH model that is, in virtually all respects, functionally equivalent to the MSH. The primary difference is solely the anatomical organization of semantic memory, and not the functional organization.

It also is not entirely clear how the OUCH model can account for some of the data for normal participants that has been reviewed in this paper. On the one hand, one can envision that it can explain the naming/categorization double dissociation by simply assuming that visual input has stronger links to category information than verbal input. It is not entirely implausible to believe that visual input and category information are more systematically related to each other as we tend to categorize objects based on what they look like. However, that begs the question of how the model would explain the cross-modal priming of categorization information data. If it were the case that visual inputs had intrinsically stronger connections to category information, then they should always provide better priming regardless of whether the second stimulus is a word or a picture because they would provide greater activation to the category information than the words would. As noted above, though, this is not the case, or at least the data is not clear on this point.

Finally, even with the ideas of privileged access and privileged relationships built in, the OUCH modal simply cannot account for the results of the three experiments presented earlier in this dissertation. The OUCH model still falls within the UCH framework, and thus the superordinate category information being accessed from both the visual and verbal inputs is the same information. Thus, there should be no difference in the semantic priming effect between the like-modal and cross-modal conditions. Furthermore, if the speed advantage that pictures have for accessing category information is solely due to privileged access or a privileged relationship, that should be a relatively constant advantage. Either way, if the OUCH model were accurate, we would expect there to be a linear distribution of reaction times in Experiment 1, which there is not.

### *5.1.3 Glaser (1992), Glaser & Glaser (1989)*

Glaser has developed a model of semantic memory based on his examinations of the processes involved with picture naming and with using stroop-like experiments (Glaser, 1992; Glaser & Glaser, 1989). His model is similar in structure to Paivio's dual code theory, but not in function. Glaser hypothesizes that there is one component of long-term memory responsible for storing information related to perceptual features and physical objects and another responsible for storing verbal information (see Figure 5.1). He refers to them as "semantic memory" and the "lexicon" respectively, and this distinction is important. Unlike dual-code theory, in Glaser's conception the lexicon (verbal memory) does not contain any semantic content. All of the semantic content resides in the semantic memory, and the lexicon only contains linguistic information needed to read, write, and understand speech. Furthermore, as with Paivio's dual-code theory, perceptual inputs have direct access to the semantic memory while words have direct access to the lexicon.

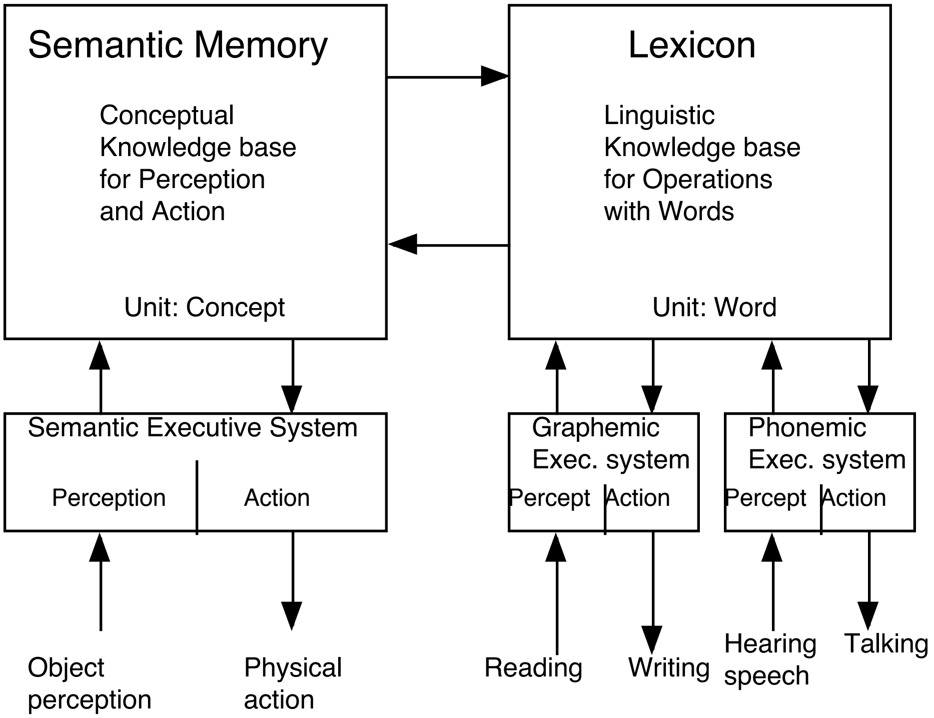


**Figure 5.1: Glaser's model of semantic memory**

(adapted from Glaser &Glaser, 1989, Figure 5a)

However, Glaser's semantic memory is not comparable to Paivio's imaginal code system for storing perceptual information. The semantic memory in Glaser's model is a relatively strict version of the UCH. Information is stored in semantic memory based on concepts rather than modality, and information from different modalities will access the same conceptual information.

Glaser's model can quite easily explain much of the data from normal people. Given that superordinate category information is stored in semantic memory and name information in the lexicon, the naming/categorization latency differences between pictures and words are quite clearly going to be related to the fact that words and pictures each have privileged access to different systems. Furthermore, the Stroop-like effects of overlaying pictures with words can also be explained by this model (see section 1.3.1.1 above). It can also rather successfully explain patients with category-specific deficits, based primarily on the fact that the semantic memory is organized categorically.

However, the model is unable to account for two pieces of data that have been discussed at length above. The first is modality-specific aphasias such as optic aphasia. In Glaser's model, every input modality's access to name information is mediated by a conceptual representation in semantic memory. In other words, regardless of whether you are looking at the pencil or holding it in your hand, you must first access the PENCIL concept in semantic memory before you can access the name "pencil." Thus, the model cannot account for an inability to name objects presented visually while the ability to name them from another input modality, such as touch, is preserved. Any damage to the connection between the semantic memory and the lexicon should affect all input modalities equally.

Secondly, the model is unable to explain the data presented in the three previous chapters. In particular, since words and pictures are both accessing the same category node in the semantic memory, the reaction times in Experiment 1 should have a linear distribution, which they do not. Furthermore, since functional information is presumably semantic and not merely linguistic, the results from Experiment 3 are inconsistent with this model. One would expect that either there would be no difference between functional categorization and taxonomic categorization or, if there is a difference, that all three conditions should be universally slower. The only way this model can explain the data from Experiment 3 is for it to presume that functional information is in some way strictly linguistic, and thus would be stored in the lexicon instead of in the semantic memory.

