STIMULUS WITH A PAST: MEMORY TASK PERFORMANCE AFFECTED BY FREQUENCY AND PROBABILITY OF INTENTIONAL ACTS INVOLVING THE STIMULUS

By

Maggie Jinghua Xiong

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in
Psychology

May, 2005

Approved:	Date:
Gordon D. Logan , Chairperson	March 15, 2005
Timothy P. McNamara	March 15, 2005
David C. Noelle	March 15, 2005
Thomas J. Pameri	March 15, 2005

ACKNOWLEDGEMENT

This research project was successfully completed thanks to many persons who helped me at various stages. I thank Dr. Gordon Logan who, as my advisor, always provided timely and insightful comments, guiding me through the steps to completion. I thank my other outstanding committee members, Dr. Tim McNamara, Dr. David Noelle, and Dr. Tom Palmeri, who provided advice and support when I was in need. I thank Dr. Jeff Franks who, above all, recommended Google for word co-occurrence counts. I thank Sean Jewett, who wrote the Google Web APIs automated search program that has been instrumental for the project. I thank members of the Google Web APIs team, who have very generously provided the project essentially unlimited access to automated Google searches. The success of this project is directly attributable to the patience and support these persons have demonstrated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Intentional History of Words

The present work aims to establish intention as an important factor in learning and memory by demonstrating the cumulative impact of specific intentional processes on memory for word stimuli. Intention, as “an aim that guides action; an objective” (The American Heritage® Dictionary of the English Language, 2000), can be internally generated, as in chocolate cravings in the middle of writing. Intention can also be externally evoked by the context, as in, there is an ice-cream truck outside, so I’m going to get ice-cream. Or, I want to go get ice-cream—intention does not require overt action. In psychology laboratories, intention is instilled into subjects through task instructions. It has been shown that the type of task that subjects perform on stimuli have specific impact on their memory about the stimuli. According to the transfer-appropriate processing principle of memory (Bransford, Franks, Morris, & Stein, 1979; Morris, Bransford, & Franks, 1977), the specific processing of specific stimuli facilitates later processing of the same stimuli, to the extent that the earlier and later events have overlapping processes. Overt task processing is only one type of intentional situation. Intention in other situations, such as evoked by the context, can also cause memory for the target to be specific to the intention. The result of such context-evoked intentional process is described in the encoding specificity principle of memory (Tulving, 1979; Tulving & Thomson, 1973), which suggests that the most efficient memory retrieval cues are cues that match the specific encoding operations performed on the input stimuli. The present work integrates the transfer-appropriate processing and encoding specificity principles of memory into the intention-specific learning principle, with the emphasis on the intentional act involving a stimulus (intention as the guide for action) and the interpretation of the stimulus as the result of the intentional act. Applying the intention-specific learning principle to the instance theory of memory (Logan, 1988), the present work arrived at novel predictions about stimulus frequency effects in memory tasks.

To achieve a goal, one must attend to the situation. Intention as goal-orientedness requires attention. Logan and Etherton (1994) suggested that attention is a sufficient, and maybe even necessary, condition for binding co-occurring elements into an instance. An instance can be considered the integral experience of intention, its target, and the context (elements that co-occurs with the target). In accordance with the instance theory of memory (Logan, 1988), the present work assumes that 1) each encounter with a stimulus is encoded and stored as a separate instance; 2) the encoding of an instance into memory is a necessary consequence of attention. To the extent that an event is fully attended, remembering is obligatory, even if the remembering may not be perfect (Craik & Lockhart, 1972; Moscovitch, 1994); 3) the retrieval of instances is a necessary consequence of attention. Attending to a stimulus causes obligatory retrieval of the instances associated with the stimulus, even if the retrieval may fail.

Applying the intention-specific learning principle to the instance theory of memory (Logan, 1988), the present work proposes that 1) the obligatorily encoded instances are specific to the intention during encoding; 2) the obligatorily retrieved instances are specific to the intention at retrieval. Imagine intention as the string that strings together instances. Instances are encoded onto a particular string, and instances are retrieved along a particular string.

The instances strung together by intentions are hereby referred to as the "intentional history" of a stimulus. Each stimulus has a past, an intentional history that registers the past engagement of the stimulus in various intentional acts, and that extends far beyond what happens in psychology laboratories (unless they are novel stimuli created in the lab, of course, like "Greebles"; Gauthier & Tarr, 1997). Following the assumptions of obligatory encoding and obligatory retrieval of instances, the present work suggests that the intentional history of a stimulus is inherent to the stimulus representation and will constrain current intentional processing of the stimulus.

A straightforward way to summarize the intentional history of a stimulus with respect to an intentional act is by the frequency with which the stimulus has been engaged in the act. And related to, but also dissociable from the intention-specific frequency, is the probability with which the stimulus has been involved in the intentional act (The relation between intention-specific frequency and intention-specific probability will be described in detail in later sections). Researchers have considered the frequency of a stimulus to be part of the stimulus representation, for example, word frequency is assumed to be part of a word's semantic representation (Forster & Davis, 1985; MacLeod & Kampe, 1996). The intentional history approach, however, suggests that the frequency information in a stimulus representation is not simply the frequency of the stimulus independent of the intentional acts at acquisition, but frequency and probability information specific to the intentional acts.

Word frequency has been a potent variable affecting various kinds of memory task performances (e.g., Gregg, 1976; Jacoby & Dallas, 1981; MacLeod & Kampe, 1996; Scarborough, Cortese, & Scarborough, 1977). However, the intention-specific principle implies that word frequency and intention-specific frequency or probability have likely been confounded in previous studies. Word frequency effects may in fact have been intention-specific frequency or probability effects. According to the intention-specific learning principle, the number of times a subject has encountered a stimulus while engaging in intentional act A will affect how he performs task A on the stimulus. However, the number of times that the subject has encountered the stimulus in intentional act A should not necessarily affect how he can perform task B on the stimulus, unless there is sufficient overlap between tasks A and B. That is, only intention-specific frequency or probability should affect current intentional processing on a stimulus, and not stimulus frequency independent of acquisition intentions. The intention-specific learning principle predicts intention-specific frequency or probability effects in memory tasks, but not general stimulus frequency effects. This is the primary hypothesis of the present work. This hypothesis is tested in the context of semantic judgments on the meanings of words, with both explicit and implicit memory tasks.

Explicit and Implicit Memory Tasks

The present work tested the intentional history hypothesis with both explicit and implicit memory tasks to show that intentional history may be a general property of memory representations for stimuli. Everything we know about memory is probed with a certain task (Roediger, Weldon, & Challis, 1989). A major distinction between the types of memory tasks is the contrast between explicit and implicit memory tasks. Typical explicit memory tasks include word recognition and recall (Gillund & Shiffrin, 1984; Gregg, 1976; Mandler, 1980). In recognition, subjects try to distinguish between words presented in the study episodes from non-studied words. In recall, subjects try to generate words that have appeared in the study episodes. Recognition and recall are referred to as episodic memory tasks, or explicit memory tasks, because in these tasks subjects explicitly try to remember particular episodes. Typical implicit memory tasks include lexical decision, word identification, and word stem completion tasks (Jacoby & Dallas, 1981; Scarborough et al., 1977; Tulving, Schacter, & Stark, 1982; for reviews, see Roediger & McDermott, 1993; Schacter, 1987). In these tasks, subjects can make word-nonword decisions faster, identify words with higher accuracy, or complete more word stems after having been exposed to the target words. The facilitation for previously presented targets words is termed repetition priming. According to the episodic or instance view of repetition priming, words are identified through representations of prior episodes or instances involving the words (see Bowers, 2000, and Tenpenny, 1995 for reviews on the abstractionist versus episodic view of repetition priming). Repetition priming arises because recent episodes are in general more accessible than older episodes (Jacoby, 1983; Jacoby, Baker, & Brooks, 1989). Or, as suggested by the instance theory (Logan, 1988, 1990), adding instances will speedup the race between instances and lead to a faster response. Repetition priming occurs irrespective of subjective awareness or control. Memory tasks involving the repetition priming phenomenon have been referred to as implicit memory tasks, in contrast to more traditional memory tasks, such as recognition and recall, in which subjects explicitly try to remember (Graf & Schacter, 1985; Schacter & Graf, 1986).

Because explicit and implicit memory task performances are differentially affected by variables such as hippocampal lesion (Cohen & Squire, 1980; Graf, Squire, & Mandler. 1984; Graf & Schacter, 1985; Shimamura & Squire, 1984, 1989; Warrington & Weiskrantz, 1978) and levels-of-processing manipulations (Craik, Moscovitch, & McDowd, 1994; Graf, Mandler, & Haden, 1982; Jacoby & Dallas, 1981), researchers have discussed whether explicit and implicit memory tasks have tapped into different memory systems. Some of the postulated system schemes include the ternary classification of episodic, semantic, and procedural systems (Tulving, 1985), the declarative versus procedural (or non-declarative) systems (Squire, 1986, 1992), the locale versus taxon systems (Nadel, 1992), and the hippocampal versus cortical learning systems (McClelland, McNaughton, & O'Reilly, 1995). Explicit memory tasks are attributed to the episodic, declarative, locale, or hippocampal system, which requires intact hippocampal function. Implicit memory tasks are ascribed to the semantic, procedural, taxon, or cortical system, which can apparently function independent of hippocampal capabilities.

Advocates of processing theories (Blaxton, 1989; Jacoby, 1983; Roediger, 1990; Roediger et al., 1989) have argued that many dissociations between explicit and implicit memory tasks can be explained in terms of the overlap between acquisition and test tasks, and that transfer-appropriate processing (Bransford et al., 1979; Morris et al., 1977) is a more parsimonious account regarding memory task performance. For example, the levels-of-processing manipulation can vary semantic versus rhyme processing between acquisition conditions. Semantic processing affects recognition and recall performance because these tasks are conceptually-driven. Semantic processing has little impact on word identification because identification is a perceptually-driven task.

Whether the dissociation between explicit and implicit memory tasks should be explained by distinct memory systems or by the type of processing involved (e.g. conceptual versus perceptual), it is important to examine the effects of intentional history with both explicit and implicit memory tasks if we wish to show that intentional history is a general property of memory representations. To the extent that the memory subserving recognition, recall, and repetition priming task performances is episodic, the intentional history of words should affect the explicit, as well as, the implicit memory task performances. But, the way in which intentional history affects explicit and implicit memory task performances may be different. For example, two measures of intentional history, the frequency and the probability of the stimulus being engaged in the acquisition acts, may have differential impact on explicit and implicit memory task performances. Differences in the effects of intentional history on explicit and implicit memory tasks can provide new clues to the mechanisms underlying the distinction between these tasks. It is therefore the secondary goal of the present work to assess the differential impact of intention-specific frequency and intention-specific probability on explicit and implicit memory tasks. In the next sections I shall briefly review word frequency effects in implicit memory repetition priming tasks, then go over word frequency effects in the explicit memory tasks of word recognition and free recall (recall without cues).

Word Frequency Effects in Repetition Priming

Word frequency effects in repetition priming have been consistently reported in lexical decision (Duchek & Neely, 1989; Forster & Davis, 1984; Scarborough et al., 1977) and perceptual identification tasks (Jacoby & Dallas, 1981; Jacoby & Hayman, 1987; Kirsner, Milech, & Stumpfel, 1986). High frequency words are responded to faster as legitimate words, or identified more accurately, than low frequency words. However, compared to high frequency words, low frequency words demonstrate greater amounts of repetition priming from having been presented in the study phase. From the instance theory point of view (Logan, 1988, 1990), one more addition to a large set of instance will have less impact on the rate of instance retrieval compared with cases where there are fewer prior instances. From another perspective, the power law of learning dictates that increments in learning should be smaller near the asymptote than near the beginning of learning.

Word frequency effects in repetition priming are of theoretical interest because they suggest that semantic representations are stored as episodic instances instead of abstract lexicons (Bowers, 2000; Tenpenny, 1995). Word frequency effects in repetition priming are readily explained from the episodic or instance perspective (Logan, 1988; 1990), yet they present a challenge for the abstractionist view of semantic representations. The abstractionist view suggests that words are represented as abstract lexical entries, with high frequency words located at the beginning of the search path and low frequency words towards the end of the search path (Forster & Davis, 1984). To account for the comparatively large amount of repetition priming low frequency words receive from prior presentation, the abstractionist view has to assume the accessibility of lexical entries to be highly labile. This goes against the abstractionist assumption that the differential accessibility of high versus low frequency words reflects fairly stable organizational properties of the lexicon. Consequently, word frequency effects in repetition priming have been considered markers of episodic memory trace retrieval.

However, word frequency effects have not been consistently observed in other implicit memory tasks such as word stem or fragment completion tasks (for a review, see Roediger & McDermott, 1993). While MacLeod (1989) and Roediger, Weldon, Stadler, and Riegler (1992) found more repetition priming for low frequency than high frequency words, Tenpenny and Shoben (1992) reported an opposite pattern, with more repetition priming for high frequency than low frequency words.

The intention-specific learning principle suggests that frequency effects in repetition priming should be explained with respect to the frequency or probability of the intentional acts being performed on the stimulus in past acquisitions, instead of the frequency of the stimulus independent of past acquisition acts. When the task used to measure repetition priming matches the intentional history invoked by the study task, intention-specific frequency or probability effects should be observed. Conversely, if the test task does not overlap with the intentional history invoked by the study task, no frequency or probability effects should be expected. From this perspective, word identification and lexical decision tasks may be special cases in which the test tasks matches well with the intentional acts in everyday word acquisition situations. Because often times reading a word entails intentional identification and lexical access processes, the acquisition act frequency and word frequency are confounded. As a consequence, intention-specific frequency effects in repetition priming would appear as word frequency effects in word identification and lexical decision tasks (Duchek & Neely, 1989; Forster & Davis, 1984; Jacoby & Dallas, 1981; Jacoby & Hayman, 1987; Kirsner et al., 1986). At the same time, it should come as no surprise that when the test task deviated from typical reading processes, as is the case in word stem or fragment completion tasks, this seemingly word frequency effect had only been inconsistently observed (MacLeod, 1989; Roediger et al., 1992; Roediger & McDermott, 1993; Tenpenny & Shoben, 1992).

