ABSTRACT

Title of dissertation: EXAMINING GROUP AND INDIVIDUAL APPROACHES TO STRATEGY USE IN FREQUENCY JUDGMENTS

Ana Maria Franco-Watkins, Doctor of Philosophy, 2004

Dissertation directed by: Assistant Professor Michael R. Dougherty, Department of Psychology

How do people make frequency judgments? One account describes how the strategy used during the memory retrieval process influences frequency judgments. The strategy that is evoked depends on the to-be judged event (Brown, 1995; 1997). Specifically, enumeration strategies were used when multiple exemplars were paired with a category. Non-enumeration strategies occurred when a single exemplar was paired with a category. Two experiments investigated the consistency with which different strategies were employed based on the characteristics of the to-be judged event. Group-level and individual-level analyses were used to examine the degree that a given strategy occurred. Analyses at the individual-level allowed for the investigation of whether the strategies observed at the group-level were consistent across individuals in the group. Exemplar typicality (typical or atypical) was manipulated within the to-be judged events to determine whether typicality played a role in strategy use and frequency judgments. Group-level analyses revealed evidence of enumeration in multiple-exemplar conditions. However, individual-level analyses revealed that few participants engaged in enumeration indicating that the group results were driven by a small number of participants. Exemplar typicality did not reliably affect strategy use or frequency judgments. Implications of using group-level and individual-level analyses are discussed.

EXAMINING GROUP AND INDIVIDUAL APPROACHES TO STRATEGY USE IN FREQUENCY JUDGMENTS

by

Ana Maria Franco-Watkins

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:

Assistant Professor Michael R. Dougherty, Chair
Associate Professor Kent L. Norman
Professor Edward F. Redish
Professor Harold Sigall
Professor Thomas S. Wallsten

Dedication

This work is dedicated to my parents, Norberto and Filomena Franco, whose continued sacrifice provided their children with greater opportunities.

Acknowledgements

I would like to thank my advisor, Mike Dougherty, for providing his guidance during the development of this line of research. I would also like to thank my committee members, Tom Wallsten, Kent Norman, Hal Sigall, and Joe Redish, for their insightful comments on prior drafts of this paper. Additionally, I would like to thank my husband, Larry Watkins, for his love and support throughout my graduate work. This material is based on work supported by the National Science Foundation under Grant SES-0134678 awarded to Dr. Michael Dougherty.

TABLE OF CONTENTS

List of Tables……………………………………………………………………..vi

List of Figures……………………………………………………………………vii

List of Figures……………………………………………………………………vii

Chapter I: Introduction……………………………………………………………….1
Brown's Experimental Paradigm……………………………………………4
Enumeration and Non-enumeration Strategies……………………………...5
Enumeration 5
Non-enumeration 6
Comparison of Enumeration and Familiarity Strategies 7
RT as a Measure to Assess Strategy Use……………………………………7
Estimation Strategies and Frequency Judgments……………………………8

Chapter II: Overview of Experiments and Research Variables……………………10
Group and Individual Approaches to Studying Strategy Use……………...11
Characteristics of the to-be Judged Event…………………………………12
Experimental Paradigm…………………………………………………… 15

Chapter III: Experiment 1…………………………………………………………..19
Experimental Hypotheses…………………………………………………19
RT 19
Frequency Judgments 20
Recall 22
Summary of Experimental Hypotheses 22
Method……………………………………………………………………….23
Participants 23
Materials 24
Design 25
Procedure 25
Results……..………………………………………………………………28
Recall 28
RTs 29
Frequency Judgments 36
Discussion….………………………………………………………………...41

Chapter IV: Experiment 2……………………………………………………………..44
Experimental Hypotheses…………………………………………………45
RT 45
Frequency Judgments 45
Recall 46
Method……………………………………………………………………….47
Participants 47

Materials 47
Design 47
Procedure 48
Results……..…………………………………………………………………49
Recall 49
RTs 51
Frequency Judgments 58
Discussion….…………………………………………………………...67

Chapter V: General Discussion……………………………………………………..69
Future Directions………………………………………………………..71
Implications……………………………………………………………..72
Appendix A……………………………………………………………..74
References………………………………………………………………80

List of Tables

Table 1: Mean Recall (Standard Error) in Different and Single-Exemplar Conditions in Experiment 1……………………………………………………….………………29

Table 2: Lower and Upper Quartile Frequency Judgments per Presentation frequency by Typicality and Exemplar presentation in Experiment ……………..………..…….39

Table 3: Mean Recall (Standard Error) in Different and Single-Exemplar Conditions in Experiment 2…………………………………………………………………….…49

Table 4: Lower and Upper Quartile Frequency Judgments per Presentation frequency by Condition Experiment 2………………………………………………………..63

List of Figures

Figure 1: Brown's Taxonomy of Strategies Used in Frequency Judgments………….2

Figure 2a: Average Median RT per Presentation frequency for Multiple-Exemplar Conditions in Experiment 1……………………………………………………….32