### *5.1.4 Summary*

The UCH models presented here make significant modifications to the basic theory in order to account for the data that is in existence. However, none of these modifications allow for these models to account for the data presented in Chapters 2-4 of this thesis. It is also not at all clear that further modifications to these models will allow them to account for this data without effectively changing them into MSH models. The OUCH model already faces this problem in its attempts to explain modality-specific deficits such as optic aphasia without theorizing direct connections from all modalities of input to all the information contained within semantic memory. The best system that could be settled on is one that retains the core OUCH principles of privileged access and privileged relationships, but has a layered architecture where the more abstract the piece of information, the farther from any particular input it resides, yet still maintaining a categorical organization.

## 5.2 Multiple semantics hypothesis models

### *5.2.1 Dual-code theory*

Probably the most well known multiple semantics model is Paivio’s Dual-Code theory (Paivio, 1971). Prior to Paivio proposing his hypothesis, most theories of semantic memory had assumed that the information was represented in a propositional fashion that was organized by category. Thus, the information that the car is blue and the information that the car needs gasoline in order to work would be stored in the same format, such as COLOR(CAR, BLUE) and NEEDS(GAS, TO RUN). The nature of the organization of information, such as a tree-like network (Collins & Quillian, 1969) or based on features (Smith, Shoben, & Rips, 1974), was the primary source of discussion rather than the nature of the information itself. In all of these models it was assumed that semantic memory was abstract and amodal.

Paivio proposed an alternative hypothesis that claimed that we used at least two different types of codes to represent information in semantic memory. One set of codes were used to represent information related to perceptual features and were called *analogue* or *imagistic* codes because they stored information as images, or in a format that preserved the perceptual features of the images being stored. We store multiple types of analogue corresponding to the various sensory modalities, and these are the codes we use to recall perceptually based information such as shape, color, feel, or sound. The other set of codes represented verbal and abstract information and were called *symbolic* codes because they stored information in a format not necessarily related to the object actually being represented, in much the same way that the word “dog” is simply a symbol and has an arbitrary relationship to an actual dog.

Paivio then proposed two different storage systems in semantic memory to store the two separate types of codes – a verbal system containing the symbolic codes and a non-verbal system containing the analogue codes. Each part contains modality-specific information, with the verbal system having word based information and the non-verbal system having sensory based information. Representations are linked both within and between subsystems.

Dual-code theory can explain the cognitive literature reasonably well by simply by making two assumptions. First, that name information is in the verbal subsystem and category information is in the visual subsystem (Johnson, Paivio, & Clark, 1996), and secondly by assuming that the visual inputs to semantic memory can more directly access information in the visual subsystem and the verbal inputs more directly access information in the verbal subsystem. Thus, it becomes apparent why people are faster to access categorization information from pictures than words, but vice versa for accessing name information.

How a multiple semantics system accounts for the cognitive neuroscientific data, and the objections that that data raises against multiple semantics hypotheses, is better discussed in the context of models proposed by those researchers, so we will put off those discussions until then.

### *5.2.2 Multiple Semantics Hypothesis*

When Beauvois (1982) presented the data on her optic aphasic patient, she argued that the cause of the impairment exhibited was due to damage to the connection between the visual and verbal semantic subsystems, and that this damage was bi-directional.

She cites the fact that patients are impaired both at naming objects presented visually and at picking out the proper object from an array of objects when presented verbally with the name as evidence for the fact that the damage is bi-directional. She cites the fact that verbal processes in general seem relatively unimpaired, e.g. people can generate names when presented with the objects in other modalities. Furthermore, when presented with an object visually, it is not so much that patients failed to generate a name as that they generated the wrong name, such as a semantically related object or a category name instead of the correct name for the object presented. Thus, verbal processes up to the level of verbal semantics seem to be reasonably intact. Finally, because patients had both preserved gesturing when presented with objects visually and visual imagery, she argues that visual processing up to the level of visual semantics appears to be preserved. Thus, the impairment must be the result of the link between separate visual and verbal semantic systems.

Shallice (1987) combines the information about optic aphasia with two other lines of evidence – namely that patients suffering from semantic access dyslexia show modality-specific priming and the modality-specific deficits  other semantic memory disorders, such as semantic dementia, sometimes show. He argues that these three pieces of evidence taken together necessitate that a semantic system with modality-specific subsystems must exist because no unitary system could show modality-specific priming, modality-specific aphasias, and modality-specific effects in degraded-store deficits like semantic dementia. In particular, he argues that if semantic memory is unitary and amodal, then damage to semantic memory should affect the representation of a particular piece of information regardless of the modality it is being accessed from. Thus, name information about an object should be unavailable regardless of whether the patient is looking at the object or holding it in their hand. Likewise, if a semantic dementia patient is unable to categorize an object when presented with its name, she should be unable to categorize the same object presented visually. Finally, pictures and words should provide equivalent priming for semantic representations in a unitary system, and thus we should not see modality-specific priming in patients with semantic access dyslexia.

I'll save the response from proponents of the unitary semantics models until I actually discuss those models. Even setting that aside, there is still the issue of patients who have category-specific deficits to be reconciled with the multiple-semantics hypothesis. Specifically, how is it that we see patients with semantic deficits that seem to be specific to one category versus another regardless of the modality in which the category is being accessed?

Shallice (1988a, 1988b) has responded to this by arguing that what is actually happening in a patient with a category specific deficit is that the patient actually has a modality specific deficit, and the category that they are also impaired with relies more heavily on that particular modality to represent its information. For example, the prototypical category specific deficits involve patients that are impaired at processing living things, but have relatively preserved behavior (naming, categorization, etc.) with non-living things. Shallice (1987) argues that most of the information about living things is stored in visual semantics, primarily because we tend to, for the most part, interact with them on a purely visual level. We look at the elephant in the picture, but we rarely get to pet one. Non-living things, on the other hand, spread their information across many more modalities because we interact with them in more modalities – on a functional level (which he places in verbal semantics), through touch, vision, audition, etc. Thus, when a patient shows a deficit in processing only living things, he really has an impairment in visual semantics, which, because of the fact that large pieces of the representations for living things are stored in visual semantics, means that living things appear to be selectively impaired relative to non-living things. The problem with this argument is that we would also expect to see deficits in processing visual information, such as an inability to name pictures, for the non-living things, which do not appear to be present (Caramazza et al., 1990; Plaut, 2002).

Thompson-Schill, Aguirre, D'Esposito, and Farah (1999) provide neuroimaging data of normal participants that seems to support Shallice's explanation of category-specific deficits. Specifically, they examined the role of the left fusiform gyrus (LFG) in accessing semantic information about living and non-living things *vis a vis* the visual/non-visual information distinction. They asked participants both visual and non-visual questions about both living and non-living things and found that when they were asking questions about living things, the LFG showed increased activation for both types of questions, but when the questions were about non-living things, the LFG only showed increased activation for the visual questions, but not for the non-visual questions.