The failure to consistently observe word frequency effects in some repetition priming tasks does not mean that frequency effects do not occur in tasks other than lexical decision and word identification. If intentional history is an inherent part of the stimulus representation, intention-specific frequency or probability effects should be reliably detected as long as the task that is used to measure repetition priming matches the intentional history invoked by the study task. Intention-specific frequency or probability effects derive directly from the episodic nature of frequency effects. It is also a natural extension of the transfer-appropriate processing principle of memory (Bransford et al., 1979; Morris et al., 1977).

In accordance with the transfer-appropriate processing principle, studies on semantic judgments (e.g. making *beautiful/ugly* judgment on the meanings of words) have shown that learning as indexed by the amount of repetition priming is contingent on the similarity between acquisition and test task processes (Franks, Bilbrey, Lien, & McNamara, 2000; Thompson-Schill & Gabrieli, 1999; Vriezen, Moscovitch, & Bellos, 1995; Xiong, Franks, & Logan, 2003). Xiong et al. determined the degree of similarity between semantic judgments from the semantic differential approach (Osgood, Suci, & Tannenbaum, 1957). According to Osgood et al., the meanings of words are represented along a limited number of semantic dimensions such as the evaluative, potency, and activity dimensions. Bipolar adjective terms such as *beautiful/ugly* represent a straight line in semantic space and can function as a “semantic scale” against which the meanings of words are measured. Xiong et al. based their semantic judgment tasks on semantic scales. Semantic judgment tasks were considered similar, that is, to share overlapping processes, if the corresponding semantic scales should elicit processing on the same semantic dimension(s). Semantic judgments were regarded as dissimilar if the corresponding semantic scales should elicit processing on orthogonal semantic dimensions. Subjects performed a semantic judgment task on a set of words at the study phase. They then performed the same, a similar, or a dissimilar judgment on these set of words as well as a new set of words at test. Repetition priming was defined as the difference in reaction times (RTs) to the old and new sets of words at test. Repetition priming was at a maximum level when the study and test judgments were the same, intermediate when the study and test judgments were similar, and negligible when the study and test judgments were dissimilar from each other.

Vriezen et al. (1995) and Thompson-Schill and Gabrieli (1999) took a different approach to similarity relations in semantic judgments and found similar patterns of results. They postulated that semantic information on objects might be organized according to domain-specific attributes (e.g., structural/visual or functional). Semantic judgments were considered similar if they accessed the same semantic domain. Both studies reported significant repetition priming between study and test judgments that accessed the same semantic domain, and little or no repetition priming when the study and test judgments tapped into different semantic domains.

In the above mentioned studies (Vriezen et al., 1995; Thompson-Schill & Gabrieli, 1999; Xiong et al., 2003), little or no repetition priming at all was detected when the study and test judgment tasks were dissimilar, even though identical stimulus items were involved. The lack of transfer between dissimilar study and test tasks indicates that distinctive task processes involving the same item do not aggregate. Results from these studies suggest that intention is an inherent part of the memory instances involving the stimulus (Logan, 1988). The cumulative effect of such intention-specific instances, then, should also be intention-specific.

Word Frequency Effects in Explicit Memory Tasks

Word frequency effects are also reported in recognition tasks. After having been presented in a study phase, low frequency words are better recognized than high frequency words (e.g. Jacoby & Dallas, 1981; MacLeod & Kampe, 1996). The dual process model of word recognition suggests that people recognize words by familiarity as well as direct retrieval of the target words (e.g. Jacoby, 1991; Mandler, 1980). Familiarity is assumed to be based on perceptual fluency, that is, the ease which people can perceive the target (Jacoby), or the degree of interitem integration, which is the integration of an event independent of its relation to other events (Mandler). Direct retrieval, on the other hand, involves retrieving the contextual details of the original study episode. Word frequency effects in recognition may be explained in terms of familiarity by assuming that people are sensitive to the absolute levels of familiarity as well as the increment in familiarity with recent exposure. The finding that low frequency words show more repetition priming than high frequency words has been taken as evidence for the familiarity based account of word frequency effects in recognition (Jacoby, Mandler). However, presuming that subjects can distinguish between items that they actually “remember” as being in the study list versus items that they “know” to be in the study list based on familiarity, a number of studies concluded that word frequency effects in recognition arise from the direct retrieval process, because word frequencies only affected the “remembered” items (Gardiner & Java, 1990; Guttentag & Carroll, 1997; Kinoshita, 1995).

With respect to word frequency effects in free recall tasks, two different patterns have emerged depending on the experimental design. When high and low frequency words are presented in separate lists during study, often with a between subject design, high frequency word are more likely to be recalled than low frequency words (see Gregg, 1976, for a review). When high and low frequency words are mixed in the same study list, the recall advantage for high frequency words disappears, or sometimes low frequency words are more readily recalled than high frequency words (Duncan, 1974; Gregg, Montgomery, & Castaño, 1980; MacLeod & Kampe, 1996). The retrieve-recognize model of recall (Kintsch, 1970, as cited by Gregg, 1976) suggests that recall starts with a retrieval process in which potential responses are generated, followed by a recognition process whereby subjects decide which of the potential responses are appropriate. The retrieve-recognize model suggests that high frequency words are better recalled than low frequency words because they are more likely to be generated than low frequency words. However, the retrieve-recognize model runs into difficulty when attempting to explain the absence of advantage for high frequency words in mixed list conditions.

Regardless of what mechanisms give rise to word frequency effects in recognition and recall tasks, what is most relevant for the present work is the consensus that word frequency effects in recognition and recall reflect episodic memory retrieval effects. Word recognition and free recall tasks rely on subjects' remembering the study episodes. These tasks should therefore be the exemplar cases for demonstrating intention specificity. In fact, task-specificity, or, intention-specificity, in explicit memory tasks has long since been documented and led to the proposal of the transfer-appropriate processing principle (Bransford et al., 1979; Morris et al., 1977), as well as the encoding-specificity principle of memory (Tulving, 1979; Tulving & Thomson, 1973). Morris et al. engaged subjects in either a semantic acquisition or a rhyme acquisition task in the study phase. Semantic acquisition led to superior performance than rhyme acquisition in a standard recognition test. However, rhyme acquisition led to better performance than semantic acquisition in a rhyming recognition test. Tulving and Thomson presented subjects target words to study in the context of weakly associated words (e.g. *ground*-COLD). Subjects recalled many more words at test when they were given the original weak associates as retrieval cues than when they were provided strong associates of the target words (e.g. *hot*-COLD) that were not presented in the study phase as retrieval cues. Given these results, intention-specific frequency effects in recognition and recall tasks should be expected.

Assessing the Intentional History of Words

The above discussion shows that both explicit and implicit memory tasks exhibit intention-specific learning effects. The type of task subjects perform on a word at the study phase has direct impact on their memory for the stimulus as assessed by repetition priming (Franks et al., 2000; Thompson-Schill & Gabrieli, 1999; Vriezen et al., 1995; Xiong et al., 2003), recognition (Bransford et al., 1979; Morris et al., 1977), and recall tasks (Tulving, 1979; Tulving & Thomson, 1973). This suggests that intention is an inherent part of the memory instances involving the stimulus (Logan, 1988). The cumulative effect of such intention-specific instances should also be intention-specific.

To examine the cumulative effect of intention-specific instances on memory for word stimuli, that is, the effect of the intentional history of words on memory for word stimuli, the present work invoked particular aspects of a word's intentional history by a semantic judgment task at study, and assessed subjects' memory for words in the study phase with repetition priming, recognition, or free recall tasks at test. The intentional history of words with respect to the study judgment task was determined, and its effects compared with those of word frequencies.

The present work estimated the intentional history of words with respect to semantic judgments by the text context the word stimuli have appeared in. For example, the frequency with which the word *dream* has appeared in the context of *beautiful* or *ugly* is presumably proportional to how often *dream* has been judged in terms of *beautiful/ugly*. This leads to two measures of the intentional history of words—the frequency and the probability with which a judgment has been made on words. These two measures are dissociable because there are situations in which the absolute frequency of a judgment being performed on a word is relatively high, but the probability of making the judgment is relatively low (e.g., making *beautiful/ugly* judgment on the word *desk*), and situations where the absolute frequency of performing the judgment is low, but the probability of making the judgment is high (e.g., making *beautiful/ugly* judgment on the word *tiara*).

Subject ratings

As a first attempt at quantifying the intentional history of words, I had subjects rate how often they had thought of a word in terms of a semantic scale, and then re-analyzed repetition priming data in Xiong et al. (2003) for task-specific frequency effects. There was significantly more repetition priming for words that received low task-specific frequency ratings than for words that had high task-specific frequency ratings. Meanwhile no difference in repetition priming was observed if the words were divided into subsets with high versus low word frequency (Kučera & Francis, 1967), or subsets with high versus low word familiarity (Gilhooly & Logie, 1981). Details of these analyses are reported in Chapter 2.

Googling

Instead of having subjects rate the frequency and probability of a judgment being performed on words, the present set of experiments relied on the co-occurrences of words in a text corpus for more objective measures of the frequency and probability with which a judgment has been performed on words. Objective measures based on co-occurrences of words are favored for two reasons: 1) it is not clear how well subjects can discriminate and give independent frequency and probability ratings; 2) ultimately, we would prefer to relate human behavior to environmental variables rather than subjective variables.

The number of co-occurrences of a semantic scale and a stimulus item in a proper text corpus presumably reflects the frequency with which the judgment has been made in the context of that item. Co-occurrences of words in text corpora have been used to develop computational models of language and memory (Landauer & Dumais, 1997; Lund & Burgess, 1996). These models were able to account for large amounts of data on language learning and memory despite their seemingly very gross measuring methods for word co-occurrences. For instance, the HAL (hyperspace analogue to language; Lund & Burgess) model counted co-occurrences of words in a 10-word sliding window, irrespective of any sentential structure. Such co-occurrence counts still had sufficient signal-to-noise ratio to give rise to a highly functional model of semantic memory.

The Internet is the largest, most general, and most up-to-date text corpus. The Google search engine provides a ready way to gather co-occurrence as well as stand-alone frequency counts of words on the Internet. When a semantic scale term and a word are entered for Google to search, for example, *beautiful dream*, the estimated number of results returned by Google represents the number of text samples in which the word *beautiful* and *dream* have appeared together, which is, in turn, presumed to correlate with the frequency of the *beautiful* judgment being made on the word *dream*.

For the present set of experiments, the frequency with which a semantic judgment has been performed on a word, for instance, the frequency with which a *beautiful/ugly* judgment has been made on the word *dream*, is calculated by adding up the estimated number of search results returned by Google for *beautiful dream* and *ugly dream*. For convenience's sake, this task-specific frequency value is labeled $F_{scale.word}$, capturing the fact that task-specific frequency is defined by the frequency of the semantic scale (which represents the semantic judgment task) and the word appearing together. In line with this denotation, the frequency of the word itself is labeled F_{word}. The probability of making a *beautiful/ugly* judgment given the word *dream* is determined by dividing the frequency of *beautiful/ugly* and *dream* appearing together by the frequency of the word *dream* itself. This task-specific probability value is labeled $P_{scale}|word$, $P_{scale}|word = F_{scale.word}/F_{word}$. In the case of a *beautiful/ugly* judgment on *dream*, the search for *dream* led to an estimate of 25,700,000 results, the search for *beautiful dream* produced an estimate of 4,970,000 results, and the search for *ugly dream* generated an estimate of 816,000 results (retrieved March 18, 2004). For the word *dream*, its word frequency F_{word} = 25,700,000, task-specific frequency $F_{scale.word}$ = 4,970,000 + 816,000 = 5,786,000, and task-specific probability $P_{scale}|word$ = 5,786,000 / 25,700,000 = 0.23.

I have no specific knowledge about what counts as one search result for Google, except the claim that Google considers the "the position and size of the search terms within the page, and the proximity of the search terms to one another" (Blachman, retrieved April 16, 2004). But whatever Google's criteria are, they are consistent. Results from the present set of experiments have proven the Google approach to the intentional history of words to be extremely functional and efficient.

Through the course of the current project, two sets of words were used for generating the word stimuli. The first set was the 1944 words from the Gilhooly and Logie (1980) norm of rated word familiarity. The second set included 2482 words from Kučera and Francis (1967), mostly nouns with some verbs and adjectives, evenly distributed across the alphabet. The second set was generated with the help of MRC Psycholinguistic Database website (retrieved August 28, 2004). With the exception of Experiment 1, an automated search program based on the Google Web API (retrieved May 29, 2004) was used to perform the Google searches. The correlation between Google word frequencies retrieved three months apart (retrieved May and August, 2004) was 0.98, $p < 0.001$. The correlation between Google word frequencies and Gilhooly and Logie familiarity ratings was 0.38, $p < 0.001$. The correlation between Google word frequencies and Kučera and Francis word frequencies was 0.63, $p < 0.001$. As a reference point, the correlation between Gilhooly and Logie familiarity ratings and Kučera and Francis word frequencies was 0.47, $p < 0.001$.

Overview of the Experiments

To address whether intention-specific frequency and probability information is inherent to a word's representation and has specific influence on memory for the words, six experiments were conducted in the presented work. They compared the effects of the intentional history of words, operationalized as the $F_{scale.word}$ and $P_{scale}|word$ values, with the effects of word frequencies, F_{word}. The experiments also compared the effects of the two different measures of the intentional history of words with each other. To test whether intentional history is a general property of memory representations, both explicit and implicit memory tasks were used when assessing the memory for the words.