Figure 2b: Average Median RT per Presentation frequency for Single-Exemplar Conditions in Experiment 1……………………………………………………….32

Figure 3a: Average RT Regression Slopes in Experiment 1………………………...34

Figure 3b: Individual RT Regression Slopes per Condition in Experiment 1……….35

Figure 4a: Average Frequency Judgments per Presentation frequency in Multiple-Exemplar Conditions in Experiment 1……………………………………………38

Figure 4b: Average Frequency Estimates per Presentation frequency in Single-Exemplar Conditions in Experiment ...…………………………………………....38

Figure 5: Individual Frequency Slopes in Experiment 1…………………………..…41

Figure 6a: Average Median RT for Multiple-Exemplar-Bounded Condition in Experiment 2……………………………………………………………………..…53

Figure 6b: Average Median RT for Multiple-Exemplar-Unbounded Condition in Experiment. 2…………………………………………………………………………53

Figure 6c: Average Median RT for Single-Exemplar-Bound Condition in Experiment 2………………………………………………………………………….54

Figure 6d: Average Median RT for Single-Exemplar-Unbound Condition in Experiment 2 ……………………………………………………………………....54

Figure 7: Average RT Regression Slope per Condition in Experiment 2…………….56

Figure 8a: Individual RT Regression Slopes in Bounded Conditions in Experiment 2………………………………………………………………………….57

Figure 8b: Individual RT Regression Slopes in Unbounded Conditions in Experiment 2……………………………………………………………………..…57

List of Figures

Figure 9a: Average Frequency Judgments for Multiple-exemplar-Bounded Condition in Experiment 2…………………………………………………………………..…60

Figure 9b: Average Frequency Judgments for Multiple-Exemplar-Unbounded Condition in Experiment 2……………………………………………………….....60

Figure 9c: Average Frequency Judgments for Single-Exemplar-Bounded Condition in Experiment 2…………………………………………………………………..…61

Figure 9d: Average Frequency Judgments for Single-Exemplar-Unbounded Condition in Experiment 2…………………………………………………………………..…61

Figure 10a: Individual Frequency Slopes per Bounded Conditions In Experiment 2……………………………………………………………………….66

Figure 10a: Individual Frequency Slopes per Unbounded Conditions In Experiment 2……………………………………………………………………….66

Chapter I: Introduction

A widely debated topic in the memory and judgment literatures is how people make judgments of frequency. For many years, the prominent view held that frequency judgments were automatically encoded and that instructional or memorial factors should not affect frequency judgments (Hasher & Zacks, 1979). As an example, participants in Flexser & Bower's (1975) experiment were either informed or not informed they would be making frequency judgments after learning nonsense words. They found that frequency judgments did not differ between those informed and not informed about the frequency judgment task.

Although research has supported the automaticity hypothesis (Flexser & Bower, 1975; Howell, 1973; Hasher & Zacks, 1979; 1984; Zacks, Hasher, & Sanft, 1982), considerable research, including that which has examined the effect of encoding (Greene, 1986; 1988; Hanson & Hirst, 1988; Hintzman, 1988) and divided attention (Naveh-Benjamin & Jonides, 1986; Sanders, Gonzalez, Murphy, Liddle, & Vitina, 1987) on frequency judgments has called into question the automaticity hypothesis. For example, Greene (1988) demonstrated that frequency judgments improved under enhanced encoding conditions. Specifically, generation instructions resulted in a frequency judgment of 3.36 compared to 3.77 for the copy condition for words that appeared four times (Experiment 6, Green, 1988). Because factors that affect memory such as encoding and divided attention have been demonstrated to influence frequency judgments, several alternative theoretical accounts regarding the memorial basis of frequency judgments have been proposed.

One account developed by Brown (1995; 1997) describes how the strategy used during the memory retrieval process influences frequency judgments. From Brown's perspective, frequency judgments might differ depending on which strategy is employed. Figure 1 presents Brown's (1995; 2002) taxonomy of frequency estimation strategies. As can be noted in Figure 1, there are many strategies that can be used to assess frequency.