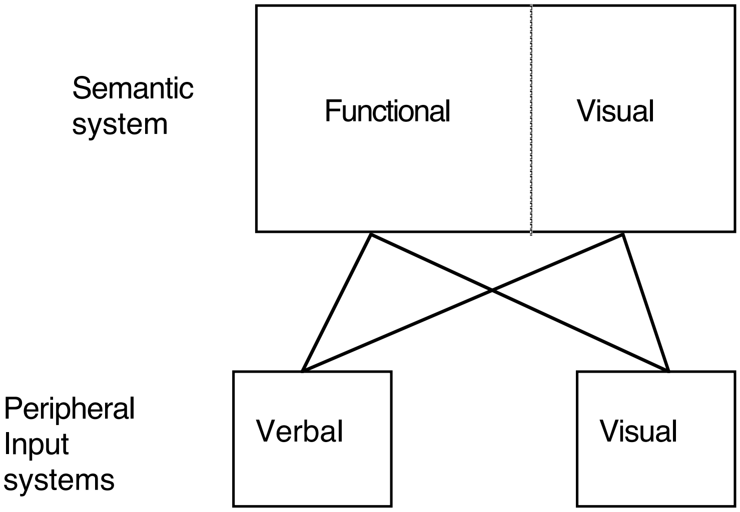
Their argument for why this would be the case actually takes Shallice's one step further. They argue that not only are the semantic representations for living things more heavily biased towards features contained in visual semantics, but they are *so* heavily biased that, in effect, all semantic access of information about living things is mediated by the visual semantic system. Thus there is LFG activation even when asking non-visual questions about living things. Non-living things, on the other hand, would have enough of their semantic representation contained in areas outside of visual semantics that visual semantics is not required for accurate access of semantic information about htem. One could potentially argue that this would imply that relatively minor damage to the visual semantic subsytem, minor enough to, for example, leave naming of pictures of non-living things relatively intact, would have a deleterious effect on a patient's ability to semantically process information about living things.

### *5.2.4 Farah & McClelland's PDP model*

Farah & McClelland (1991) built a parallel distributed processing (PDP) model based on the sensory-functional theory (SFT; see section 1.2.2.4 above). Their model had separate verbal and visual input systems and a semantic layer with its units divided between visual semantics and functional semantics in a 3:1 ratio. Living things were represented in the semantic layer with an average ratio of visual to functional units of 16.1:2.1 while non-living things were represented with an average ration of 9.4:6.7. They found that if they lesioned the visual semantic units, the model was impaired at processing semantic information about living things and if they lesioned the functional semantic units the model was impaired at processing semantic information about non-living things.

It is unclear how much Farah and McClelland's model can be generalized to the other phenomena discussed above because their model differs from the standard versions of the multiple semantics hypothesis in a rather significant fashion. Normally, multiple semantics models hypothesize that an input modality only has direct access to its corresponding semantic modality. For example, visual input would only have direct access to visual semantics, and likewise visual semantics could only be directly accessed via visual input. Farah and McClelland's model eliminates this restriction and provides for direct access of functional semantics from visual input and of visual semantics from verbal input (see Figure 5.3).

**Figure 5.3: Farah & McClelland (1991) PDP model of semantic memory**



(Based on Figure 1 of Farah & McClelland (1991))

Under such a system, it is difficult to see how optic aphasia can arise in brain-damaged patients. While theoretically possible there would have to be damage both to the connection between visual and verbal semantics as well as the connection between visual input and verbal semantics. Unless it could be shown that these pathways are anatomically close together, this explanation would be, at best, unparsimonious. One could make the counter-argument that, based on Thompson-Schill et al. (1999) that visual semantics has to mediate all semantic retrievals involving living things, but optic aphasics' naming deficits are for all objects, not just living things.

This version of the MSH is also effectively equivalent to the OUCH model, except that the visual and verbal semantic information is anatomically separated, whereas the OUCH model mixes them in together.

### *5.2.5 Summary*

In general, the versions of the multiple semantics perspective presented here can account for much of the data presented above. If one assumes that, for whatever reason, pictures in general have easier access to semantic information than words do, then the data on picture/word categorization differences, increased semantic relatedness effects for pictures, and increased ability to remember words follows relatively straighforwardly.

One piece of data still problematic for the multiple semantics accounts to explain is that cross-modal priming data provided by Guenther et al. (1980). They found that when participants were making categorization judgments about objects (i.e. is the object a living or non-living thing) that were presented as both pictures and words, that, while cross-modal priming existed, it was essentially equivalent regardless of whether words or pictures were being presented first. Furthermore, it was equivalent regardless of whether the objects presented were living things or non-living things. According to the standard multiple-semantics account of category-specific deficits in brain-damaged patients, at the very least one would expect there to be greater priming in the like-modal conditions for living things when both items are pictures and non-living things when both items are words, but this is not the case. In both cases there is greater priming for living things. While, though the difference is admittedly greater when both stimuli are pictures.

Furthermore, one could argue that, according to the multiple semantics theory of category-specific deficits, there should also be greater priming for living things when the first item is a picture regardless of the modality of the second item, and likewise for words priming knowledge of non-living things. If living things rely so much more on that part of their representation that is stored in visual semantics, then one would presume that seeing a picture of a living thing would activate this entire representation, thus priming a much larger portion of the overall representation than when a word is presented first, and thus there should be a larger priming effect for living things when a picture is presented first. One can make a symmetrical argument regarding words priming non-living things.

On the other hand, one can argue that precisely because living things rely more heavily on their visual semantic components as the bulk of their overall representation in semantic memory, then that would imply that, in all likelihood, categorical information would also be stored in visual semantics, and thus even when a word is presented first, it needs to activate the entire visual semantic representation in order to access that categorization information, and thus the following stimulus would still benefit from the same priming effect as if a picture had been the first stimulus. In this case, then, one would still expect, all other things being equal, that the pictures should be responded to faster for living things than the words and vice versa for non-living things. There is, however, no such interaction, and in fact participants responded slower across the board to non-living things than to living things, though the difference was more pronounced when the first item presented was a picture than when it was a word.

# CHAPTER 6: SPLITTING THE DIFFERENCE

As noted in several places in Chapter 5, once one makes the changes necessary for versions of either the UCH or the MSH to explain the available data, what you end up with is a set of models that are all essentially equivalent to each other. Functionally, the models end up being the same. The primary differences exist in how information in semantic memory is anatomically organized. However, even on this dimension, there is some significant convergence among the models towards something that isn't quite a unitary content hypothesis, but isn't quite a multiple semantics hypothesis, either.