As suggested earlier, intention-specific frequency or probability effects are expected only if the task with which memory for the stimulus is measured matches the intentional history invoked by the study task. For repetition priming, use of the same task at test as used in the study phase should result in maximum effects. For recognition and recall, a cued procedure could be employed, in which subjects are instructed to use the task they have just performed in the study phase as a retrieval cue. However, I speculate that people naturally use intentional acts as retrieval cues for remembering things (how many times have you asked yourself "who have I told this joke to"?). Subjects will most likely use the study task as a retrieval cue for the recognition and recall tasks even without explicit cueing. The present work chose not to use explicit cueing but instead gave subjects standard recognition and free recall instructions (e.g., respond *old* to words that have been presented in the previous block). This might cause the intention-specific frequency or probability effects to be weaker because subjects might not always think about the task they have just done, so the results would be a mixture of cued and uncued performance. But, with standard recognition and free recall instructions, the experiments can afford more generalizability, which is important if we wish to generalize the results to word frequency effects in previous studies.

In all experiments to be reported here, the study phases involved making semantic judgments on the meanings of words. The study task should invoke a particular aspect of the intentional history of words with respect to the judgment task. The test phases of Experiment 1 and 2 repeated the same semantic judgment task on words as in the study phase. Repetition priming for items with high and low F_{word}, $F_{scale.word}$ and $P_{scale}|word$ values with respect to the study judgment task was compared. Experiment 3 and 4 measured subjects' sensitivity in distinguishing between old and new items in recognition tests. Recognition sensitivities for items with high and low F_{word}, $F_{scale.word}$ and $P_{scale}|word$ values were compared. Finally, Experiment 5 and 6 focused on the effects of $F_{scale.word}$ and $P_{scale}|word$ in free recall tests, with mixed and separate list designs, respectively.

The primary goal of the present work is to establish intentional history as part of a word's representation instead of simple word frequency. This entails re-attributing the effects of word frequencies with regard to memory task performances to intentional history. Therefore, one, or both measures of the intentional history of words, $F_{scale.word}$ and $P_{scale}|word$, should reproduce the traditional word frequency effect pattern in memory tasks. For all the experiments, the intentional history hypothesis predicts intention-specific frequency or probability effects analogous to traditional word frequency effects, but not word frequency effects themselves. Results like these would suggest intentional history to be part of the stimulus representation, but not stimulus frequency information independent of acquisition intentions.

It is also likely that two measure of the intentional history of words, $F_{scale.word}$ and $P_{scale}|word$, should have differential impact on explicit and implicit memory tasks. If this is the case, it will indicate that explicit and implicit memory processes integrate information differently, thus supporting the proposition of multiple memory systems. On the other hand, if $F_{scale.word}$ and $P_{scale}|word$ affects explicit and implicit memory tasks similarly, it should raise questions about the necessity of postulating multiple memory systems.

CHAPTER II

INTENTION-SPECIFIC FREQUENCY EFFECT IN XIONG ET AL. (2003) DATA

Studies on word frequency effects in repetition priming have produced mixed results. Studies using lexical decision and word identification tasks have generally reported more repetition priming for low frequency words than high frequency words (Jacoby & Dallas, 1981; Jacoby & Hayman, 1987; Kirsner et al., 1986; Scarborough et al., 1977). However, studies using word stem or fragment completion tasks have not always found greater facilitation for low frequency words (for a review, see Roediger & McDermott, 1993). The present work suggests that learning and memory is specific to the intentional acts at stimulus acquisition, the frequency information in stimulus representation should be specific to the frequency of the acquisition acts involving the stimulus, rather than frequency of the stimulus independent of acquisition acts. To test this hypothesis, this study collected intention-specific frequency ratings on words from subjects and analyzed the amount of repetition priming with respect to intention-specific frequency ratings in the same-transfer conditions in the Xiong et al. (2003) study.

Eight experiments in Xiong et al. (2003) included same-transfer conditions, that is, conditions in which the study and test semantic judgment tasks were repeated. Judgments in these experiments were based on six different semantic scales. Subjects in the current study rated how frequently they had thought of a word's meaning with respect to the semantic scales (e.g. *fast/slow*) when they had seen or heard the word. The rating is labeled $F_{scale}|word$, rather than $F_{scale.word}$, because subjects gave estimates of the relative frequency of a judgment task being performed when they had encountered a word, instead of the absolute frequency of performing the judgment on the word. If frequency effects in repetition priming are specific to the intentional acts at past stimulus acquisitions, we would expect words with low $F_{scale}|word$ ratings with respect to a judgment to show more repetition priming in the judgment task than words with high $F_{scale}|word$ ratings.

Method

Subjects

Eighteen Vanderbilt undergraduate students participated for experimental credits for courses. All subjects were native speakers of English. They had normal or corrected-to-normal vision. All subjects signed a consent form for participation.

Tasks and stimuli

Six semantic judgment tasks were used in the Xiong et al. (2003) study. They were based on the semantic scales of *active/passive*, *fast/slow*, *pleasant/unpleasant*, *strong/weak*, *valuable/worthless*, and *young/old* (Osgood et al., 1957). Subjects' task in the current study was to rate how often they had thought of a word's meaning in terms of these semantic scales. A subset of 130 words was selected from the words used in the Xiong et al. study. The words ranged from four to eight letters in length. The mean word frequency was 177 per million, *SD* = 156 (Kučera & Francis, 1967).

Procedure

Subjects rated how often they had thought about a word's meaning in terms of a semantic scale when they had seen or heard the word. A five-point rating scheme was used. A rating of one indicated that the subject had *very rarely* considered the word's meaning in terms of the semantic scale. A rating of five suggested that the subject had *very frequently* thought of the word in terms of the scale.

> When you have seen or heard the word
>
> airplane
>
> how often have you thought of its meaning in terms of FAST/SLOW?   5
>
> 1=very rarely 2=rarely 3=sometimes 4=frequently 5=very frequently

Figure 1. An example trial for the $F_{scale}|word$ rating.

Instructions for the subjects included examples. Subjects performed six blocks of semantic scale ratings, one for each scale. The order of the blocks was Latin-square counterbalanced. There were 130 trials in each block. Subjects initiated the blocks. The leading question and reminder of the rating scheme remained on screen throughout a block. The semantic scale being rated was presented in upper case font at the end of the leading question. Figure 1 illustrated what a trial looked like. On each trial, a word was presented in lower case font in between the two lines of the leading question. Subjects responded by pressing the "h", "j", "k", "l", and ";" keys, which were labeled with the ratings of 1 through 5. The rating that a subject had just given was shown after the leading question for 300 ms before the word and the rating both disappeared. The next trial started 500 ms later with the presentation of a new word. It took a subject anywhere between 40 to 60 minutes to finish.

Results and Discussion

Mean frequency ratings were calculated for each semantic scale and word combination, collapsing across subjects. There were altogether 6 (scale) × 130 (word) mean frequency ratings. The mean frequency rating and standard deviation for each scale, collapsing across words and subjects, are listed in Table 1. Given the relative frequency rating scale ranged from 1 to 5 from least frequent to most frequent, for every semantic scale, words that received mean frequency ratings of 3 or higher were assigned to the high $F_{scale}|word$ rating list. Words that received mean frequency ratings less than 3 were assigned to the low $F_{scale}|word$ rating list. Table 1 contains the number of words that were assigned to each list.

There were eight experiments in Xiong et al. (2003) that involved the same 2 (test judgment, same or different) × 2 (item type, old or new) design. In that work, subjects performed a semantic judgment, for example, *pleasant/unpleasant* judgment, on 40 unique words during study. They were then tested on the same judgment (same-transfer condition) as well as a different judgment task (cross-transfer condition) on words. The same versus different judgment task conditions were blocked. Half of the words for each test judgment were old items from the study phase, half of them were new. Repetition priming was calculated as the difference between RTs to old and new items at test. Same-transfer conditions produced the greatest amounts of repetition priming. The current study focused on intention-specific frequency effects in same-transfer conditions.

Table 1. Mean $F_{scale}|word$ Ratings and the Number of Words in High versus Low $F_{scale}|word$ rating lists

	Mean frequency rating		$F_{scale}|word$ list	
Semantic scale	*M*	*SD*	high	low
active/passive	2.21	0.07	24	106
fast/slow	2.11	0.08	25	105
pleasant/unpleasant	2.90	0.08	53	77
strong/weak	2.52	0.07	36	94
valuable/worthless	2.61	0.07	47	83
young/old	2.08	0.07	22	108

RTs in the eight same-transfer conditions (Xiong et al., 2003) were sorted depending on whether they were RTs to words in the high versus low $F_{scale}|word$ rating list with respect to the semantic judgment task. RTs to both high and low $F_{scale}|word$ rating words were then sorted again according to the old/new status of the words. As can be seen from Table 1, there are much fewer words with high $F_{scale}|word$ than low $F_{scale}|word$ ratings. This led to an average of only five observations from each subject in the high $F_{scale}|word$ rating cells, and approximately 12 observations in the low $F_{scale}|word$ rating cells. The following analyses were thus collapsed across all semantic-transfer conditions.

Subjects' mean RTs at test were analyzed with a 2 ($F_{scale}|word$, high or low) × 2 (item type, old or new) repeated measures ANOVA. Means for the ANOVA are displayed in Figure 2. Items with high $F_{scale}|word$ ratings were responded to 119 ms quicker than items with low $F_{scale}|word$ ratings. Old items were responded to 128 ms faster than new items at test. Both the main effects of $F_{scale}|word$ and item type were significant, $F_{(1,255)}$'s = 49.09 and 52.55, p's < 0.001, and partial eta square η^2's = 0.16 and 0.17, respectively. More critically, consistent with the task-specific learning hypothesis, old items with low $F_{scale}|word$ ratings showed 102 ms more repetition priming than old items with high $F_{scale}|word$ ratings. The interaction between $F_{scale}|word$ and old new item type was highly reliable, $F_{(1,255)} = 11.40$, $p = 0.001$, and $\eta^2 = 0.04$.

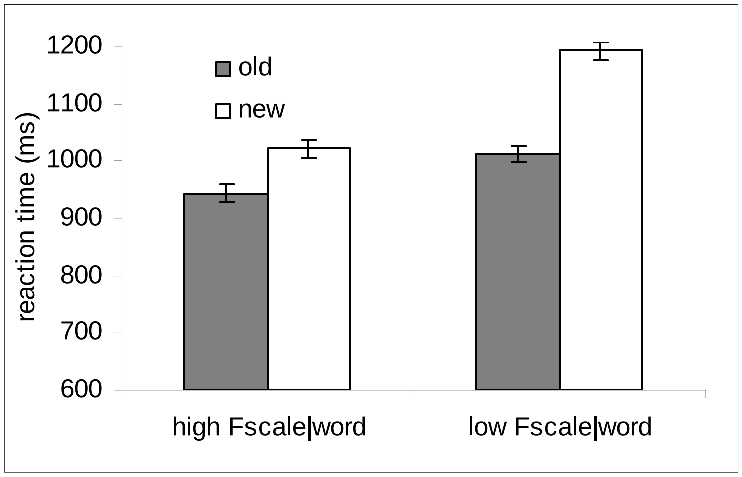

Figure 2. Mean reaction times for items with high versus low intention-specific frequency (*$F_{scale}|word$*) ratings. Error bars represent stand error from the ANOVA.

To show that the frequency effects were specific to the semantic judgment tasks, rather than a general effect of word frequency or word familiarity, the same analysis was repeated with word frequency (Kučera & Francis, 1967) and with word familiarity ratings (Gilhooly & Logie, 1980). For the word frequency analysis, test RTs in same-transfer conditions were sorted into high word frequency ($M = 376$ per million, $SD = 189$) versus low word frequency ($M = 31$ per million, $SD = 17$) halves. The amount of repetition priming did not differ between high and low frequency words. No interaction between word frequency and old new item type was observed, $F_{(1,255)} < 1$, *ns*. Similarly, when test RTs in same-transfer conditions were sorted according to high versus low word familiarity ratings ($M = 6.32$ and 5.45, $SD =$ 0.27 and 0.38, respectively), no difference in repetition priming was detected between words with high versus low familiarity ratings, $F_{(1,255)} < 1$, *ns*. Repetition priming in semantic judgments was not affected by word frequency or familiarity ratings. Frequency effects in repetition priming were specific to the intentional acts involving the stimulus at past acquisitions.

CHAPTER III

EXPERIMENT 1

Word frequency effects (more repetition priming for low frequency words than for high frequency words) have been reported in repetition priming studies, albeit inconsistently (for a review, see Roediger & McDermott, 1993). The present study hypothesized that intentional history is an inherent part of stimulus representation, and that frequency effects should be specific to the frequency of the intentional acts involving the stimulus at past acquisitions, rather than frequency of the stimulus in the environment. The post hoc study on data from Xiong et al. (2003) supported this hypothesis. However, the intention-specific frequency measure in the post hoc study ($F_{scale}|word$) confounded the frequency and the probability of a judgment being performed on words. To go beyond that, the present set of experiments relied on word co-occurrence counts from Google searches to measure the frequency and probability of a judgment being performed on words. This afforded us the opportunity to disambiguate the effects of intention-specific frequency ($F_{scale.word}$) and intention-specific probability ($P_{scale}|word$) on learning and memory.

Experiment 1 examined the effects of intention-specific frequency and intention-specific probability on repetition priming with a 2 ($F_{scale.word}$, high versus low) × 2 ($P_{scale}|word$, high versus low) × 2 (item type, old versus new) within-subject design. If frequency effects in repetition priming are specific to the frequency of past acquisition acts, we should expect more repetition priming for words with low $F_{scale.word}$ values than for words with high $F_{scale.word}$ values, but there might not necessarily be more repetition priming for words with low $P_{scale}|word$ values than for words with high $P_{scale}|word$ values.