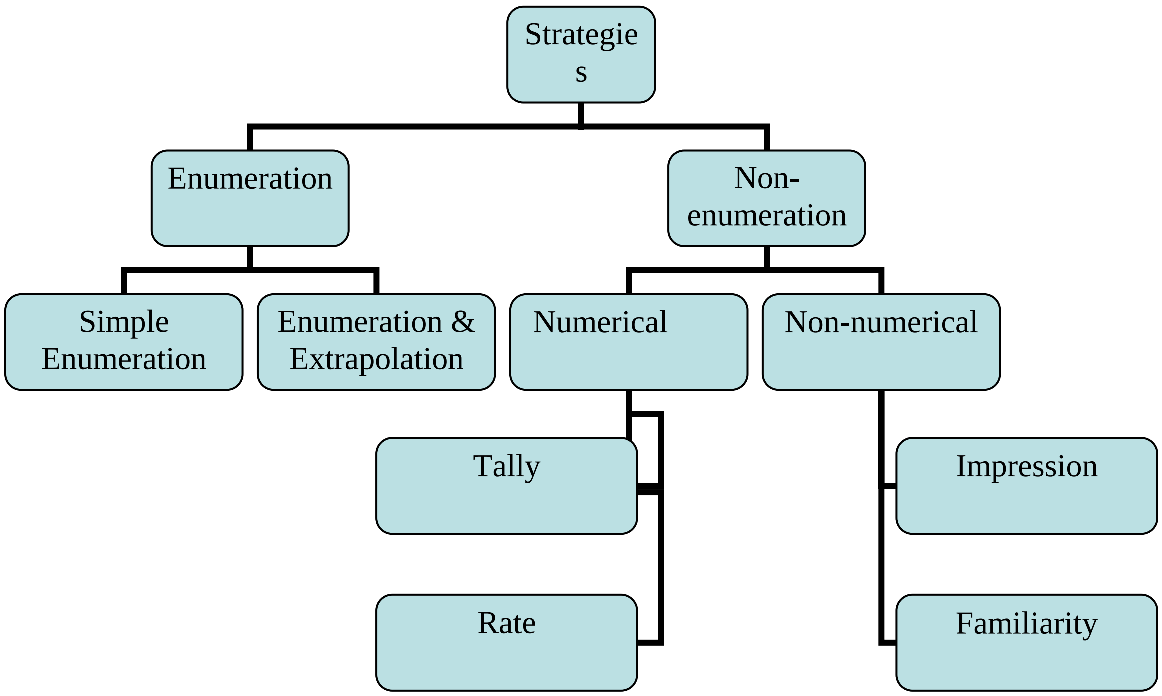

Figure 1: Brown's Taxonomy of Strategies Used in Frequency Judgments

Among the memory-based strategies, two have received great attention in the literature: enumeration and familiarity. As elaborated on below in greater detail, the primary difference between enumeration and familiarity strategies is in the type of retrieval process upon which each one is assumed to rely upon. Enumeration is assumed to be based on a recall-like process where specific examples are retrieved from memory whereas familiarity is assumed to be based on a recognition-like process that does not involve the recall of specific examples. As is commonly found in the memory literature for recall and recognition processes, enumeration-based and familiarity-based frequency judgments processes can often be distinguished on the bases of reaction time and accuracy measures. Because these measures have been used to asses strategy use, they are described in the sections outlining Brown's experimental paradigm and results.

Furthermore, the taxonomy of strategies presented in Figure 1 suggests at least two complications to studying frequency judgment processes. The first complication is that for any given task, participants can employ any number of strategies for deriving a frequency judgment. Thus, an important first step in studying frequency judgment processes is that one must understand the antecedent conditions that trigger the use of a particular strategy. To some extent, Brown's (1995; 1997) research lays the ground work for identifying some of these antecedent conditions. The second complication suggested by the taxonomy is that, again for any given task, the same antecedent conditions might trigger different frequency estimation strategies across participants. This later complication suggests that data analytic techniques used for determining strategy use need to be focused at the level of the individual participants.

The purpose of this paper is to determine the consistency with which different strategies are employed across participants within a given judgment task and to determine the degree to which different antecedent conditions trigger different strategies. It is common in psychological research to test a theory based on the data from a group of participants (i.e., using the overall group average). However, in many cases it is necessary to explore sources of variability at the level of individual participants. Analyses at the individual-level provide insight into the degree to which the group's performance is consistent across individuals versus the degree to which the overall group averages are driven by a small number of participants within the sample. Because several strategies can be used (as suggested by Brown's taxonomy and research), individual-level analysis seems particularly relevant for examining to what degree a particular strategy is used in frequency judgments. Although one might anticipate that certain strategies might be more common in some experimental conditions, rigorous tests have not been performed to examine the consistency with which particular experimental conditions evoke particular strategies across individuals. In the next section, Brown's experimental paradigm is presented as a judgment task that is presumed to evoke different strategies based on the characteristics of the to-be judged event.

Brown's Experimental Paradigm

In Brown's (1995) experiments, participants study a list of category-exemplar pairs (e.g., CITY-London) where categories are presented at frequencies of 0, 2, 4, 8, 12 or 16. In the single-exemplar condition, one exemplar was used consistently each time the category appeared. For example, London appeared each time the category city occurred during the study phase (e.g., CITY-London, CITY-London). In the multiple-exemplar condition, different exemplars were used each time the category occurred. For example, CITY-London, CITY-Paris, CITY-New York, and CITY-Boston were presented if the category CITY appeared four times. After studying the category-exemplar pairs, the category label alone (e.g., CITY) was presented during the test phase. Participants were asked to make frequency judgments regarding how often the *category* appeared during study phase. The use of single-exemplar and multiple-exemplar conditions allowed Brown to distinguish between two broad classes of estimation strategies: Enumeration and non-enumeration strategies.