We particularly started to get a hint of what appears to be the best model to account for the available data when discussing the changes to the OUCH model that were necessary to reconcile it with much of the existing data. Namely, a system that is unitary in the respect that it lacks an overall organization along any particular dimension, but has multiple semantics features in that its structure is layered such that information more closely associated with a particular perceptual modality will tend to be strongly associateed (i.e., have "privileged access") to the inputs from that modality while more abstract information will not be more likely to be associated with any particular input modality.

What we seem to have converged on is a system that takes advantage of the relationships between input modalities and particular types of information such that these types of information are very strongly linked to specific input modalities. For example, shape and color information are going to be very strongly related to visual input because they are very systematically related to that input. Likewise, name information is going to be strongly related to verbal inputs for the same reason.

This organizational system builds up through various levels of abstraction such that, while each successive layer is less systematically related to any particular input modality than the layer before it, i.e. is more abstract, it is still very strongly related to the information contained in the previous layer (see Figure 6.1). Thus, some information seems to exist in semantic memory in a very modality-specific fashion, such as in the MSH. However, other information that is less strongly tied to a particular input modality appears to reside in semantic memory in a more abstract, amodal fashion, *a la* the UCH.

## 6.1 EPAM

I set out with the goal of building a model that encapsulated these relevant features using the EPAM framework (Feigenbaum, 1958). EPAM uses a decision-tree architecture to model the processes and structures contained in memory. Essentially, EPAM takes as input an ordered list of features. At each level of the tree, it tests a feature or group of features in the list and, based upon the results of the list, traverses down a particular branch of the tree until it settles on a decision as to what that list of features represents.

The decision was made to attempt to structure an MSH model in EPAM with multiple decision trees, each representing a modality-specific semantic subsystem. An immediate problem arose in that EPAM appeared to not be powerful enough to provide an MSH model due to issues with how to connect the various semantic subsystems. It would not be sufficient to simply connect the relevant leaf nodes to each other as those would essentially be where exemplars would be stored. Thus, while it was possible to connect the leaf node BEAGLE from visual semantics to its appropriate name node in verbal semantics, it was not possible to connect the superordinate category node DOG from visual semantics to its name node in verbal semantics. Finally, EPAM has no mechanism for learning such associations. EPAM is very good at learning things that can be represented as decision trees, but learning these types of complex associations is, at best, very difficult for it.

Creating a UCH model in EPAM also proved to be an unfruitful direction in which to proceed as EPAM has no mechanism allowing for inputs from multiple different locations, meaning that inputs from different modalities could not be represented. While the issue could be fudged by having the first node in the decision tree test for what modality the input was in, I then found myself with the same problem I had encountered attempting to create an MSH model in EPAM. This is because this initial test would effectively create a *de facto* MSH architecture, with the attendant problems mentioned above.

Consequently, the attempt to create an EPAM model of semantic memory based upon the data and issues outlined thus far in the thesis was abandoned and I started to explore neural network models of semantic memory.

## 6.2 Graded specialization

Fortunately, there is one last model that I have not discussed yet in this thesis that instantiates precisely these properties into its conception of how the semantic system is organized (McGuire & Plaut, 1997; Plaut, 2002). This model presumes that semantic memory starts out with no *a priori* organization, either based on categories or on input modalities. As mentioned in Chapter 1, the primary purpose of semantic memory is to map multiple input modalities onto multiple output modalities. To accomplish this goal, two assumptions are made. The first is that Caramazza et al.'s (1990) notion of privileged access and privileged relationship will arise in a neural network due primarily to the systematicity of the relationship between, for example, visual input and shape or color information. The second assumption is that neurons like short connections. Thus, rather than having a fully connected network, units are only connected to other units that are close by.

Thus, in training this model, a semantic system develops where units that are closer to the visual inputs respond more selectively to visual information, or in other words are modality-specific for vision, and units that are closer to the tactile inputs are modality-specific for touch, etc. Units that are close to the phonology outputs are more specialized for naming while units that are closer to the action outputs are more specialized for gesturing. Units that are equidistant from both inputs respond equally to both types of inputs, or in other words in an amodal fashion. Consequently, the system naturally sorts itself out so that damage to connections between the visual inputs and phonological outputs leads to the types of behaviors evidenced in patients with optic aphasia, namely impaired naming with preserved gesturing from vision and both preserved naming and preserved gesturing from touch.

Essentially, this model is the OUCH model with the semantic system layered based on abstractness in order to fully reflect the effect of privileged relationships (see Figure 6.1). As a model based upon the OUCH model, it is able to easily explain much of the same data. It explains optic aphasia through the existence of damage to the connections from the visual inputs to those parts of semantic memory near the phonological outputs.

Category-specific deficits can be explained by simply applying the notion of privileged relationships to the fact that pieces of information from the same category are going to have very privileged relationships to each other, and thus are likely to be topographically close to each other in this layered system. If each layer represents a horizontal dimension of the abstractness of the information contained within the layers, then categories could easily be encapsulated as vertical dimensions of difference. Since the model was specifically designed to provide an account of optic aphasia, this prediction has not been explicitly tested, but there is no reason to believe it to be implausible based on the other underlying assumptions of the model.

How well does the model account for the data from normal people? Explaining the naming/categorization differences for words and pictures is a relatively straightforward result of naming information residing closer to the verbal inputs while category information would be more abstract, and potentially closer to the visual inputs. Likewise, the Stroop-like effects of pairing up words and pictures and the differences that result depending on whether the task is to name or categorize the stimulus can be explained in a similar fashion.

The major question is whether the model can account for the data presented in Chapters 2-4 above, which has proved to be the most troublesome data to explain under the UCH framework. Since the model presented in Plaut (2002) was designed to provide an account of optic aphasia, it only provides for naming and miming outputs, and thus cannot be directly applied to the data presented in Chapters 2-4. Extrapolating, however, we can see that the model has just enough MSH qualities to potentially provide a fit for the data in Experiment 1.

You might recall that the primary difficulty any UCH model has in explaining why there is a deviation from linearity across the three modality conditions hinges on the fact that every UCH model posits that, in normal people, the semantic information accessed is the same regardless of the input modality. Thus, regardless of whether the input is a picture or a word, they would both end up accessing the same superordinate category information. Therefore, there cannot be a conversion step needed, which would cause the non-linearity. There also should not be differential semantic priming between the like-modality and cross-modality conditions, which is the most likely explanation of the non-linearity. Put another way, the data from Experiment 1 very clearly seems to indicate that, absent any other task demands, words and pictures will access different semantic representations of superordinate categories. This is a difference that simply cannot be accounted for under the UCH.