Method

Subjects

Twenty-four Vanderbilt undergraduate students participated in the experiment for experimental credits for courses. All subjects were native speakers of English. They had normal or corrected-to-normal vision. All subjects signed a consent form for participation. Data from one subject were replaced due to extremely long RTs (more than 2.5 *SD* away from mean RTs for all subjects).

Tasks and stimuli

The study and test tasks involved making *beautiful/ugly* judgments on the meanings of words. A total of 104 nouns were used, 96 of which were experimental items. Experimental items included four equal subsets of words in accordance with the factorial combination of high versus low $F_{scale.word}$ and $P_{scale}|word$ values with respect to *beautiful/ugly* judgments. The average $F_{scale.word}$ and $P_{scale}|word$ values for each subset of words are listed in Table 2. The average word frequencies (F_{word}) are also included for reference. Efforts were made to equate the numbers of items that could be judged *beautiful* versus *ugly* for each subset of words.

Table 2. Average word frequency (F_{word}), intention-specific frequency ($F_{scale.word}$), intention-specific probability ($P_{scale}|word$), and repetition priming (RP) for Experiment 1 word subsets.

subset	F_{word}	$F_{scale.word}$	$P_{scale}\|word$	RP (ms)
1	13,224,167	2,862,892	0.23	89
2	154,025,000	4,474,792	0.04	86
3	975,625	222,466	0.24	120
4	3,707,836	124,753	0.04	133

Procedures

During the study phase, subjects performed *beautiful/ugly* judgment on 48 experimental items, 12 from each of the four subsets. After a 30-sec non-filled break, subjects initiated the test phase. The test phase involved making the same judgment on all the old items from the acquisition, as well as 48 new items. There were four filler trials at the beginning of both the study and test phases. Presentation order of items was randomized for each subject. The counterbalanced factors were: old versus new item sets, the left-right hand assignment of *beautiful* and *ugly* responses, and whether the test phase had the same response mapping as acquisition or not.

Subjects were tested individually. Task instructions and materials were presented on 8088 personal computers in 80-column lowercase font, in the default display color of green or orange. Each letter was about 2.5 × 2.5mm or 2.5 × 5mm (with extender) in size. The instructions included examples and stated that there was no right or wrong answer in the judgments. Subjects were asked to respond as soon as they had the impression of the meaning of words with regard to the judgment instead of spending a lot of time thinking. Subjects read the instructions and were then presented words one at a time. A reminder for hand response mapping remained on the screen throughout the blocks. Subjects indicated different responses by pressing the “z” or "/?" key on the keyboard.

Each trial started with a ready signal (*) that appeared in the middle of the screen for 500 ms. It was followed by the presentation of a word which remained on the screen until the subject responded. The interval from the previous trial response to the next ready signal was 500 ms. The presentation order, response type, and RT were recorded for experimental items.

Results and Discussion

All RTs from the test phase were entered into the analyses for there were no absolute right or wrong answers to *beautiful/ugly* judgments. Median RTs from each subject were used instead of mean

RTs because the RT distributions in the semantic judgments tended to be skewed towards long RTs. The median RTs were analyzed with a 2 ($F_{scale.word}$, high versus low) × 2 ($P_{scale}|word$, high versus low) × 2 (item type, old versus new) repeated measures ANOVA. Means for the ANOVA are displayed in Figure 3.

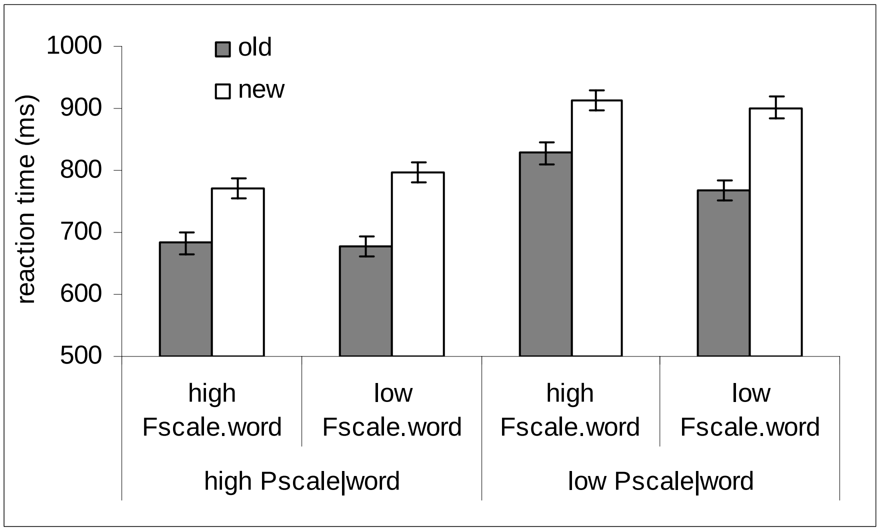

Figure 3. Mean reaction times for items with high versus low intention-specific frequency ($F_{scale.word}$) and high versus low intention-specific probability ($P_{scale}|word$). Error bars stand for stand error from the ANOVA.

Words with a high probability of being judged in terms of *beautiful/ugly*, that is, words with high $P_{scale}|word$ values, were responded to 121 ms faster than words with low $P_{scale}|word$ values. Old items were responded to 107 ms quicker than new items at test. The main effects of $P_{scale}|word$ and item type were highly significant, $F_{(1,23)}$'s = 20.31 and 44.18, p's < 0.001, η^2's (partial eta squares) = 0.47 and 0.66. More importantly, consistent with the task-specific learning hypothesis, words with low $F_{scale.word}$ values received 39 ms more repetition priming than words with high $F_{scale.word}$ values. The interaction between $F_{scale.word}$ and old new item type was significant, $F_{(1,23)} = 4.22$, $p = 0.05$, $\eta^2 = 0.16$. In contrast, the amount of repetition priming did not differ between words with low $P_{scale}|word$ values and words with high $P_{scale}|word$ values, $F_{(1,23)} < 1$, *ns*. No three way interaction between the factors was observed, $F_{(1,23)} < 1$, *ns*.

Because there was no absolute right or wrong answer in the judgment task, response consistency was analyzed as an indicator of response validity (Xiong et al., 2003). If an item received the same *beautiful* or *ugly* response in the test phase as in acquisition, response to that item was deemed consistent. Percentage consistency rates for old items at test were analyzed with a 2 ($F_{scale.word}$, high versus low) × 2 ($P_{scale}|word$, high versus low) × 2 (response type, *beautiful* versus *ugly*) repeated measures ANOVA. The only significant effect was the main effect of $P_{scale}|word$, $F_{(1,22)} = 13.16$, $p = 0.001$, $\eta^2 = 0.37$. There were a greater proportion of consistent responses for old words with high $P_{scale}|word$ values than for old words with low $P_{scale}|word$ values, 93% and 87%, respectively. The relatively high consistency rate, even for items with low $P_{scale}|word$ values, suggested good response validity despite the sometimes odd judgments (imagine making *beautiful/ugly* judgment on *membrane*). The fact that consistency rates did not differ statistically between old words with high $F_{scale.word}$ values and old words with low $F_{scale.word}$ values excluded response consistency as a potential confounding factor behind the $F_{scale.word}$ effect in repetition priming.

One concern with Experiment 1 is that both the $F_{scale.word}$ and $P_{scale}|word$ values were confounded with word frequency, F_{word} (see Table 2 for detailed values). However, while both $F_{scale.word}$ and $P_{scale}|word$ values for the word subsets co-varied with F_{word} values, they exhibited completely independent influences on RT and repetition priming. This implies that F_{word} could not have been responsible for the whole pattern of results in this experiment. While it is very unlikely that the main effect of $P_{scale}|word$ on RT was due to F_{word} (high $P_{scale}|word$ value items, while having low F_{word} values, led to much quicker responses), it is nonetheless possible that the $F_{scale.word}$ effect was in fact an F_{word} effect.

A linear regression analysis was therefore performed, with repetition priming as the dependent variable, and F_{word}, $F_{scale.word}$, and $P_{scale}|word$ as the predictor variables. Repetition priming was calculated for each word, collapsing across subjects. Repetition priming was calculated as $RT_{acq} - RT_{old}$ rather than $RT_{new} - RT_{old}$ because the latter comparison was between subjects (Repetition priming, when calculated as $RT_{acq} - RT_{old}$, yielded identical pattern of results as the above ANOVA). The regression analysis used the backward variable elimination method. The two variables that remained in the regression analysis were $F_{scale.word}$ and $P_{scale}|word$, with $R^2 = 0.09$, $F_{(2,93)} = 4.64$, $p = 0.01$. The standardized coefficients (β's) for $F_{scale.word}$ and $P_{scale}|word$ were -.28 and -.17, t's = -2.80 and -1.71, p's = 0.006 and 0.09. This indicated that it was indeed the intention-specific frequency, $F_{scale.word}$, instead of F_{word} that was driving the frequency effect.

In summary, Experiment 1 showed that learning as indexed by the amount of repetition priming was specific to the acquisition intentional act involving the word stimulus. Furthermore, Experiment 1 demonstrated that the effects of $F_{scale.word}$ and $P_{scale}|word$ on semantic judgments were dissociable. While $F_{scale.word}$ affected the amount of repetition priming, $P_{scale}|word$ had large impact on semantic judgment RTs. It is likely that $F_{scale.word}$ and $P_{scale}|word$ should have dissociable influences in other types of memory tasks as well. This would be addressed in the up-coming experiments.

CHAPTER IV

EXPERIMENT 2

The regression analysis in Experiment 1 implied that it was intention-specific frequency ($F_{scale.word}$) rather than word frequency (F_{word}) that gave rise to the frequency effect. Experiment 2 directly tested the effects of $F_{scale.word}$ and F_{word} by varying the $F_{scale.word}$ and F_{word} values of words factorially. A strong intentional history approach would argue that all instances are encoded with their specific intention, and that intention-free representations of objects do not exist. It would predict the amount of learning to be a function of the frequency of the intentional act involving the stimulus at past acquisitions rather than the frequency of the stimulus. This position would be supported if there was more repetition priming for items with low $F_{scale.word}$ values than for items with high $F_{scale.word}$ values, and at the same time there was the same amount of repetition priming for items with high F_{word} values as for items with low F_{word} values. A weaker position would acknowledge intention-specific as well as more abstract representations of objects. The weaker position should be adopted if the amount of repetition priming differed between items with high and low $F_{scale.word}$ values as well as between items with high and low F_{word} values.

Method

Subjects

Thirty-two Vanderbilt undergraduate students participated in the experiment for experimental credits for courses. All subjects had English as their native language and normal or corrected-to-normal vision. All subjects signed a consent form for participation. Data from two subjects with extremely long RTs (more than 2.5 *SD*s away from the mean RT across all subjects) were replaced.

Tasks and stimuli

As in Experiment 1, both the study and test phases involved making *beautiful/ugly* judgments on the meanings of words. There were altogether eight filler items and 144 experimental items in this experiment. Experimental items included four equal subsets of words in accordance with the factorial combination of high versus low $F_{scale.word}$ values with respect to *beautiful/ugly* judgments and high versus low F_{word} values. The average $F_{scale.word}$ and F_{word} values for the word subsets are listed in Table 3. The average task-specific probabilities ($P_{scale}|word$) are also included in Table 3 for reference. Efforts were made to equate the numbers of items that could be judged *beautiful* versus *ugly* for each subset of words.

Procedures

The experimental procedures remained the same as in Experiment 1.

Table 3. Average word frequency (F_{word}), intention-specific frequency ($F_{scale.word}$), intention-specific probability ($P_{scale}|word$), and repetition priming (RP) for Experiment 2 word subsets.

subset	F_{word}	$F_{scale.word}$	$P_{scale}\|word$	RP (ms)
1	2,625,556	214,497	0.10	156
2	2,809,167	68,359	0.03	205
3	832,944	200,864	0.24	108
4	772,139	67,240	0.10	242

Results and Discussion

All RTs from the test phase were included in the analyses. A 2 ($F_{scale.word}$, high versus low) × 2 (F_{word}, high versus low) × 2 (item type, old versus new) repeated measures ANOVA was performed on subjects' median RTs. Figure 4 displays the means from the ANOVA. Items with high $F_{scale.word}$ values were responded to 126 ms faster than items with low $F_{scale.word}$ values, $F_{(1,31)} = 23.57$, $p < 0.001$, $\eta^2 = 0.43$. Items with high F_{word} values were responded to 97 ms slower than words with low F_{word} values, $F_{(1,31)} = 13.23$, $p = 0.001$, $\eta^2 = 0.30$. Old items were responded to 178 ms faster than new items at test, $F_{(1,31)} = 25.15$, $p < 0.001$, $\eta^2 = 0.45$. More critically, items with low $F_{scale.word}$ values received 91 ms more repetition priming than items with high $F_{scale.word}$ values. This interaction between $F_{scale.word}$ and old new item type was statistically reliable, $F_{(1,31)} = 5.33$, $p < 0.05$, $\eta^2 = 0.15$. In contrast, repetition priming did not differ between items with high versus low F_{word} values. No interaction between F_{word} and old new item type was indicated, $F_{(1,31)} < 1$, *ns*. The three way interaction between $F_{scale.word}$, F_{word}, and item type approached significance, $F_{(1,31)} = 4.20$, $p = 0.05$, $\eta^2 = 0.12$.

As a measure of response validity, the percentages of old items that had received the same response at study and test were analyzed with a 2 ($F_{scale.word}$, high versus low) × 2 ($P_{scale}|word$, high versus low) × 2 (response type, *beautiful* versus *ugly*) repeated measures ANOVA. Words with high $F_{scale.word}$ values generated a greater proportion of consistent responses than words with low $F_{scale.word}$ values, 87% and 83%, respectively, $F_{(1,29)} = 4.57$, $p < 0.05$, $\eta^2 = 0.14$. No other effects were significant.