Enumeration and Non-enumeration Strategies

Brown (1995; 2002) described various frequency estimation strategies. Although the taxonomy presented in Figure 1 is not exhaustive, it represents some of the basic strategies summarized by Brown. The strategies are classified into two general groupings: enumeration and non-enumeration.

Enumeration

Enumeration strategies involve the retrieval and counting of specific event occurrences. If asked to determine the number of times the category CITY appeared on the list, a simple enumeration strategy would be to recall and count the exemplars associated with the category CITY, and simply output the number counted as the estimate. For example, a participant using simple enumeration who recalled London, Paris, New York would output "three" as their estimate of how often CITY occurred. A variant to simple enumeration would be to extrapolate from the number of cities recalled. Extrapolation might occur when a participant realizes that more exemplars were associated with the category CITY than were recalled.[1] As a result of extrapolation, the frequency judgment does not equal to the number of exemplars

[1] The realization that more exemplars exist than the retrieved exemplars is similar to the metacognitve process of feeling-of-knowing (Hart, 1965).

retrieved, but rather is adjusted upward to account for the fact that the participant knows that he or she was unable to recall all of the exemplars that were presented.

Non-enumeration

Although there are several types of non-enumeration strategies (e.g., tally, rate, impression; see Brown, 2002 for a description of these), only the familiarity strategy is described here. Familiarity is expressed as frequency relevant information but in a non-numerical manner. Thus, familiarity does not necessitate direct access to numerical information. The single-exemplar condition in Brown's paradigm is one where the category CITY was paired multiple times with the same exemplar (e.g., London). Rather than retrieving and counting individual exemplars, one could presumably rely on the overall memory strength or feeling of familiarity associated with the category CITY or the word pair CITY-London. Because numerical information is not produced by familiarity, the feeling of familiarity has to be mapped onto a response scale before the participant can provide a numerical estimate. [2]

Familiarity is best exemplified by the global matching multiple-trace models of memory. An underlying assumption of these models is that each event is stored as a separate trace in memory. Judged frequency is based on the contribution of all traces in memory. These traces are not veridical copies of the event, but degraded representations of the event. A probe containing elements of the event is used to access memory. Traces most similar to the retrieval probe provide a larger

[2] A key difference between familiarity-based and enumeration-based frequency judgments is access to numerical information. Familiarity requires a mapping from a covert feeling to an overt response scale whereas enumeration already contains numerical information. This suggests that familiarity processes should be dependent on the properties of the response scale whereas enumeration processes should not. I address the role of response scales in frequency judgments in a subsequent section.

contribution to the overall familiarity. Traces dissimilar to the probe might be activated, but provide smaller contributions to overall familiarity. Hence, the greater the number of similar traces stored in memory, the larger the frequency estimate (Hintzman, 1988). Hence, if CITY-London occurred twelve times during the study list, then the frequency judgment would be greater than if CITY-London occurred four times.

Comparison of Enumeration and Familiarity Strategies

Enumeration (simple and extrapolation) are assumed to take more time than non-enumeration (e.g., familiarity) strategies. Brown (1995) assumed that enumeration involved a recall-like process in that enumeration entails the retrieval of multiple events from memory in a sequential manner. Thus, under enumeration, the total time to retrieve events increased as the number of recalled events increases. In contrast, the familiarity process was assumed to be similar in nature to a recognition process, which is based on an activation associated with probing memory that occurs in parallel. As a result, the assessment of familiarity is assumed to be a relatively fast process and the retrieval time remains constant regardless of the number of events contributing to the familiarity signal (cf., Hintzman, 1984, 1988). In the next section, I focus on the use of reaction time (RT) to assess which strategy was used for the frequency judgment.

RT as a Measure to Assess Strategy Use

Brown (1995; 1997) used RT and verbal reports to infer which strategy participants used. In Brown's (1995) Experiment 1, participants were asked to report their thoughts while making frequency judgments. The verbal reports revealed that 57% of the responses by the participants in the multiple-exemplar condition were classified as enumeration whereas enumeration was not used at all by participants in the single-exemplar condition. As noted earlier, an assumption of enumeration is that retrieval of events occurs in a sequential manner. Brown hypothesized that it should require more time to retrieve exemplars when people use enumeration. Consequently, as the presentation frequency of a category increases, so should the time to retrieve exemplars from that category. Indeed, Brown’s (1995) Experiments 2 and 3 demonstrated that RTs increased with presentation frequency in the multiple-exemplar condition. Brown interpreted the verbal reports and RT data as converging evidence for the occurrence of enumeration strategies under multiple-exemplar conditions. The multiple-exemplar condition is thought to be more memorably distinct because a person could presumably recall different exemplars associated with the category. The recall of different exemplars could be used as information when making frequency judgments.

In contrast to the multiple-exemplar condition, RTs suggested that enumeration strategies were not used for the single-exemplar condition. As stated earlier, the familiarity process is fast and all traces are assumed to be activated in parallel. As a result the RT slope should remain relatively flat regardless of the number of traces for a given event. In fact, Brown (1995) found that RT slopes remained relatively flat across all presentation frequencies. The single-exemplar condition is thought to be less memorably distinct because a single exemplar appeared with the categories at various frequencies. Although not impossible, it becomes more difficult to use the exemplars to infer frequency when the same exemplar is always paired with a category. In fact, Brown found evidence that frequency judgments differed between single and multiple-exemplar conditions.