The graded specialization modal can account for this data by presuming that, in fact, the representations of the superordinate categories being retrieved are not the same. This fits with the underlying assumption of the graded specialization model that information is associated with other information based on how systematically the two pieces of information are related to each other. Along these lines, one could argue that visual inputs will access superordinate categories based on visual similarity – similar to a prototype. This follows from the assumption of privileged relationships since things that look like tend to be categorized together, especially in the natural world. Thus, the systematicity of the visual similarity will lead to there being superordinate category representations that are close to the visual inputs.

Verbal inputs, on the other hand, would need to access a more abstract representation of superordinate category. Just as there is no systematic relationship between a word and the object it represents, neither is there a systematic relationship between a word and the superordinate category of the object the word represents. Thus, superordinate category information for words will necessarily have to exist at a more abstract layer than it would for pictures. Thus, words and pictures would be accessing different categorical representations. This could lead to there being a semantic priming effect in the like-modal conditions that does not exist in the cross-modal conditions, which explains the deviation from linearity found in Experiment 1.

The graded specialization model has a less easy time accounting for the data in Experiment 3. It is easy enough to explain why pictures take longer to categorize based on function simply by presuming that the function an object serves, while related to its form, is going to necessarily be more abstract. Consequently, functional information would reside at a deeper layer of semantic memory in relation to visual inputs than superordinate category information would.

What is less easy to understand is why words should be relatively faster at retrieving function information than superordinate category information. Presumably, the data indicates that function information is more systematically related to words than superordinate category information. While not an obvious proposition, it is not an absurd one, either. One can imagine that when we refer to a man-made object, it is typically in terms of either needing a particular function to be fulfilled, as in, “Do you have a pencil?”

or, “Bake in the oven at 350° for 35 minutes.” It is with less frequency that we think of man-made objects in terms of their superordinate categories, such as, “I need new furniture,” or, “These kitchen utensils stink.”

The brain imaging data discussed in section 1.3.1.5 above also seems to provide somewhat compelling evidence in favor of the graded specialization model. Martin and Chao (2001) review a number of studies providing evidence for selective activation of various brain areas depending upon the category of information being processed, and also depending on the type of information being processed. For example, Martin et al. (1996) found that areas of the brain adjacent to premotor cortex were selectively activated when participants were naming tools while areas adjacent to visual processing centers became activated when animals were named. Lest we think that it is simply the case that conceptual information about tools is stored close to premotor cortex, though, Martin et al. (1995) found that brain areas adjacent to color processing centers were selectively activated when participants were generating color words about tools, and areas close to motion processing centers were activated when participants generated action words.

The result that is not as frequently discussed, and which is of interest to us here, is that while there do appear to be areas of the brain that become selectively active depending upon the category and type of information being processed, there is also a very significant area of the brain that is activated for any type of semantic processing, regardless of the category or the type of information being processed (Tyler et al., 2003). This is precisely the pattern of activation that one would expect from the graded-specialization model: peripheral areas of semantic memory that are specific to input and output modalities, and central areas that are more modality-general.

## 6.3 The grand finale

Thus, we have reached a point where we can say with a reasonable degree of confidence that the most likely model of how information is organized in semantic memory is a blend of the MSH and UCH approaches. We can envision a system, such as the graded specialization system, whereby semantic memory is organized along two primary dimensions (see Figure 6.2). The first dimension is based on input modality and involves a layered topography. Those layers closest to the inputs from a particular modality will be more specific to that modality. The more layers you get away from a particular modality, the more abstract the information contained within a given layer.

The second dimension is based on categories and cuts vertically through the layers of abstraction. Information from a given category is necessarily going to be more strongly connected to other information from that category than to information from other categories, which leads to this particular dimension of organization.

This conception of the organization of information in semantic memory is able to account for the widest variety of data, both from cognitive psychology as well as cognitive neuroscience. It is also able to account for the data from the three experiments presented in this thesis that stricter versions of both the UCH and MSH had difficulties explaining without resorting to very *ad hoc* explanations of how specific pieces of information were stored.

# REFERENCES

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 80 (3), 1-46.

Beauvois, M. (1982). Optic aphasia: a process of interaction between vision and language. Philosophical Transactions of the Royal Society of London. 298, 35-47.

Bloom, P. (1996). Intention, history, and artifact concepts. Cognition, 60(1), 1-29.

Bousfield, W. A., Esterson, J., & Whitmarsh, G. A. (1957). The effects of concomitant colored and uncolored pictorial representations on the learning of stimulus words. Journal of Applied Psychology. 41, 165-168.

Caramazza, A. (2000). Minding the facts: a comment on Thompson-Schill et al.'s "A neural basis for category and modality specificity of semantic knowledge." Neuropsychologia, 38, 944-949.

Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? Cognitive Neuropsychology. 7, 161-189.

Caramazza, A. & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. Neurocase Special Issue: Category-specific deficits, 4(4-5), 339-351.

Cohen, J. D., MacWhinney, B., Flatt, M. & Provost, J. (1993).PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. Behavior Research Methods, Instruments, & Computers, 25(2), 257-271.

Collins A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior. 8, 240-248.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. Psychological Review, 100(4), 589-608.

Ehri, L. C. (1976). Do words really interfere in naming pictures? Child Development, 47(2), 502-505.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. Psychological Review, 87, 215-251.

Farah, M. J. & McClelland, J. L. (1991) A computational model of semantic memory impairment: modality specificity and emergent category specificity. Journal of Experimental Psychology: General. 120, 339-357.

Feigenbaum, E. A. (1960) An information processing theory of verbal learning. Unpublished doctoral dissertation, Carnegie Institute of Technology.

Fodor, J. A. (1983). Modularity of mind: an essay on faculty psychology. Cambridge, MA: MIT Press.

Forde, E. M. E., & Humphreys, G. W. (1999). Category-specific recognition impairments: A review of important case studies and influential theories. Aphasiology. 13(3), 169-193.

Fraisse, P. (1969). Why is naming longer than reading? Acta Psychologica, 30, 96-103.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. Psychological Bulletin, 123(1), 71-99.

Gainotti, G. & Silveri, M. C. (1996). Cognitive and anatomical locus of lesion in a patient with a category-specific semantic impairment for living beings. Cognitive Neuropsychology, 13(3), 357-389.