The difference in response consistency rates between words with high versus low $F_{scale.word}$ values could not have been responsible for the different amounts of repetition priming between words with high versus low $F_{scale.word}$ values. In this experiment, consistent responses were made over 300 ms faster than inconsistent responses (812 versus 1154 ms). With a greater proportion of consistent responses, old items with high $F_{scale.word}$ values should have a shorter average RT than old items with low $F_{scale.word}$ values. When compared to RTs for new items, old items with high $F_{scale.word}$ values should have shown more, instead of less, repetition priming than old items with low $F_{scale.word}$ values. That is, if consistency rates had had any impact on repetition priming in this experiment, we should have observed an $F_{scale.word}$ effect opposite to what was shown in the results.

A linear regression analysis as in Experiment 1 was performed, with repetition priming as the dependent variable, and F_{word}, $F_{scale.word}$, and $P_{scale}|word$ as the predictor variables. Repetition priming was calculated as $RT_{acq} - RT_{old}$ for each word, collapsing across subjects. The regression analysis used the backward variable elimination method. The only variable that remained in the regression analysis were $F_{scale.word}$, with $R^2 = 0.08$, $F_{(1,142)} = 11.76$, $p = 0.001$. The standardized coefficient (β) for $F_{scale.word}$ was -.28, $t = -3.43$, $p = 0.001$. This supported the hypothesis that intention-specific frequency rather than word frequency underlies frequency effects in repetition priming.

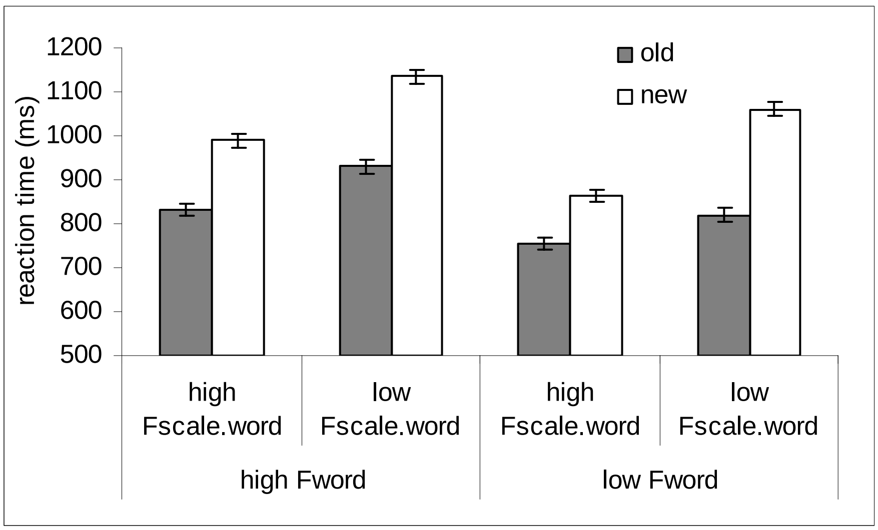

Figure 4. Mean reaction times for items with high versus low intention-specific frequency ($F_{scale.word}$) and high versus low word frequency (F_{word}). Error bars stand for stand error from the ANOVA.

Overall, Experiments 1 and 2 demonstrated that repetition priming in semantic judgments varied as a function of $F_{scale.word}$, the frequency with which the intentional act under investigation had been performed on words. At the same time, Experiments 1 and 2 showed that repetition priming did not correlate with F_{word}, the word frequency in the environment. These results are consistent with the hypothesis that intentional history is part of a word's representation but not word frequency independent of intentional acts.

CHAPTER V

EXPERIMENT 3

Experiments 1 and 2 demonstrated that frequency effects in the repetition priming paradigm, a typical implicit memory measure, were intention-specific. Frequency effects have also been consistently found in recognition tests, a traditional explicit memory measure. After subjects studied a list of words, they could more accurately recognize low frequency words than high frequency words (Gillund & Shiffrin, 1984; MacLeod & Kampe, 1996). This experiment examined whether frequency effects in recognition tests were also specific to the intentional acts at past stimulus acquisitions. Like in Experiment 2, this experiment was designed to disentangle the effect of intention-specific frequency ($F_{scale.word}$) and word frequency (F_{word}) by varying them factorially. If explicit memory for words was intention-specific, we should expect subjects to be better at recognizing words with low $F_{scale.word}$ values than word with high $F_{scale.word}$ values, and yet not necessarily better at recognizing words with low F_{word} values than word with high F_{word} values.

Method

Subjects

Twenty-four Vanderbilt undergraduate students participated in the experiment for experimental credits for courses. All subjects had English as their native language and normal or corrected-to-normal vision. All subjects signed a consent form for participation. Data from one subject were replaced because of extremely low sensitivity in distinguishing old from new items ($hit\% - false\ alarm\% < 0$).

Tasks and stimuli

The study phase involved making *strong/weak* judgments on the meanings of words. In the test phase subjects decided whether words had been presented during acquisition or not. There were six filler items and 144 experimental items in the experiment. Experimental items included four equal subsets of words in accordance with the factorial combination of high versus low $F_{scale.word}$ values with respect to *strong/weak* judgments and high versus low F_{word} values. Efforts were made to equate the numbers of items that could be judged *strong* versus judged *weak* for each subset of words.

Procedures

Subjects performed a block of *strong/weak* judgments on the meanings of words during study. They were instructed to respond as quickly as possible while avoiding errors when a word was presented. Subjects were not informed of the following recognition test. After four filler trials at the beginning of the acquisition, there were 72 experimental trials, 18 for each of the four subsets of words. There was a 30-sec non-filled break between study and test phases. Subjects initiated the test phase when they were ready.

They were instructed to respond *old* to words that had been presented in the previous block and respond *new* to words that had not been presented. The first four trials in the test phase were filler trials. The rest 144 trials included words from the study phase as well as 72 new words. All trials in the study and test phases started with a ready signal (*) that appeared in the middle of the screen for 500 ms. A word was then presented and remained on the screen until the subject responded by pressing the “z” or “?/” key. The interval from the previous trial response to the next ready signal was 500 ms. Presentation order of the experimental items was randomized for each subject. The counterbalanced factors were: old versus new item sets, the left-right hand assignment of *strong* and *weak*, and of *old* and *new* responses.

Table 4. Average word frequency (F_{word}), intention-specific frequency ($F_{scale.word}$), intention-specific probability ($P_{scale}|word$), study phase RT, and recognition performance (d') for Experiment 3 word subsets. Numbers in parenthesis are standard errors from the ANOVA.

subset	F_{word}	$F_{scale.word}$	$P_{scale}\|word$	RT_{study} (ms)	d'
1	3,368,611	449,731	0.13	1026(38)	2.43
2	3,310,278	189,308	0.06	885(38)	2.49
3	1,262,056	415,306	0.33	969(38)	1.86
4	1,207,389	172,428	0.14	1035(38)	2.32

Results and Discussion

Recognition performance was measured by subjects’ sensitivity in distinguishing old from new items. Table 4 lists the d's ($Z_{hit} - Z_{false\ alarm}$) for the four subsets of words. The Z-scores for hits and false alarms were based on the percentage hit and false alarm rates averaged across subjects. Figure 6 shows subjects’ recognition sensitivity as calculated by *hit% − false alarm%.* A 2 ($F_{scale.word}$, high versus low) × 2 (F_{word}, high versus low) repeated measures ANOVA was performed on the latter sensitivity measure (See Figure 5). Words with low $F_{scale.word}$ values were better recognized than words with high $F_{scale.word}$ values, $F_{(1,23)} = 7.14$, $p = 0.01$, $\eta^2 = 0.24$. Words with high F_{word} values were better recognized than words with low F_{word} values, $F_{(1,23)} = 22.35$, $p < 0.001$, $\eta^2 = 0.49$. The interaction between $F_{scale.word}$ and F_{word} was marginally significant, $F_{(1,23)} = 3.59$, $p = 0.07$, $\eta^2 = 0.14$.

In order to see whether recognition performance was contingent on how long it took subjects to process different subsets of words, subjects’ median RTs in the study phase were analyzed with a 2 ($F_{scale.word}$, high versus low) × 2 (F_{word}, high versus low) repeated measures ANOVA. Means for the ANOVA are shown in Table 4. The pattern of results did not correspond with subjects’ recognition performance on the four subsets of words. Words with high $F_{scale.word}$ values were responded to slightly slower (37 ms) than words with low $F_{scale.word}$ values, but the difference was only marginally significant, $F_{(1,23)} = 3.49$, $p = 0.08$, $\eta^2 = 0.13$. No main effect of F_{word} was observed, $F_{(1,23)} = 1.42$, *ns*. The interaction between $F_{scale.word}$ and F_{word} was statistically significant, $F_{(1,23)} = 7.51$, $p = 0.01$, $\eta^2 = 0.25$. Such a pattern of results suggested that differences in recognition performance for the four subsets of words were not simply consequences of differential encoding times.

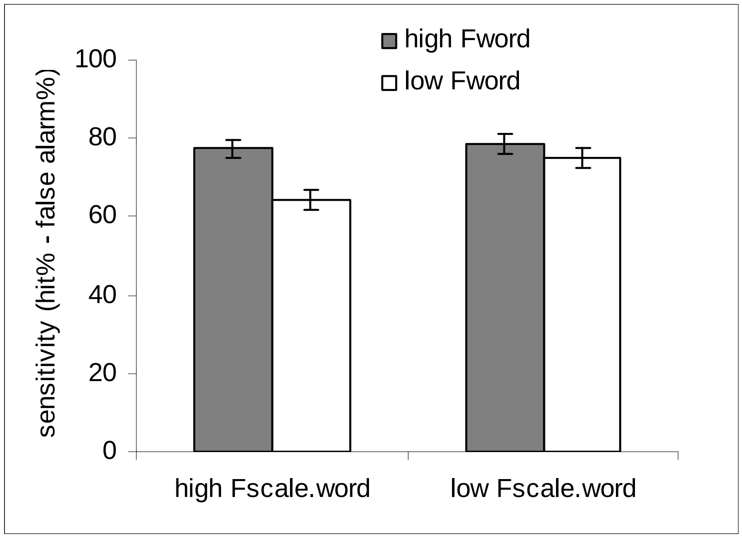

Figure 5. Recognition sensitivity (*hit% − false alarm%*) for items with high versus low intention-specific frequency ($F_{scale.word}$) and high versus low word frequency (F_{word}). Error bars stand for stand error from the ANOVA.

The fact that words with high F_{word} values were much better recognized than words with low F_{word} values prompted an alternative interpretation for the observed recognition results. A highly significant word frequency effect was observed in this experiment, yet it was the exact opposite of traditional word frequency effects in recognition (Gillund & Shiffrin, 1984; MacLeod & Kampe, 1996). High frequency words had been traditionally associated with worse instead of better recognition performance than low frequency words. A plausible explanation is that $P_{scale}|word$, instead of F_{word}, led to the difference between high and low F_{word} value items. As shown in Table 4, words with high F_{word} values also had lower $P_{scale}|word$ values. Rather than words with high F_{word} values being better recognized than words with low F_{word} values, it is arguable that low $P_{scale}|word$ value items were better recognized than high $P_{scale}|word$ value items. In fact, there was a perfect inverse relation between average $P_{scale}|word$ values and *d*'s (see Table 4) for the four subsets of words. $P_{scale}|word$ might be the most important factor in determining word recognition performance. This speculation was tested with a linear regression analysis.

The dependent variable of the linear regression analysis was the recognition sensitivity for each item, calculated as the percentage of times the item was corrected recognized as being old (hit) minus the percentage of times the item was incorrectly identified as being old (false alarm). Different subjects contributed to the percentages of hits and false alarms. The predictor variables were F_{word}, $F_{scale.word}$, and $P_{scale}|word$. The regression analysis used the backward variable elimination method. The only variable that remained in the regression analysis were $P_{scale}|word$, with $R^2 = 0.09$, $F_{(1,142)} = 13.54$, $p < 0.001$. The standardized coefficient (β) for $P_{scale}|word$ was -.30, $t = -3.68$, $p < 0.001$. This supported the hypothesis that intention-specific probability rather than word frequency underlies frequency effects in word recognition.

CHAPTER VI

EXPERIMENT 4

Experiment 3 suggested that the probability with which a judgment had been performed on a word ($P_{scale}|word$) was the most important factor in determining how well the word could be recognized on a later occasion. However, Experiment 3 was not specifically designed to test the effect of $P_{scale}|word$ on word recognition. Experiment 4 re-examined the findings of Experiment 3 by varying the $F_{scale.word}$ and $P_{scale}|word$ values of words factorially. Words with low $F_{scale.word}$ or $P_{scale}|word$ values were expected to be better recognized than words with high $F_{scale.word}$ or $P_{scale}|word$ values.

Method

Subjects

Sixteen Vanderbilt undergraduates participated in the experiment for experimental credits for courses. They had English as their native language and normal or corrected-to-normal vision. All subjects signed a consent form for participation. Data from two subjects were replaced due to lack of sensitivity in distinguishing old from new items ($hit\% - false\ alarm\% < 0$).

Tasks and Stimuli

The study task was a *strong/weak* judgment on the meanings of words. The test task was a recognition test on words presented during study. There were six filler items and 144 experimental items in the experiment. The experimental items consisted of four equal subsets of words in accordance with the factorial combination of high versus low $F_{scale.word}$ values and high versus low $P_{scale}|word$ values with respect to *strong/weak* judgments. Table 5 contains the average F_{word}, $F_{scale.word}$, and $P_{scale}|word$ values for each subset of words.