Estimation Strategies and Frequency Judgments

In addition to examining strategy use within single and multiple-exemplars conditions, Brown (1995, 1997) demonstrated that frequency judgments were underestimated compared to the actual frequencies in the multiple-exemplar condition and were overestimated in the single-exemplar condition. Brown attributed this difference to differential use of the response scale between enumeration and non-enumeration strategies and to the possibility that people using enumeration fail to adequately adjust their judgment to account for non-recalled examples. Underestimation was presumed to occur because enumeration tends to lead to fewer items recalled than were presented and because the adjustment process failed to account for the number of non-recalled examples (cf. Tversky & Kahneman, 1974).

Differences between enumeration and non-enumeration processes lead to differences in how participants use the response scale. In the multiple-exemplar condition, the presence of an upper bound on the response scale should not influence frequency judgments. The lack of a response scale effect on frequency judgments is assumed to be due to the fact that enumeration strategies are based on numerical information (a count of recalled events). However, familiarity is presumed to be a non-numerical strategy such that the response scale provides a range on to which to map familiarity into a numerical response. Thus, manipulations of the response scale should be reflected in frequency judgments only when familiarity strategies are used.

In Brown's (1995) Experiment 3, the response scale's upper bound was manipulated across three levels: 16, 24 or no specified bound. Brown's results indicated that as the upper bound of the scale increased, so did frequency judgments in the single-exemplar condition (Brown, 1995). In particular, the bound of 16 resulted in underestimation compared to the presentation frequency. The bound of 24 resulted in a pattern mirroring the presentation frequencies. The unspecified upper bound resulted in overestimation of frequency judgments compared to the presentation frequencies. However, the multiple-exemplar conditions did not reveal differences in frequency judgments for the three levels of response scale. Frequency judgments were underestimated in comparison to the presentation frequencies regardless of the response scale used in the multiple-exemplar conditions.

Chapter II: Overview of Experiments and Research Variables

Brown argued that the to-be judged event affects strategy use. In particular, enumeration strategies are common when the event (i.e., category) is presented with multiple examples (i.e., multiple-exemplar condition). Non-enumeration strategies (e.g., familiarity) are common when the event is presented with similar examples (i.e., single-exemplar condition). Furthermore, strategy use determines whether or not the response scale bound influences frequency judgments. Brown argued that RTs provide a means of indirectly assessing strategy use. RT slopes increased as presentation frequencies increased when enumeration strategies were used. No increase in RT slope was found when a familiarity strategy was used.

In the next sections, I describe group and individual approaches to studying strategy use and how the characteristics of the to-be-judged event might also affect strategy use and frequency judgments. Next, an overview of experimental paradigm used in two experiments is described. Finally, I address how the current research differs from Brown's research.

Group and Individual Approaches to Studying Strategy Use

In the Brown (1995; 1997) experiments, RTs were averaged across individuals to produce mean RTs and RT slopes.[3] The assumption behind using mean RT to infer strategy use is that the mean reflects the majority of participants within a group.

Although group level analyses certainly provide insight into frequency judgment strategies, a more in-depth analysis can be achieved by examining how many individuals engage in a certain strategy. Brown's use of verbal protocols to

[3] It should be noted that median RTs also produced the same pattern of results.

identify individual differences in strategy use was a important first step, however, given the unreliability of verbal reports and questions about the validity of findings based on verbal reports (Nisbett & Wilson, 1977) a necessary second step is to examine individual differences behaviorally through the use of reaction time data.

One of the primary questions addressed in this paper is the degree to which analyses based on the individual participants support the conclusions drawn from an analysis of the group. Individual and group RT slopes between frequency judgments and presentation frequency are computed. Comparison of the group and individual RT slopes allows one to determine whether a few outliers or the majority of participants are contributing to the results. The individual analyses allow us to determine to what degree a specific strategy is evoked within an experimental condition. Thus, a strong test of the consistency with which a strategy is used is to examine strategy use at both the group-level and the individual-levels of analyses. If enumeration is common in the multiple-exemplar condition but not in the single-exemplar condition as indicated by Brown's results, then it should also be common at the individual-level of analysis.

In addition to the group and individual approaches to examining strategy use, the characteristics of the to-be judged event plays an integral role in which estimation strategy is likely to be used. The characteristics of the to-be judged event in Brown's research as well as specific factors that affect memory processes are described in the next section.

Characteristics of the to-be Judged Event

A critical characteristic of Brown's research was that the stimuli consisted of exemplars that represent a natural category. Unnatural categories (e.g., based on color or texture properties) did not result in use of enumeration strategies (Conrad, Brown, & Dashen, 2003). Additionally, enumeration strategies were not present when random word pairs were used (Brown, 1997; Dougherty & Franco-Watkins, 2003).