Glaser, W. R. (1992). Picture naming. Cognition. 42, 61-105.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. Journal of Experimental Psychology: General. 118, 13-42.

Guenther, R. K., and Klatzky, R. L. (1977). Semantic classification of pictures and words. Journal of Experimental Psychology: Human Learning and Memory. 3, 498-514.

Guenther, R. K., Klatzky, R. L., and Putnam, W. (1980). Commonalities and Differences in Semantic Decisions about Pictures and Words. Journal of Verbal Learning and Verbal Behavior. 19, 54-74.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996) Cognitive components of picture naming. Psychological Bulletin. 120, 113-139.

Komatsu, L. K. (1992). Recent views of conceptual structure. Psychological Bulletin, 112(3), 500-526.

Komatsu, L. K., Kerner, R. S., Lee, K. P., Thompson, J. R., & Gottlieb, J. F. (1997). Creators' Intentions do not play a major role in artifact categorization. Annual meeting of the American Psychological Society (APS). Washington, D.C.

Kosslyn, S. (1994). Image and brain: The resolution of the imagery debate. Cambridge, MA: The MIT Press.

Lauro-Grotto, R., Piccini, C. & Shallice, T. (1997). Modality-specific operations in semantic dementia. Cortex, 33(4), 593-622.

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? Journal of Memory & Language, 66(1), 79-85.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. Science, 270(5233), 102-105.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. Nature, 379, 649-652

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structures and processes. Current Opinion in Neurobiology, 11, 194-201.

McCarthy, R. A. & Warrington, E. K. (1988). Evidence for modality-specific meaning systems in the brain. Nature. 334, 428-430.

McGuire, S. & Plaut, D. C. (1997). Systematicity and specialization in semantics: a computational account of optic aphasia. Proceedings of the 19th annual conference of the cognitive science society. Hillsdale, NJ: Lawrence Erlbaum Associates.

Paivio, A. (1971). Imagery and verbal processes. New York: Holt, Rinehart & Winston.

Pellegrino, J. W., Rosinski, R., Chiesi, H., & Siegal, A. (1977). Picture-word differences in decision latency: An analysis of single and dual-memory models. Memory & Cognition. 5, 383-396.

Plaut, D. C. (2002). Graded modality-specific specialization in semantics: A computational account of optic aphasia. Cognitive Neuropsychology. 19 (7), 603-639.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. Psychological Review, 103(1), 56-115.

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. Psychological Review, 74(5), 392-409.

Potter, M. C., and Faulconer, B. A. (1975). Time to understand pictures and words. Nature. 253, 437-438.

Pylyshyn, Z. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin. 80, 1-24.

Riddoch, M. J., & Humphreys, G. W. (1987). Visual object processing in optic aphasia: A case of semantic access agnosia. Cognitive Neuropsychology, 4(2), 131-185.

Riddoch, M. J., Humphreys, G. W., Coltheart, M., & Funnell, E. (1988). Semantic systems or system? Neuropsychological evidence re-examined. Cognitive Neuropsychology. 5, 3-25.

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General. 104, 192-233.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. Child Development. 26, 247-253.

Saffran, E. M., Coslett, H. B., & Keener, M. T. (2003). Differences in word associations to pictures and words. Neuropsychologia, 41, 1541-1546.

Sailor, K. M. (1995). The effects of context on comparative judgment and property verification. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL. As cited in Seifert (1997).

Seidenberg, M. S., & McClelland, J. L. (1989). Visual word recognition and pronunciation: A computational model of acquisition, skilled performance, and dyslexia. In Galaburday, A. L. (Ed.) From reading to Neurons, 255-303. Cambridge, MA: The MIT Press.

Seifert, L. (1997). Activating Representations in Permanent Memory: Different Benefits for Pictures and Words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 23, 1106-1121

Shallice, T. (1987). Impairments of semantic processing: Multiple dissociations. In M. Coltheart, R. Job, & G. Sartori (Eds.) The cognitive neuropsychology of language. London: Lawrence Earlbaum Associates Ltd.

Shallice, T. (1988a). From neuropsychology to mental structure. New York, NY: Cambridge University Press.

Shallice, T. (1988b). Specialisation within the semantic system. Cognitive Neuropsychology. 5, 133-142.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. Journal of Verbal Learning and Verbal Behavior. 6, 156-163.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 81(3), 214-241.

Smith, M.C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. Journal of Experimental Psychology: General. 109, 373-392

Springer, S. P., and Deutsch, G. (1993). Left Brain, Right Brain. New York: W.H. Freeman.

Stenberg, G., Radeborg, K., and Hedman, L. R. (1995). The picture superiority effect in a cross-modality recognition task. Memory & Cognition. 23, 425-441.

Theios, J., and Amrhein, P.C. (1989). Theoretical Analysis of the Cognitive Processing of Lexical and Pictorial Stimuli: Reading, Naming, and Visual and Conceptual Comparisons. Psychological Review. 96, 5-24.

Thompson-Schill, S. L., Aguirre, G. K., D'Esposito, M., & Farah, M. J. (1999). A neural basis for category and modality specificity of semantic knowledge. Neuropsychologia. 37, 671-676.

Tranel, D., Kemmerer, D., Adolphs, R., Damasio, H., & Damasio, A. R. (2003). Neural correlates of conceptual knowledge for actions. Cognitive Neuropsychology, 20(3-6), 409-432.

Tyler, L. K., Bright, P., Dick, E., Tavares, P., Pilgrim, L., Fletcher, P., Greer, M., & Moss, H. (2003). Do semantic categories activate distinct cortical regions? Evidence for a distributed neural semantic system. Cognitive Neuropsychology, 20(3-6), 541-559.

Uyeda, K. M., & Mandler, G. (1981). Prototypicality norms for 28 semantic categories. Behavior Research Methods & Instrumentation, 12(6), 587-595.

Warrington, E. K., & McCarthy, R. A. (1983). Category specific access dysphasia. Brain, 106, 859-878.

Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge: further fractionation and an attempted integration. Brain, 110, 829-854

Warrington, E. K., & Shallice, T. (1979). Semantic access dyslexia. Brain. 102, 43-63.

Warrington, E.K., & Shallice, T. (1984). Category-specific semantic impairments. Brain, 107, 829-853.

# APPENDIX A

## Section A.1: Stimuli for Experiment 1

Stimuli for Experiment 1 were selected by collecting survey data in which participants were asked to name the first 5 members of the given categories that came to mind. Items named on at least one-third of the surveys and for which pictures could be found were included as stimuli. Where possible, those items were matched with their rated category membership in Battig & Montague (1969) and Uyeda & Mandler (1980). The first two columns in the table below contain the rank ordering of the given item in those two lists. As can be seen, the vast majority of items for which there were ratings came from the most typical members of their respective categories.