Table 5. Average word frequency (F_{word}), intention-specific frequency ($F_{scale.word}$), intention-specific probability ($P_{scale}|word$), study phase RT, and recognition performance (d') for Experiment 4 word subsets. Numbers in parenthesis are standard errors from the ANOVA.

subset	F_{word}	$F_{scale.word}$	$P_{scale}|word$	RT_{study}(ms)	d'
1	2,265,833	617,417	0.28	1028(63)	2.14
2	6,956,667	662,378	0.10	800(63)	2.93
3	531,250	147,431	0.28	897(63)	2.61
4	1,938,611	145,972	0.09	971(63)	3.05

Procedures

The experimental procedures remained identical to the procedures in Experiment 3.

Results and Discussion

Recognition performance was measured by subjects' sensitivity in distinguishing old from new items. Table 5 includes the *d′*s ($Z_{hit} - Z_{false\ alarm}$) for the four subsets of words. The *Z*-scores for hits and false alarms were based on the percentage hit and false alarm rates averaged across subjects. The *d′*s did not correspond at all with the average F_{word} values. Figure 6 displays subjects' recognition sensitivity as calculated by *hit% − false alarm%*. A 2 ($F_{scale.word}$, high versus low) × 2 ($P_{scale}|word$, high versus low) repeated measures ANOVA was performed on the percent sensitivity measure. Both $F_{scale.word}$ and $P_{scale}|word$ had significant main effects on subjects' recognition performance, $F_{(1,15)}$'s = 5.55 and 13.32, *p*'s < 0.05 and 0.01, η^2's = 0.27 and 0.47, respectively. Items with low $F_{scale.word}$ or $P_{scale}|word$ values were better recognized than items with high $F_{scale.word}$ or $P_{scale}|word$ values. This was congruent with past studies on word frequency effects in recognition, which had found low frequency words to be better recognized than high frequency words (Gillund & Shiffrin, 1984; MacLeod & Kampe, 1996). The interaction between $F_{scale.word}$ and $P_{scale}|word$ approached significance, $F_{(1,15)} = 3.43$, $p = 0.08$, $\eta^2 = 0.19$. The $P_{scale}|word$ effect was smaller for words with low $F_{scale.word}$ values than for words with high $F_{scale.word}$ values. Given the high level of performance for words with low $P_{scale}|word$ values, the interaction could easily be the consequence of a ceiling effect.

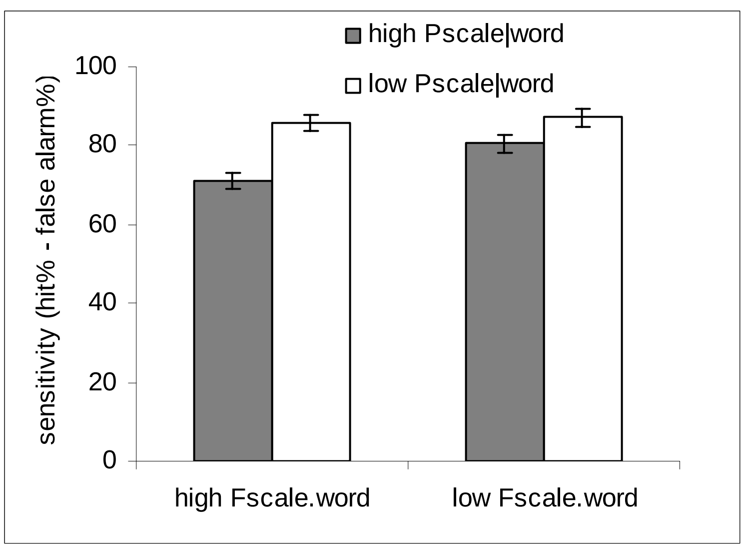

Figure 6. Recognition sensitivity (*hit% - false alarm%*) for items with high versus low intention-specific frequency ($F_{scale.word}$) and intention-specific probability ($P_{scale}|word$). Error bars stand for stand error from the ANOVA.

As in Experiment 3, subjects' word recognition performance did not depend on their RTs to the words at the study phase (see Table 5). Subjects' median RTs in making *strong/weak* judgment on the four subsets of words were analyzed with an ANOVA. The resulting pattern was quite different from the pattern in the recognition test. $F_{scale.word}$ had no impact on subjects' median RTs, $F_{(1,15)} < 1$, *ns*, whereas $P_{scale}|word$ did. Items with high $P_{scale}|word$ values were responded to 78 ms slower than items with low $P_{scale}|word$ values, $F_{(1,15)} = 8.58$, $p = 0.01$, $\eta^2 = 0.36$. Closer examination shows that $P_{scale}|word$ had opposite effects for items with high and low $F_{scale.word}$ values—when the $F_{scale.word}$ values were high, items with high $P_{scale}|word$ values were responded to slower than items with low $P_{scale}|word$ values; when the $F_{scale.word}$ values were low, items with high $P_{scale}|word$ values were actually responded to faster than items with low $P_{scale}|word$ values. The interaction between $F_{scale.word}$ and $P_{scale}|word$ was significant, $F_{(1,15)} = 5.68$, $p < 0.05$, $\eta^2 = 0.28$.

A linear regression analysis also pointed to $P_{scale}|word$ as the most important factor in word recognition. The linear regression analysis had the recognition sensitivity for an item as the dependent variable, which was calculated as the percentage of hits minus percentage of false alarms across subjects. The predictor variables were F_{word}, $F_{scale.word}$, and $P_{scale}|word$. The regression analysis used the backward variable elimination method. Like in Experiment 3, the only variable that remained in the regression analysis were $P_{scale}|word$, with $R^2 = 0.08$, $F_{(1,142)} = 12.92$, $p < 0.001$. The standardized coefficient (β) for $P_{scale}|word$ was -.29, $t = -3.59$, $p < 0.001$.

In Experiment 3, the $P_{scale}|word$ values of words were related to recognition performance for the words, with words having low $P_{scale}|word$ values being better recognized than words having high $P_{scale}|word$ values. While not designed to study the effect of $P_{scale}|word$, Experiment 3 nonetheless suggested that $P_{scale}|word$ is the main factor affecting recognition task performance. Results from Experiment 4 supported this observation. Experiment 4 also found $F_{scale.word}$ to have some effect on recognition performance, but to a much lesser degree than $P_{scale}|word$. Word frequencies seemingly had a large impact (although in the opposite direction from traditional word frequency effects) on recognition performance in Experiment 3. However, the regression analysis suggested that the seemingly word frequency effect was actually an intention-specific probability effect. In Experiment 4, when the $P_{scale}|word$ values and F_{word} values were un-confounded, the relation between word frequency and recognition performance disappeared. Both Experiment 3 and 4 supported the hypothesis that intentional history is a general property of word representations, and not simple word frequency independent of intentional acts.

CHAPTER VII

EXPERIMENT 5

Word frequency effects have also been found in free recall, a benchmark explicit memory test. Two patterns of word frequency effects have appeared in free recall, though, depending on the experimental procedures (Duncan, 1974; Gillund & Shiffrin, 1984; Gregg, 1976; Gregg et al., 1980; MacLeod & Kampe, 1996). When high and low frequency words were presented in separate lists, typically with a between-subject design, high frequency words were more likely to be recalled than low frequency words. In contrast, when high and low frequency words were presented in a mixed list, the recall advantage for high frequency words disappeared, and low frequency words tended to be better recalled than high frequency words. Experiment 5 tested whether frequency effects in free recall was intention-specific with a mixed-list, within-subject design. If explicit memory for words was contingent on the intentional acts at past stimulus acquisitions, we should expect better recall performance for words with low intention-specific frequency and (or) intention-specific probability than for words with high intention-specific frequency and (or) intention-specific probability.

Method

Subjects

Thirty-two Vanderbilt undergraduates participated in the experiment for experimental credits for courses. They had English as their native language and normal or corrected-to-normal vision. All subjects signed a consent form for participation.

Tasks and stimuli

The study task involved making a *strong/weak* judgment on the meanings of words. For the test task subjects were instructed to write down as many words from the study block as they could remember. There were 40 experimental items and four filler items in the experiment. The experimental items consisted of four equal subsets according to the factorial combination of high versus low $F_{scale.word}$ and high versus low $P_{scale}|word$ values. Table 6 listed the average $F_{scale.word}$ and $P_{scale}|word$ values for the subsets. The average word frequency, F_{word}, is also included for reference.

Procedures

Subjects performed a *strong/weak* judgment on the meanings of words at study. Subjects were asked to respond as quickly as possible while avoiding errors. They were not informed of a later recall test. There were two filler trials at the beginning of the study block and two filler trials at the end. Each trial started with a ready signal (*) that appeared in the middle of the screen for 500 ms. Then a word was presented and remained on the screen until the subject responded by pressing the “z” or “?/” key. The interval from the previous trial response to the next ready signal was 500 ms. The presentation order of experimental items was randomized for each subject. The left-right hand mapping assignment for *strong* and *weak* responses was counterbalanced. There was a 30-sec non-filled break after the acquisition. Subjects initiated the test phase when they were ready. They were given ten minutes to write down on a piece of paper as many words from the study phase as they could remember. When ten minutes had passed, subjects were told that the experiment was over and they were debriefed.

Table 6. Average word frequency (F_{word}), intention-specific frequency ($F_{scale \cdot word}$), intention-specific probability ($P_{scale}|word$), study phase RT, and percentages of words recalled for Experiment 5 word subsets. Numbers in parenthesis are standard errors from the ANOVAs.

subset	F_{word}	$F_{scale.word}$	$P_{scale}\|word$	RT_{study} (ms)	% recalled
1	11,246,000	2,994,700	0.27	1066(23)	27(3)
2	16,714,545	2,456,818	0.15	1107(23)	34(3)
3	3,790,000	1,056,900	0.28	1132(23)	30(3)
4	9,423,000	1,085,820	0.12	1061(23)	33(3)

Results and Discussion

For the free recall test, the exact reproduction, inaccurate spelling (e.g., "brethern" for *brethren*), and plural form of experimental items were counted as correct, but derivatives (e.g., "master" for *mastery*) and synonyms (e.g., "influx" for *inflow*) were not. Table 6 lists the percentages of items recalled for the four subsets of words. As predicted, subjects recalled words with low $P_{scale}|word$ values more often than words with high $P_{scale}|word$ values. A 2 ($F_{scale.word}$, high versus low) × 2 ($P_{scale}|word$, high versus low) repeated measures ANOVA confirmed that the effect of $P_{scale}|word$ on free recall was significant, $F_{(1,31)} = 4.77$, $p < 0.05$, $\eta^2 = 0.13$. Recall performance for words with low $F_{scale.word}$ values did not differ statistically from recall performance for words with high $F_{scale.word}$ values, $F_{(1,31)} < 1$, *ns*. No interaction between $P_{scale}|word$ and $F_{scale.word}$ was observed, $F_{(1,31)} < 1$, *ns*.

Free recall performance for the four subsets of words did not correlate with study phase RTs for the words (see Table 6). Subjects' median RTs in making *strong/weak* judgments at the study phase were analyzed. The resulting pattern was markedly different from the result pattern in free recall performance. The ANOVA showed no main effect of either $F_{scale.word}$ or $P_{scale}|word$, yet the interaction between $F_{scale.word}$ and $P_{scale}|word$ was significant, $F_{(1,31)} = 6.12$, $p < 0.05$, $\eta^2 = 0.17$. The pattern of the interaction was opposite of the pattern in Experiment 4—when the $F_{scale.word}$ values were high, items with high $P_{scale}|word$ values were responded to faster than items with low $P_{scale}|word$ values; when the $F_{scale.word}$ values were low, items with high $P_{scale}|word$ values were responded to slower than items with low $P_{scale}|word$ values.

A linear regression analysis was performed, with the percentage of times an item was correctly recalled as the dependent variable, collapsing across subjects, and F_{word}, $F_{scale.word}$, and $P_{scale}|word$ as the predictor variables. The regression analysis used the backward variable elimination method. The regression analysis failed to find a significant effect of any of the predictor variables.

The null result of the regression analysis notwithstanding, the observation that subjects were more likely to recall words with low $P_{scale}|word$ values than words with high $P_{scale}|word$ values paralleled established results on free recall with mixed frequency lists (Duncan, 1974; Gillund & Shiffrin, 1984; Gregg, 1976; Gregg et al., 1980; MacLeod & Kampe, 1996). When high frequency and low frequency words were presented in the same list, there was a tendency for subjects to recall low frequency words better than high frequency words. By dissociating the probability of a task being performed on a word and the frequency of the word, this experiment showed that the probability of a task being performed was the important factor in determining explicit memory task performance. In fact, the typical word frequency effect could be effectively reversed by varying the $P_{scale}|word$ values of words, such as was the case between word subsets 1 and 2, and between subsets 3 and 4 in Table 6.

CHAPTER VIII

EXPERIMENT 6

Experiment 6 measured free recall performance in separate-list, between-subject conditions. High word frequency has traditionally been associated with better recall performance when high versus low frequency words were presented in separate lists (Duncan, 1974; Gillund & Shiffrin, 1984; Gregg, 1976; Gregg et al., 1980; MacLeod & Kampe, 1996). In this experiment, high versus low intention-specific frequency ($F_{scale.word}$) and high versus low intention-specific probability ($P_{scale}|word$) items were presented in separate lists rather than mixed in the same list. If intentional history is a general property of memory representations, we should expect better recall performance for words with high $F_{scale.word}$ or $P_{scale}|word$ values than for words with low $F_{scale.word}$ or $P_{scale}|word$ values.

This experiment examined the simple effects of $F_{scale.word}$ and $P_{scale}|word$ separately rather than the factorial combination of $F_{scale.word}$ and $P_{scale}|word$, because the number of subjects required for the full between-subject factorial design would be prohibitive. Given the lack of interaction between $F_{scale.word}$ and $P_{scale}|word$ in the mixed-list, within-subject free recall Experiment 5, this focus on simple effects was reasonable. However, even with a focus on simple effects only, a complete between-subject design comparing the effects of high versus low $F_{scale.word}$ values and high versus low $P_{scale}|word$ values would still require a large number of subjects. The current experiment therefore employed two different semantic judgment tasks. They made it possible to gather data from the same subject (rather than a within-subject design factor) on both the $F_{scale.word}$ and the $P_{scale}|word$ comparisons while keeping between-subjects the most critical comparison, high versus low $F_{scale.word}$ values and high versus low $P_{scale}|word$ values.