In memory research, it is well documented that characteristics of the stimuli affect encoding and retrieval processes. One characteristic that affects encoding quality is word concreteness. Encoding was enhanced for concrete words (e.g., chair) compared to abstract words (e.g. love). Pavio and Csapo (1969) postulated that concrete words can be encoded in terms of verbal and imaginable properties whereas abstract words can be encoded only in terms of their verbal properties.

A second characteristic is the frequency with which a word appears in the English language.[4] High-frequency words (e.g. hotel) and low-frequency words (e.g. atlas) have differential effects on recall and recognition processes. In recall, a greater number of high-frequency words were recalled compared to low-frequency words (cf., Clark & Gronlund, 1996). Conversely, low-frequency words were better recognized (higher hit rate and lower false alarm rate) than high-frequency words (Glanzer, Adams, Iverson, & Kim, 1993). It should be noted that the recall advantage for high-frequency words occurred on a consistent basis when the list consisted of only high or low-frequency words (i.e., pure list) but not when high and low-frequency words were mixed in the same list (Gillund & Shriffrin, 1984). Despite the latter result, the research suggests that manipulations of normative word frequency produced a dissociation between recognition and recall processes (Macleod & Kampe, 1996).

---

[4] Normative word frequency will be used from henceforth to describe the frequency with which a word or exemplar occurs in the English language and presentation frequency will be used to refer to the manipulated frequency within the experimental paradigm.

As noted earlier, the enumeration strategy is presumed to involve a recall process whereas the familiarity strategy is assumed to involve recognition process. The effects of normative word frequency on recall and recognition raise the interesting possibility that the frequency with which an exemplar occurs within a category might affect strategy use. One can examine the normative word frequency within a category by examining the number of times an exemplar was generated for a specific category. Typical and atypical exemplars can be inferred on the basis of the frequency of generation in category norms (e.g., Battig & Montague, 1969). Typical exemplars would be examples that were generated most often for a particular category and atypical exemplars would be the examples generated least often for a particular category.

If one assumes that normative word frequency applies to exemplar typicality, several questions can be asked within the strategy use framework. For example, given that low-frequency words are recalled less, will atypical exemplars decrease the ability to use enumeration and result in a greater reliance on a familiarity strategy in multiple-exemplar conditions? Because low-frequency words are recognized more often, will atypical exemplars result in greater frequency judgments than typical exemplars? Two experiments were designed to investigate the consistency of strategy use based on the characteristics of the to-be judged events: exemplar presentation (single or multiple) and exemplar typicality (atypical or typical). Before describing each experiment, a general overview of the experiments is provided in the next section. Greater detail of each experiment methodology is provided in the method section of the experiment.

Experimental Paradigm

The experimental paradigm was similar to Brown’s experimental paradigm whereby participants studied many category-exemplar pairs and then estimated the number of times that the category occurred during the study phase.

A total of 25 categories were used in the present experiments. Each category had a potential 12 exemplars that could be paired with the category during the study phase. Appendix A provides a complete list of categories and exemplars used in Experiments 1 and 2. The exemplars in the experiments differed on the basis of exemplar typicality with either typical or atypical exemplars used with each category. The Battig and Montague (1969) category word norm was used to determine exemplar typicality. Typical exemplars constituted the top 12 exemplars within each category. Atypical exemplars were those exemplars generated by less than 10 participants based on a total of 442 participants. Exemplar typicality was manipulated between participants. Similar to Brown’s experimental paradigm, single or multiple exemplars were paired with a category label. Exemplar presentation (single or multiple) was manipulated between participants.

Participants were presented with each category-exemplar pair on a computer screen for 3s. These pairs were associated with various presentation frequencies of 0, 2, 4, 8, or 12. Five categories were assigned to each level of presentation frequency. Every participant was exposed to all presentation frequencies, but participants studied the category-exemplar pairs in a random order such that each participant had a unique presentation.

Participants were assigned to one of four experimental conditions. In the typical-single-exemplar condition, one typical exemplar occurred with each repetition of the category (e.g., OCCUPATION-doctor). In the atypical-single-exemplar condition, one atypical exemplar occurred when the category was repeated (e.g., OCCUPATION-agent). In the typical-multiple-exemplar condition, multiple typical exemplars were paired with category. A different typical exemplar occurred each time the category was presented (e.g., OCCUPATION-doctor, OCCUPATION-lawyer, OCCUPATION-teacher, and OCCUPATION-dentist). In the atypical-multiple-exemplar condition, multiple atypical exemplars occurred with each category. A different atypical exemplar occurred with each repetition of the category (e.g., OCCUPATION-agent, OCCUPATION-umpire, OCCUPATION-butler, and OCCUPATION-miner).