The visual complexity and visual familiarity ratings were gathered by presenting the pictures of the items to participants and asking them to provide ratings of complexity and familiarity on a 10-point scale. The last two columns provide the average ratings for each item.

| Category | Item | Battig & Montague (1969) | Uyeda & Mandler (1980) | Visual complexity | Visual familiarity |
|---|---|---|---|---|---|
| bird | bluejay | 4 | 2 | 5.29 | 6.50 |
| | buzzard | 33 | | 5.54 | 6.69 |
| | canary | 8 | 10 | 4.60 | 7.29 |
| | cardinal | 3 | 15 | 5.71 | 8.10 |
| | condor | | | 6.53 | 5.40 |
| | crow | 6 | 9 | 5.78 | 6.67 |
| | eagle | 5 | 8 | 5.13 | 7.94 |
| | finch | 31 | | 6.18 | 7.47 |
| | goose | 44 | | 6.39 | 7.56 |
| | hawk | 10 | 18 | 6.50 | 7.27 |
| | oriole | 13 | 24 | 5.94 | 6.22 |
| | parakeet | 9 | 11 | 5.20 | 6.80 |

| | | | | | |
|---|---|---|---|---|---|
| | parrot | 14 | 12 | 6.71 | 5.65 |
| | partridge | | | 6.82 | 5.00 |
| | pigeon | 15 | 5 | 5.33 | 8.42 |
| | robin | 1 | 3 | 6.00 | 7.56 |
| | sparrow | 2 | 4 | 5.59 | 7.06 |
| | wren | 12 | 17 | 5.58 | 6.25 |
| cat | angora | | | 5.43 | 7.93 |
| | bobcat | | | 6.78 | 6.30 |
| | jaguar | | | 6.60 | 7.50 |
| | lynx | | | 6.92 | 3.92 |
| | manx | | | 5.31 | 7.46 |
| | puma | | | 6.45 | 7.27 |
| | siamese | | | 5.41 | 7.29 |
| | wildcat | | | 6.45 | 5.50 |
| dog | akita | | | 6.60 | 8.20 |
| | beagle | | | 4.95 | 7.47 |
| | boxer | | | 5.06 | 7.00 |
| | chihuahua | | | 6.08 | 8.08 |
| | collie | | | 5.82 | 8.18 |
| | dachshund | | | 5.50 | 6.95 |
| | dalmation | | | 5.20 | 7.27 |
| | doberman | | | 4.06 | 7.59 |
| | greyhound | | | 5.06 | 7.11 |
| | husky | | | 5.53 | 7.88 |
| | kelpie | | | 5.46 | 6.00 |
| | labrador | | | 4.95 | 7.47 |
| | mastiff | | | 4.53 | 6.94 |
| | pitbull | | | 5.82 | 6.65 |
| | pointer | | | 5.50 | 6.25 |
| | pomeranian | | | 5.94 | 6.00 |
| | poodle | | | 6.47 | 7.89 |
| | rottweiler | | | 4.75 | 7.94 |
| | saluki | | | 5.88 | 6.19 |
| | samoyed | | | 5.63 | 7.00 |
| | schnauzer | | | 5.64 | 7.14 |
| | sheepdog | | | 6.56 | 8.00 |
| | spaniel | | | 5.89 | 6.95 |
| | terrier | | | 5.95 | 7.00 |
| fish | bass | 2 | | 6.06 | 6.47 |
| | carp | 16 | | 5.81 | 5.44 |
| | catfish | 5 | | 5.17 | 5.22 |

| | | | | | |
|---|---|---|---|---|---|
| | chub | | | 5.06 | 6.63 |
| | crappie | 33 | | 5.00 | 6.82 |
| | goldfish | 9 | | 5.10 | 7.80 |
| | grouper | | | 6.64 | 4.79 |
| | guppy | 19 | | 7.17 | 6.83 |
| | haddock | 31 | | 5.76 | 4.65 |
| | herring | 4 | | 5.50 | 7.00 |
| | mackerel | 25 | | 6.22 | 6.00 |
| | perch | 6 | | 6.87 | 6.33 |
| | sailfish | 26 | | 5.40 | 6.47 |
| | salmon | 7 | | 4.67 | 6.87 |
| | sardine | 38 | | 5.36 | 5.69 |
| | shark | 3 | | 5.21 | 6.86 |
| | smelt | | | 4.71 | 6.29 |
| | snapper | 44 | | 5.15 | 5.00 |
| | trout | 1 | | 4.67 | 7.39 |
| | tuna | 8 | | 5.08 | 5.69 |
| | walleye | 42 | | 5.69 | 5.88 |
| flower | carnation | 3 | | 5.14 | 6.93 |
| | daisy | 4 | | 4.29 | 6.43 |
| | goldenrod | | | 7.00 | 5.53 |
| | hyacinth | 33 | | 5.58 | 6.67 |
| | hydrangea | | | 4.77 | 5.15 |
| | lilac | 15 | | 6.11 | 6.00 |
| | lily | 8 | | 5.60 | 5.47 |
| | orchid | 6 | | 6.40 | 6.73 |
| | pansy | 9 | | 6.18 | 6.59 |
| | rose | 1 | | 6.30 | 8.00 |
| | snapdragon | | | 6.10 | 4.80 |
| | tulip | 2 | | 5.09 | 7.82 |
| | violet | 5 | | 5.52 | 6.57 |
| fruit | apple | 1 | 2 | 2.67 | 9.33 |
| | apricot | | 13 | 4.90 | 5.90 |
| | banana | 4 | 4 | 3.43 | 9.07 |
| | blueberry | 22 | 16 | 5.75 | 6.00 |
| | cherry | 7 | 7 | 2.69 | 8.06 |
| | cranberry | | | 3.53 | 7.00 |
| | grapefruit | 9 | 9 | 4.53 | 7.40 |
| | grapes | 6 | 8 | 4.71 | 8.21 |
| | lemon | 10 | 14 | 2.84 | 7.74 |
| | orange | 2 | 1 | 4.64 | 7.45 |