Method

Subjects

Thirty-two Vanderbilt undergraduate students participated in the experiment for experimental credits for courses. All subjects were native speakers of English. They had normal or corrected-to-normal vision. All subjects signed a consent form for participation. Two subjects' data were replaced. One of them had study phase RTs that were almost 4 *SD*s away from the mean RT across all subjects. The other subject recalled an exceptionally high percentage of words at test, about 3 *SD*s away from the mean percentage across subjects.

Tasks and stimuli

The study tasks involved making *beautiful/ugly* and *strong/weak* judgments on the meanings of words. For the test task subjects were instructed to write down as many words from the study block as they could remember. The experiment used a total of 16 filler items and 160 experimental items, 80 for each judgment task. Experimental items for each judgment consisted of four subsets. Word subsets 1 and 2, and subsets 5 and 6, were constructed for the comparison between high and low $F_{scale.word}$ values. They varied on $F_{scale.word}$ values (HF versus LF) and were equated in terms of $P_{scale}|word$ values. Subsets 3 and 4, and subsets 7 and 8, were constructed for the comparison between high and low $P_{scale}|word$ values. They varied on $P_{scale}|word$ values (HP versus LP) and were equated in terms of $F_{scale.word}$ values. Detailed values are reported in Table 7.

Table 7. Average word frequency (F_{word}), intention-specific frequency ($F_{scale.word}$), intention-specific probability ($P_{scale}|word$), study phase RT, and percentages of words recalled for Experiment 6 word subsets. Numbers in parenthesis are standard errors from independent sample t-tests.

subset	F_{word}	$F_{scale.word}$	$P_{scale}\|word$	RT_{study} (ms)	% recalled
beautiful/ugly					
1	3,435,500	329,810	0.10	1108(133)	31(4)
2	499,800	49,842	0.10	1201(133)	28(4)
3	483,650	127,595	0.27	890(104)	45(5)
4	4,018,000	146,970	0.04	1198(104)	31(5)
strong/weak					
5	2,770,000	311,535	0.11	1249(157)	36(2)
6	402,200	48,941	0.12	1094(157)	36(2)
7	545,250	142,885	0.26	1308(120)	30(3)
8	2,392,000	130,445	0.06	1069(120)	34(3)

Procedure

Each subject completed four study-test sequences in this experiment, for example, *beautiful/ugly* (HF) – recall – *strong/weak* (LF) – recall – *beautiful/ugly* (HP) – recall – *strong/weak* (LP) – recall. High versus low $F_{scale.word}$ and high versus low $P_{scale}|word$ conditions for a judgment task were tested between subjects. As in the above example, *beautiful/ugly* and *strong/weak* judgments alternated between study-test sequences to minimize interference in recall across sequences. A subject would finish the two $F_{scale.word}$ sequences (HF and LF) before moving on to the $P_{scale}|word$ sequences (HP and LP). The order of *beautiful/ugly* and *strong/weak* judgments and the order of $F_{scale.word}$ and $P_{scale}|word$ sequences were counterbalanced. Also counterbalanced was the left-right hand mapping assignment for *strong*/*beautiful* (assigned to the same hand) and *weak/ugly* responses.

The instruction for the semantic judgment study blocks indicated which semantic judgment task subjects should perform. The instruction stated that subjects would be asked to recall the words later, and that when a word was presented subjects should respond as quickly as possible while avoiding errors. The semantic judgment study blocks started with two filler trials, followed by 20 experimental trials, and finally two filler trials at the end. The presentation order of experimental items was randomized. Each trial started with a ready signal (*) that appeared in the middle of the screen for 500 ms. Then a word was presented and remained on the screen until the subject responded by pressing the "z" or "?/" key. The interval from the previous trial response to the next ready signal was 500 ms.

After each study judgment block, there was a 30-sec break during which subjects did simple additions and subtractions on the computer. Subjects were then instructed to write down on a blank sheet of paper as many words from the previous block as they could remember. They had four minutes to write down the words. When four minutes had passed, subjects went on to do another two minutes of simple additions and subtractions. The next study-test sequence started after the math.

Results and Discussion

For free recall tests, the exact reproduction, inaccurate spelling (e.g., "brethern" for *brethren*), and plural form of experimental items were counted as correct, but derivatives (e.g., "master" for *mastery*) and synonyms (e.g., "influx" for *inflow*) were not. There were a few rare cases (amounting to 1% of all recalled words) in which subjects put down a word from a previous study-test sequence. Such cases were not included in correct recalls.

Consistent with the hypothesis, when subjects had performed *beautiful/ugly* judgments on the meanings of words at study, they were more likely to recall words with high $P_{scale}|word$ values (subset 3) than words with low $P_{scale}|word$ values (subset 4). This difference in recall performance was significant according to the independent sample t-test, $t_{(30)} = 3.17$, $p < 0.01$. This was congruent with the traditional finding in free recall with separate lists. High frequency words were more readily recalled than low frequency words when they were studied separately (Duncan, 1974; Gillund & Shiffrin, 1984; Gregg et al., 1980; MacLeod & Kampe, 1996). When $P_{scale}|word$ values were equated between word subsets (subsets 1 and 2), recall performance did not differ between words with high versus low $F_{scale.word}$ values, $t_{(30)} < 1$, *ns*.

The recall advantage for words with high $P_{scale}|word$ values in *beautiful/ugly* judgment was not a result of prolonged encoding time during study. As reported in Table 7, words with high $P_{scale}|word$ values had actually led to shorter median RTs than words with low $P_{scale}|word$ values, $t_{(30)} = 2.96$, $p < 0.01$. Median study phase RTs for words with high versus low $F_{scale.word}$ values did not differ statistically, $t_{(30)} < 1$, *ns*.

When subjects had made *strong/weak* judgments on the meanings of words, no significant effects of either $P_{scale}|word$ or $F_{scale.word}$ were observed, $t_{(30)} = 1.07$ and $t_{(30)} < 1$, respectively, *ns*. The lack of effect for $P_{scale}|word$ in *strong/weak* judgment was most likely a floor effect. Across the Kučera and Francis

(1976) word set from which stimuli for this experiment were selected, the median $P_{scale}|word$ value for *strong/weak* judgment was 0.19. In comparison, the median $P_{scale}|word$ value for *beautiful/ugly* judgment was 0.11. A $P_{scale}|word$ value of 0.26 was very high for *beautiful/ugly* judgment, but it was not so high for *strong/weak* judgment. Words with higher $P_{scale}|word$ values might be necessary for producing the $P_{scale}|word$ effect in *strong/weak* judgments, especially with a between-subject design.

Linear regression analyses on recall performance for the items produced similar results as the above analyses. The regression analyses had the percentage of times an item was correctly recalled following *beautiful/ugly* or *strong/weak* judgments as the dependent variable, collapsing across subjects. The regression analyses had F_{word}, $F_{scale.word}$, and $P_{scale}|word$ as the predictor variables. The backward variable elimination method was used. For the recall test following *beautiful/ugly* judgments, the only variable that significantly affected the recall of an item was $P_{scale}|word$, with $R^2 = 0.10$, $F_{(1,78)} = 8.95$, $p < 0.01$. The standardized coefficient (β) for $P_{scale}|word$ was .32, $t = 2.99$, $p < 0.01$. For the recall test following *strong/weak* judgments, no significant effect of any of the predictor variables was observed.

Experiment 5 reproduced the traditional pattern of word frequency effects in mixed-list recall with intention-specific probability instead of word frequency. Correspondingly, the current experiment showed that it is possible to reproduce the traditional pattern of word frequency effects in separate-list recall with intention-specific probability rather than word frequency. When subjects had performed *beautiful/ugly* judgments on the meanings of words at acquisition, the traditional word frequency effect in free recall was once again reversed by the variation in $P_{scale}|word$ values. As is shown in Table 7, words in subset 3 had lower word frequencies, yet they were more easily recalled than words in subset 4. Also notice that for word subsets 1 and 2, when the $P_{scale}|word$ values were equated, the F_{word} and $F_{scale.word}$ values were necessarily confounded. Yet even with combined force, F_{word} and $F_{scale.word}$ failed to significantly affect free recall performance.

CHAPTER IX

GENERAL DISCUSSION

The primary hypothesis in the present work is that intentional history is an inherent part of a word's representation, and that frequency effects in both implicit and explicit memory tasks should be specific to the frequency or probability of intentional acts at past acquisitions of a word, instead of the frequency of the word independent of acquisition acts. The present work tested this hypothesis by comparing the effects of intention-specific frequency ($F_{scale.word}$) and intention-specific probability ($P_{scale}|word$) to the effects of word frequency (F_{word}) in repetition priming, recognition, and free recall tests. The intention-specific frequency and probability are operationalized by the co-occurrences of the word stimulus and the semantic scale (e.g. *beautiful/ugly*) on the Internet, with the assumption that the number of times the word and the semantic scale co-occur should be proportional to the number of times people have intentionally processed the meaning of the word with respect to the semantic scale.

Consistent with the hypothesis, memory task performance, whether measured as repetition priming in semantic judgments, or sensitivity in recognizing old versus new items, or the percentage of items recalled, was affected by the frequency or probability of the intentional act involving the word, and not by word frequency. More specifically, in repetition priming tasks, words with low $F_{scale.word}$ values showed more repetition priming than words with high $F_{scale.word}$ values. In recognition tests, words with low $P_{scale}|word$ values, and to a lesser degree low $F_{scale.word}$ values, were more accurately recognized than words with high high $P_{scale}|word$ or $F_{scale.word}$ values. And finally, in free recall tests, when words with high and low $P_{scale}|word$ values were mixed in the same block at acquisition, words with low $P_{scale}|word$ values were more readily recalled than words with high $P_{scale}|word$ values. When words with high and low $P_{scale}|word$ values were presented separately during study, words with high $P_{scale}|word$ values were more likely to be recalled. The patterns of these intention-specific frequency and probability effects mirrored the traditional word frequency effect patterns in repetition priming (Duchek & Neely, 1989; Forster & Davis, 1984; Jacoby & Dallas, 1981; Jacoby & Hayman, 1987; Kirsner et al. 1986; MacLeod & Kampe, 1996; Scarborough et al., 1977), recognition (Gardiner & Java, 1990; Guttentag & Carroll, 1997; Jacoby & Dallas; Kinoshita, 1995; MacLeod & Kampe), and free recall tests (Duncan, 1974; Gregg, 1976; Gregg et al., 1980; MacLeod & Kampe).

The secondary goal of the present work is to compare the effects of the two different aspects of intentional history, intention-specific frequency and intention-specific probability, to examine whether they differentially affect explicit and implicit memory tests (Graf & Schacter, 1985; Schacter & Graf, 1986). The results showed dissociable influence of the two factors in explicit and implicit memory tests. While intention-specific frequency had major impact on repetition priming in semantic judgments, intention-specific probability turned out to be the primary factor in determining recognition and recall performances.

Intentional History versus Word Frequency

The present results found significant effects of the intentional history on memory for words, and no effect at all of simple word frequency. These results supported the idea that frequency information associated with a stimulus representation is specific to the intentional acts involving the stimulus. They suggest that it may be necessary to re-conceptualize word frequency effects with respect to past intentional acts involving words. The intentional history of words presents a parsimonious account for the existence of word frequency effects, as well as the lack of it in various memory tasks. According to the intentional history account, word frequency effects are found in word identification and lexical decision tasks because similar intentional processes are involved in everyday word acquisition situations, in the study phases of the experiments, and in the test phases (Duchek & Neely, 1989; Forster & Davis, 1984; Jacoby & Dallas, 1981; Jacoby & Hayman, 1987; Kirsner et al., 1986; Scarborough et al., 1977). In contrast, word frequency effects have not been consistently observed in word stem or fragment completion tasks (MacLeod, 1989; Roediger et al., 1992; Tenpenny & Shoben, 1992; for a review, see Roediger & McDermott, 1993), presumably because the tasks in which repetition priming was measured deviated from everyday word acquisition situations. In other words, word frequency effects are found in word identification and lexical decision tasks because word frequencies happen to match the intention-specific frequencies in these tasks. Word frequency effects are more difficult to find in word-stem or fragment completion tasks because word frequencies no longer match the intention-specific frequencies for these tasks.

Memory Systems versus Transfer Appropriate Processing

Advocates of multiple memory systems attribute the dissociation between explicit and implicit memory tasks to distinct memory systems these tasks involve (McClelland et al., 1995; Nadel, 1992; Schater & Tulving, 1994; Squire, 1986, 1992; Tulving, 1985; Tulving & Schacter, 1990). Processing theorists (Blaxton, 1989; Jacoby, 1983; Roediger, 1990; Roediger et al., 1989; Weldon & Roediger, 1987), on the other hand, argue that many dissociations between explicit and implicit memory tasks could be explained in terms of the overlap between study and test tasks in accordance with the transfer-appropriate processing principle (Bransford et al., 1979; Morris et al., 1977).

Consistent with the processing account of explicit and implicit memory tasks, the present work demonstrated that the intention-specific learning applies to both explicit and implicit memory tasks. However, the present work also indicated that there are important differences between explicit and implicit memory tasks in terms of how information is utilized, beyond what type of information (e.g., conceptual versus perceptual) is accessed. In the present work, repetition priming varied as a function of intention-specific frequency of words, whereas recognition and recall performances were mostly contingent on the intention-specific probability of words. This difference indicates that explicit and implicit memory tasks might involve distinctive modes of processing. As Sherry & Schacter (1987) stated, and echoed by Nadel (1992), the term multiple memory systems "refers to the idea that two or more systems are characterized by fundamentally different rules of operation" (p. 440). McClelland et al. (1995) proposed the distinction between the hippocampal and cortical systems because of differences in the modes of processing (i.e., rapid versus slow learning). According to results in the present work, while intention-specific learning should be a general principle for learning and memory, the differential impact of intention-specific frequency and intention-specific probability on explicit and implicit memory tasks is consistent with the idea of multiple memory systems operating on different rules of learning and retrieval.