After viewing all the category-exemplar pairs, a filler task was completed. The filler task consisted of 10 trivia questions that were not related to the studied category-exemplar pairs. In the test phase, the *category* label (e.g., OCCUPATION) appeared on the screen and participants estimated the number of times the category appeared during the study phase. Participants assessed the frequency judgments of each category in a random order. These frequency judgments were estimated using labeled keys on the computer keyboard (see method section of Experiment 1) or using the keypad to estimate a number on a box on the screen (see method section of Experiment 2).

Additionally, the scale provided to participants for their frequency judgments differed in the two experiments. In Experiment 1, participants received instructions that their frequency judgment should be a number from 0 to 24. In Experiment 2, the response scale's upper bound was manipulated. Some participants received instructions to respond using a response scale of 0 to 12 while others were given an open ended scale.

Participants recalled exemplars presented during the study phase in the final task of the experiment. This task occurred after all frequency judgments had been collected for all 25 categories. The category label appeared on the screen and the participant typed all the exemplars they could recall per category. Category labels were presented in a random order for the recall task.

Although I use Brown's experimental paradigm as the basis for the present experiments, the research outlined below differed from Brown's (1995; 1997) studies in several ways. First, strategy use was examined both at the group-level and individual- level to shed light on the consistency of strategy use across participants. Secondly, the typicality of the exemplar within the to-be-judged event was manipulated. Exemplar typicality was used to determine whether factors that differentially affect memory retrieval processes also influence strategy use and frequency judgments. Lastly, a correction for a potential methodological flaw in Brown's studies was implemented. Specifically, Brown failed to counterbalance categories across presentation frequencies. Some categories naturally contain more exemplars than other categories. A potential problem arises if the categories that contain more exemplars are used for larger presentation frequencies. It becomes difficult to discern whether frequency judgments are based on a certain categories containing more exemplars or based on presentation frequencies. In order to eliminate this problem, categories were counterbalanced across presentation frequencies in the two experiments presented in this paper. The next section presents Experiment 1.

Chapter III: Experiment 1

Experiment 1 involved the manipulation of exemplar typicality and exemplar presentation. If typicality affects retrieval processes in the same manner that normative word frequency affects retrieval, then a specific pattern of results are expected. These results are described below in terms of RT, frequency judgment, and exemplar recall.

Experimental Hypotheses

RT

Several assumptions must be stated in order to establish the pattern of results. If the multiple-exemplar condition leads to enumeration strategies, then enumeration will be present in both typical and atypical-multiple-exemplar conditions. However, participants in the typical–multiple-exemplar condition should be able to recall more exemplars than participants in the atypical-multiple-exemplar condition. An experiment conducted by Williams and Durso (1986) provided support for this prediction. Although, participants studied both typical and atypical exemplars, they recalled more typical than atypical exemplars.

Use of enumeration should be reflected in RTs. At the group-level, the average RTs per presentation frequency should result in an increased RTs as presentation frequency increases resulting in a positive RT slope. If participants in the atypical-multiple-exemplar condition are not able to enumerate as many exemplars as the typical-multiple-exemplar condition, then the corresponding RT slope should be shallower than the typical-multiple-exemplar condition. Notwithstanding, both slopes should demonstrate increased RT as presentation frequency increases. At the individual-level, the individual RT slopes should be larger in the typical-multiple-exemplar condition than the atypical-multiple-exemplar condition. Both multiple-exemplar conditions should reveal that most participants engaged in enumeration when compared single-exemplar conditions.

However, if participants in atypical-multiple-exemplar condition cannot enumerate due to the inability to retrieve atypical exemplars, then the RT slope should be relatively flat (reflecting use of a familiarity strategy). Additionally, the individual RT slopes should reflect the inability to use enumeration in the atypical-multiple-exemplar condition.

Participants in the single-exemplar conditions should use non-enumeration strategies as demonstrated in previous experiments conducted by Brown (1995). Typicality should not affect RTs between typical and atypical-single-exemplar conditions if the underlying retrieval process is familiarity. Because familiarity is assumed to be a relatively fast retrieval process the corresponding RT slopes for the single-exemplar conditions should be relatively flat and this pattern should hold at the individual and group levels.

The expected pattern in terms of frequency judgments is described next.

Frequency Judgments

Frequency judgments should be less than the actual presentation frequencies in both the typical and atypical-multiple-exemplar conditions (Brown, 1995; 1997). Brown assumed that underestimation is due to insufficient extrapolation in frequency judgments. Specifically, participants fail to sufficiently account for unrecalled exemplars. Additionally, if both conditions use enumeration, the typical condition will result in larger frequency judgments than the atypical condition. The difference between typical and atypical-multiple-exemplar conditions is based on the assumption that the atypical condition recalls fewer exemplars and does not sufficiently extrapolate for unrecalled exemplars. It is an open question whether frequency estimates should differ between the multiple-exemplar conditions if the atypical-multiple-exemplar does not use enumeration.