| | | | | | |
|---|---|---|---|---|---|
| | peach | 5 | 5 | 3.50 | 7.58 |
| | pear | 3 | 3 | 2.50 | 9.36 |
| | strawberry | 16 | 6 | 3.53 | 9.33 |
| | watermelon | 17 | 21 | 4.79 | 7.07 |
| horse | appaloosa | | | 6.17 | 6.25 |
| | arabian | | | 5.23 | 6.08 |
| | bay | | | 5.38 | 6.00 |
| | clydesdale | | | 6.46 | 6.69 |
| | mustang | | | 6.81 | 7.00 |
| | palomino | | | 4.73 | 7.91 |
| | shire | | | 4.44 | 6.89 |
| | suffolk | | | 5.00 | 6.18 |
| | thoroughbred | | | 5.69 | 7.62 |
| tree | balsa | | | 5.29 | 7.47 |
| | banyan | | | 8.24 | 6.35 |
| | birch | 6 | 6 | 6.91 | 8.18 |
| | cedar | 22 | 10 | 6.25 | 7.19 |
| | chestnut | 21 | 24 | 6.64 | 9.00 |
| | dogwood | 8 | 27 | 6.53 | 8.13 |
| | elm | 4 | 8 | 6.00 | 7.70 |
| | fir | 14 | 4 | 5.72 | 8.60 |
| | magnolia | 27 | 28 | 6.79 | 8.26 |
| | mahogany | 40 | | 6.73 | 8.26 |
| | maple | 2 | 7 | 6.27 | 8.14 |
| | oak | 1 | 1 | 6.95 | 8.43 |
| | pine | 3 | 2 | 5.78 | 7.56 |
| | poplar | 19 | 26 | 5.12 | 7.35 |
| | redwood | 10 | 3 | 5.73 | 8.14 |
| | spruce | 9 | 11 | 5.43 | 8.33 |
| | sycamore | 17 | 14 | 6.33 | 7.50 |
| | tamarack | | | 7.14 | 8.86 |
| | willow | 12 | 20 | 6.15 | 7.77 |
| | yew | | | 7.00 | 8.60 |
| vegetable | asparagus | 9 | 10 | 4.25 | 8.13 |
| | broccoli | 10 | 6 | 5.24 | 7.94 |
| | cabbage | 12 | 15 | 6.21 | 4.43 |
| | carrot | 1 | 1 | 3.20 | 8.70 |
| | cauliflower | 14 | 18 | 4.36 | 8.09 |
| | celery | 11 | 9 | 5.71 | 8.00 |
| | corn | 3 | 7 | 5.31 | 8.38 |
| | greenbeans | 23 | 2 | 2.94 | 7.88 |

| | | | | |
|---|---|---|---|---|
| lettuce | 7 | 8 | 6.15 | 7.54 |
| peas | 2 | 5 | 3.80 | 8.05 |
| potato | 5 | 17 | 2.14 | 8.64 |

## Section A.2: Stimuli for Experiment 3

The selection criteria for Experiment 3 are outlined in Chapter 4 above. The visual complexity and familiarity ratings were gathered in the same manner as those gathered in Experiment 1, as described in section A.1 above.

| Category | Item | Battig & Montague (1969) | Uyeda & Mandler (1980) | Visual complexity | Visual familiarity |
|---|---|---|---|---|---|
| cleaning | broom | | | 3.68 | 7.96 |
| | mop | | | 3.48 | 7.52 |
| | sponge | | | 2.69 | 8.33 |
| | vacuum | | | 5.94 | 7.97 |
| communication | radio | | | 6.19 | 6.57 |
| | telephone | | | 4.52 | 8.12 |
| | television | | | 3.96 | 8.81 |
| | walkietalkie | | | 5.85 | 7.31 |
| | book | | | 6.73 | 7.19 |
| cooking | bowl | 9 | 18 | 3.47 | 7.27 |
| | toaster | 22 | 23 | 5.22 | 7.35 |
| | fryingpan | 15 | 9 | 2.50 | 9.06 |
| | pot | 56 | 6 | 3.68 | 7.50 |
| | grill | | | 5.17 | 7.16 |
| | microwave | | | 6.07 | 8.33 |
| | oven | 26 | 27 | 5.12 | 7.36 |
| | stove | | 26 | 6.14 | 6.46 |
| cutting | knife | | | 2.68 | 8.81 |
| | saw | | | 2.58 | 7.87 |
| | scissors | | | 3.00 | 9.04 |
| firstaid | bandaid | | | 2.17 | 8.26 |
| | crutch | | | 2.77 | 6.90 |
| | gauze | | | 4.15 | 6.54 |
| hygiene | deodorant | | | 3.07 | 7.76 |
| | razor | | | 3.38 | 6.96 |
| | toothbrush | | | 4.33 | 7.87 |

| | | | | | |
|---|---|---|---|---|---|
| | brush | | | 3.75 | 7.29 |
| lighting | candle | | | 3.54 | 7.86 |
| | flashlight | | | 3.89 | 7.78 |
| | lamp | | | 4.50 | 7.14 |
| | lightbulb | | | 2.81 | 9.22 |
| | spotlight | | | 4.28 | 6.22 |
| shelter | house | 1 | 1 | 7.00 | 7.96 |
| | igloo | 14 | 14 | 4.96 | 4.96 |
| | shack | 10 | 23 | 7.32 | 5.12 |
| | tent | 3 | 24 | 4.10 | 7.25 |
| | cave | 4 | 29 | 5.80 | 5.60 |
| sitting | beanbag | | | 2.88 | 7.44 |
| | bench | 27 | 19 | 2.61 | 6.76 |
| | chair | 1 | 1 | 2.69 | 8.72 |
| | couch | 7 | 3 | 3.85 | 8.39 |
| | recliner | | | 4.62 | 7.96 |
| | stool | 10 | 18 | 2.70 | 7.26 |
| storage | backpack | | | 3.60 | 8.36 |
| | box | | | 2.33 | 7.93 |
| | briefcase | | | 4.21 | 7.52 |
| | crate | | | 3.94 | 7.33 |
| | duffelbag | | | 3.97 | 7.70 |
| | purse | | | 3.48 | 6.04 |
| | shelf | | | 3.87 | 7.50 |
| | suitcase | | | 4.32 | 7.57 |
| | closet | | | 4.27 | 7.00 |
| | bag | | | 3.93 | 7.94 |
| transportation | airplane | 3 | 14 | 6.15 | 7.26 |
| | bicycle | 6 | 11 | 6.21 | 8.45 |
| | boat | 8 | 17 | 6.74 | 5.76 |
| | bus | 2 | 4 | 5.62 | 7.88 |
| | car | 1 | 1 | 5.89 | 8.04 |
| | train | 4 | 12 | 6.77 | 6.60 |
| | truck | 5 | 3 | 5.61 | 7.48 |
| | van | 39 | | 5.88 | 6.97 |
| | wagon | 10 | 24 | 4.68 | 6.79 |
| | scooter | 9 | 20 | 5.05 | 5.41 |