The distinction between intention-specific frequency and intention-specific probability also addresses a long standing question about word frequency effects in repetition priming and recognition. Word frequencies have seemingly similar effects in repetition priming as in recognition tasks. Low frequency words show more repetition priming than high frequency words. They are also better recognized than high frequency words (Jacoby & Dallas, 1981; MacLeod & Kampe, 1996). Researchers have discussed whether word frequency effects in repetition priming and recognition share the same episodic origin (Duchek & Neely, 1989; Forster & Davis, 1984; Kinoshita, 1995). In the present work, frequency effects in repetition priming were intention-specific frequency effects, whereas "frequency effects" in recognition were largely intention-specific probability effects. This implies that while word frequency effects in both repetition priming and recognition tasks are episodic in origin, they arise from rather different mechanisms. Word frequency effects in repetition priming are not the same as word frequency effects in recognition tasks.

Abstract versus Episodic Representations

In accordance with the instance theory of memory (Logan, 1988), the present work assumes that each encounter with a stimulus is encoded and stored as a separate instance. Intention is instrumental in the formation of instances because intention as goal-orientedness requires attention. And attention has been suggested to be a sufficient condition for binding co-occurring elements into instances (Logan & Etherton, 1994). Building on previous studies on repetition priming (Franks et al., 2000; Thompson-Schill & Gabrieli, 1999; Vriezen et al., 1995; Xiong et al., 2003), recognition (Bransford et al., 1979; Morris et al., 1977), and recall (Tulving, 1979; Tulving & Thomson, 1973), the present work inferred that intention is an inherent part of memory instance, and theorized that stimulus representation should include a collection of intention-specific instances.

The experiment results supported this cumulative, intention-specific, instance view of memory representations. Intention-specific frequency or probability effects were found in both explicit and implicit memory tasks, whereas no effects of word frequency were observed. These results are hard to explain if words are represented as abstract lexical entries. The present work provides strong evidence for the instance, or episodic, view of semantic representations, and presents a challenge for the abstractionist view of semantic representations (for reviews see Bowers, 2000; Tenpenny, 1995).

The present work, however, with the blocked study and test design, does not exclude the possibility of a short-lived, abstract component of word representation that is not affected by semantic context or task demand. Some studies showed that with masked prime or very brief presentation of the prime frequency had no effect on repetition priming (e.g., Forster & Davis, 1984; Versace & Nevers, 2003). Repetition priming under such conditions were short-lived, fading into non-significance within just seconds. It was assumed to be due to transient variations in the activation level of abstract representations.

Priming studies on morphologically related words also supported the existence of abstract word representations. Fowler, Napps, and Feldman (1985) found the same amount of repetition priming from inflected or derived words to their base forms as from base forms to base forms, even when the inflected words and the base forms did not fully overlap in terms of orthographic or phonological representations (e.g. from "clarify" to "clear"). This was considered evidence for a common lexical entry for morphologically related and derived words.

Instead of siding with one or the other in the abstractionist-versus-episodic debate, the present work speculates that there may be at least two layers to word representations. The morphological studies (e.g., Fowler et al., 1985) may have reflected a more abstract layer of word representations, whereas studies like Jacoby and colleagues' (Jacoby, 1983; Jacoby et al., 1989), studies on semantic judgments (Thompson-Schill & Gabrieli, 1999; Vriezen et al., 1995; Xiong et al., 2003), and the present work have reached into a much richer, episodic, layer of semantic representations. Upon reviewing the abstractionist-versus-episodic debate, Tenpenny (1995) called for a hybrid model to account for all the empirical findings. The multi-layer view presents just such a model. Such a view exists implicitly in neurocomputational models of semantic memory in the form of separate layers of processing units (McClelland & Rumelhart, 1985; Rumelhart & Todd, 1993). It is consistent with the claim that there were multiple components to repetition priming (Forster & Davis, 1984; McKone, 1995; McNamara, 2005). It also parallels the distinction between activation patterns in left perisylvian (word form) areas and sensorimotor cortices in response to words (for a review, see Pulvermüller, 1999).

Context, Intentional History, Transfer, and Stroop

The present work suggests that intentional history is an inherent part of a stimulus' representation, and operationalized the intentional history of words by Google word co-occurrence counts. The fact that word co-occurrence counts could produce meaningful experimental results is a perfect illustration of the power of encoding context. People may not have a particular task when they read a word in everyday life, but there is always an encoding context in which the word appears, and that encoding context can decide how people intentionally process the words. This has implications for other psychological studies.

For example, Franks et al. (2000) investigated transfer between various judgments on words and speculated that characteristics of the words and/or the required judgment might cue subjects into making certain judgments on words even if the experimental instruction did not ask for such judgments. For instance, when presented with the word *diamond*, a subject very likely would think of its meaning as being *valuable*, even though the required study task was *strong/*weak judgment. Franks et al. referred to such unsolicited judgments as “automatically elicited intentional processes”. If the automatically elicited intentional processes in the study phase matched the required judgment task at test, repetition priming would be observed. The present work indicates that the intentional history of words may be the “characteristics of the words” that can interact with the encoding context and lead to unsolicited intentional processes. If a word has a high frequency or probability of being involved in an intentional act in the past, as is the case with *diamond* and *valuable* judgment, such habitual intentional process may be easily elicited, even if it is not required by the task at hand. Repetition priming due to the match between unsolicited intentional processes and test task should not be taken as evidence for transfer between the required study and test tasks. Transfer studies may need to take into account the intentional history of the stimuli when interpreting the results.

As another example, the famous Stroop phenomenon (Stroop, 1935) can also be interpreted from the intentional history perspective. In Stroop task, subjects try to name the color of the ink a word is written in instead of reading the word. This proves to be a difficult task, because it is difficult to inhibit the reading response and name the ink. From the intentional history perspective, word-reading is presumably extremely high in intention-specific frequency and probability, whereas color-naming is low in intention-specific frequency and probability. In accordance with the above mentioned automatically elicited intentional process idea, intentional acts with a high intention-specific frequency or probability may dominate over intentional acts with low intention-specific frequency and probability. Stroop-like phenomenon may be observed whenever there is a large imbalance between the intention-specific frequencies or probabilities of two intentional acts.

Summary

Building on previous works that have suggested learning and memory to be intention-specific (Bransford et al., 1979; Morris et al., 1977; Tulving, 1979; Tulving & Thomson, 1973), the present work applied the intention-specific principle to the instance theory of memory (Logan, 1988), and theorized that the intentional history of a stimulus is an inherent part of the stimulus representation. The present work defined the intentional history of a stimulus with the frequency, and the probability of an intentional act involving the stimulus. The effects of the intention-specific frequency and probability were tested and compared with the effects of word frequency in repetition priming, recognition, and recall tasks. The results supported the intentional history view of stimulus representations. The present work highlighted the intentional and episodic nature of memory representations in general, while at the same time suggested important differences in the modes of processing between implicit and explicit memory tasks. The difference between intention-specific frequency and intention-specific probability effects should present interesting clues and important constraints on computational models of memory processing.

REFERENCES

Blachman, N. (n.d.). Google guide: An interactive tutorial on making searching even easier. Retrieved April 16, 2004, from http://www.googleguide.com

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 657-668.

Bowers, J. S. (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin and Review, 7,* 83-99.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak, & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory* (pp. 331-355). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Craik, F. I. M., Moscovitch, M., & McDowd, J. M. (1994). Contributions of surface and conceptual information to performance on implicit and explicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 864-875.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science, 210,* 207-210.

Duchek, J. M., & Neely, J. H. (1989). A dissociative word-frequency × levels-of-processing interaction in episodic recognition and lexical decision tasks. *Memory & Cognition, 17,* 148-162.

Duncan, C. P. (1974). Retrieval of low-frequency words from mixed lists. *Bulletin of the Psychonomic Society, 4,* 137-138.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 680-698.

Fowler, Napps, & Feldman, (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory and Cognition, 13,* 241-255.

Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer-appropriate processing (TAP) and repetition priming. *Memory and Cognition, 28,* 1140-1151.

Gauthier, I., Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37,* 1673-1682.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1944 words. *Behavior Research Methods & Instrumentation*, 12, 395-427.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1-67.

Google Web APIs. Retrieved May 29, 2004, from http://www.google.com/apis/

Graf, P., Mandler, G., & Haden, P. E. (1982). Simulating amnesic symptoms in normal subjects. *Science, 218,* 1243-1244.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 501-518.

Graf, P., & Schacter, D. L. (1989). Unitization and grouping mediate dissociations in memory for new associations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 930-940.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Human, Learning & Memory, 10,* 164-178.

Gregg, V. H. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall & Recognition* (pp. 183-216). Oxford, England: John Wiley & Sons.

Gregg, V. H., Montgomery, D. C., & Castaño, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior, 19,* 240-245.

Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language, 37,* 502-516.

Intentional history of words (2004, March 18-20). Retrieved March 18-20, 2004, from http://www.google.com

Jacoby, L. L. (1983). Perceptual enhancement: persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9,* 21-38.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

Jacoby, L. L., Baker, J. G., & Brooks, L. R. (1989). Episodic effects on picture identification: Implications for theories of concept learning and theories of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 275-281.

Jacoby, L. L., & Hayman, C. A. (1987). Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13,* 456-463.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306-340.

Kinoshita, S. (1995).The word frequency effect in recognition memory versus repetition priming. *Memory & Cognition, 23,* 569-80.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of Human Memory*. New York: Academic Press.

Kirsner, K., Milech, D., & Stumpfel, P. (1986). Word and picture identification: Is representational parsimony possible? *Memory & Cognition, 14,* 398-408.

Kučera, H., & Francis, W. N. (1967). *Computational Analysis of Present-day American English.*

Providence, RI: Brown University Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211-240.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological-Review, 95,* 492-527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22,* 1-35.

Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition,20,* 1022-1050.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments and Computers, 28*, 203-208.

MacLeod, C. M. Word context during initial exposure influences degree of priming in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition,15,* 398-406.

MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition,22,* 132-142.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252-271.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-457.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114,* 159-188.

McKone, E. (1995). Short-term implicit memory for words and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition,21,* 1108-1126.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press/Taylor & Francis.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.

Moscovitch, M. (1994). Memory and working with memory: evaluation of a component process model and comparisons with other models. In D. L. Schacter & E. Tulving (Eds.), *Memory System 1994* (pp. 269-310). Cambridge, M.A.: The MIT Press.

MRC Psycholinguistic Database. Reretrieved August 28, 2004, from http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

Nadel, L. (1992). Multiple memory systems: what and why. *Journal of Cognitive Neuroscience, 4,* 179-188.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22,* 253-336.

Roediger, H. L. III (1990). Implicit memory: retention without remembering. *American Psychologist, 45,* 1043-1056.

Roediger, H. L. III, & McDermott, K. B. (1993). implicit memory in normal human subjects. In H. Spinnler, & F. Boller (Eds.), *Handbook of Neuropsychology, 8,* 63-131. Amsterdam: Elsevier.

Roediger, H. L. III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: a processing account. In H. L. Roediger, III, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: essays in honor of Endel Tulving* (pp. 3-41). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Roediger, H. L., Weldon, M. S., Stadler, M. L., & Riegler, G. L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory and Cognition,18,*1251-1269.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In S. Kornblum, & D. E. Meyer (Eds.), *Attention and Performance 14: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (pp. 3-30). Cambridge, MA, US: The MIT Press.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1-17.

Schacter, D. L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13,* 501-518.

Schacter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 432-444.

Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory System 1994* (pp. 1-38). Cambridge, M.A.: The MIT Press.

Shimamura, A. P. & Squire, L. R. (1984). Paired-associate learning and priming effects in amnesia: a neuropsychological study. *Journal of Experimental Psychology: General, 113,* 556-570.

Shimamura, A. P., & Squire, L. R. (1989). Impaired priming of new associations in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 4,* 721-728.

Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review, 94,* 439-454.

Squire, L. R. (1986). Mechanisms of memory. Science, 232, 1612-1619.

Squire, L. R. (1992). Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience, 4,* 232-243.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2,* 339-363.

Tenpenny, P. L., & Shoben, E. J. (1992). Component processes and the utility of the conceptually-driven/data-driven distinction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 25-42.

The American Heritage® Dictionary of the English Language (4th ed.). (2000). Boston, MA: Houghton Mifflin Company.

Thompson-Schill, S. L., & Gabrieli, J. D. E. (1999). Priming of visual and functional knowledge on a semantic classification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 41-53.

Tulving, E. (1979). Relation between encoding specificity and levels of processing. In L. S. Cermak, & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory* (pp. 405-428). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385-398.

Tulving, E., Schacter, D. L., & Stark, H. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 8,* 336-342.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,*359-380.

Versace, R., & Nevers, B. (2003). Word frequency effect on repetition priming as a function of prime duration and delay between the prime and the target. *British Journal of Psychology, 94,* 389-408.

Vriezen, E. R., Moscovitch, M., & Bellos, S. A. (1995). Priming effects in semantic classification tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 933-946.

Warrington, E. K., & Weiskrantz, L. (1978). Further analysis of the prior learning effect in amnesic patients. *Neuropsychologica 16,* 169-177.

Xiong, M. J., Franks, J. J., & Logan, G. D. (2003). Repetition priming mediated by task similarity in semantic classification. *Memory & Cognition, 31,* 1009-1020.