Brown (1995; 1997) demonstrated that non-enumeration strategies lead to overestimation of presentation frequency in the single-exemplar conditions. However, research has demonstrated that multiple-exemplar conditions result in underestimation of frequency. Hintzman and Stern (1978) demonstrated that names used in multiple sentences resulted in lower frequency judgments compared to names used in the same sentences. For example, the single-sentence condition judged the frequency to be about 6 and the multiple-sentence condition judged the frequency to be closer to 4.5 for names repeated six times. Thus, there is evidence that multiple-exemplar conditions should result in underestimation regardless of whether an enumeration or familiarity strategy is used.

Because only the category is presented at test, one of two situations could occur for the single-exemplar conditions. In one case, the category alone is used to probe memory. In the other case, the exemplar information is included with the category to probe memory.[5] It is possible that the exemplar information would be included because one exemplar occurred with each presentation of the category. The category should act as cue for a particular exemplar resulting in the recollection of the

[5] The different hypotheses arise based on whether the participants use item only information (e.g., category) or item-context information (e.g., category-exemplar).

category and exemplar and then both the category and exemplar would be used to probe memory.

There should be no difference between typical and atypical conditions in frequency judgments if the category alone is used to probe memory. This reflects the fact that the categories occur at the same presentation frequencies and the same response scale is used. However, if exemplar information is also included in the probe memory, then frequency judgments for the atypical condition should be larger than the typical condition demonstrating the advantage for low-frequency words found in normative word frequency effects. Normative frequency effects should also be found in exemplar recall.

Recall

In terms of the cued-recall data collected at the end of experiment, typical conditions should recall more exemplars than the atypical conditions. This would replicate the normative word frequency effects for recall where high-frequency words are recalled more often than low-frequency words.

Because of the number of experimental hypotheses, a summary of the expected pattern of results is provided in the next section.

Summary of Experiment 1 Experimental Hypotheses

If enumeration occurs in the multiple-exemplar conditions, both conditions should show an increased RT slope as presentation frequency increases. At the group-level analysis, the atypical condition's RT slope should be attenuated compared to the typical condition's slope. Investigation of the individual RT slopes ought to provide further evidence of whether or not enumeration occurs in the multiple-exemplar conditions. If enumeration occurs, individuals within the multiple-exemplar conditions should produce larger RT slopes compared to individuals in the single-exemplar conditions.

There should be no differences in RT slope for the single-exemplar conditions regardless of typicality. Group-level and individual-level analyses should reveal the same pattern of RT slopes for the single-exemplar conditions. The RT slope is presumed to be relatively flat for the single-exemplar conditions because familiarity is independent of the number of exemplars stored in memory.

In terms of frequency judgments, the multiple-exemplar conditions should underestimate frequency and single-exemplar conditions should overestimate frequency compared to the actual presentation frequencies. If the typical-multiple-exemplar condition exhibits an advantage over atypical-multiple-exemplar condition based on enumeration, then frequency judgments will be greater in the typical than the atypical condition. If non-enumeration is used in the atypical-multiple-exemplar condition, then it is uncertain whether differences in frequency judgments would emerge. A reversed pattern should occur in the single-exemplar conditions where atypical produce larger frequencies than the typical condition. This assumes that the exemplar is also used to probe memory, otherwise, there should be no difference based on typicality.

Method

Participants

One hundred and four University of Maryland students participated in this experiment as partial credit towards fulfillment of course requirements.

Materials

Four Pentium III (or greater) computers were used to present and collect all experimental material. DirectRT and MediaLab software was used to program the experiment. This software works in conjunction to allow for precise timing of material presentation as well as RT and data collection. Each computer was flanked by a partition in order to ensure that participants could not view any of the other computer screens except for the one presented in front of them.

Category names and exemplars were selected from the Battig and Montague (1969) norms. A complete list of the 25 categories and exemplars are presented in Appendix A. Additionally, the number of people (in Battig & Montague, 1969) who generated the exemplar is provided to the right of the exemplar. Twelve typical and 12 atypical exemplars were assigned to each category. Typical exemplars consisted of the top 12 exemplars generated most often in a category (with a few exceptions where 12 exemplars from the top 15 were used). Atypical exemplars consisted of exemplars that less than 10 respondents generated. In the Batting and Montague (1969) non-exemplars were included in the list. For example, the word “criminal” would be a non-exemplar for the category fruit. Only exemplars representing the specified category were included. Additionally, items that were more familiar to students because the norms were constructed were not included as an atypical exemplar. For example, Afghanistan was considered as an atypical exemplar for the category country in 1969 norms. In light of recent world events, Afghanistan would not be considered atypical.

Category-exemplar norms were cross-validated with the Kucera and Francis (1967) normative word frequency norm. The cross-validation was used to signify that atypical and typical exemplars correspond to low and high-frequency words. Using the exemplars presented in Appendix A, typical and atypical exemplars were examined in terms of normative word. Typical exemplars occurred at an average of 43.7 times per million (range of 1-431; with the exception of one exemplar at 569). Atypical exemplars occurred at an average of 10.8 per million (range of 1-72; with two exceptions: one exemplar occurred at 123 and another at 371). Hence, the cross-validation results indicated that the typicality of an exemplar is interrelated with its normative word frequency.

Five categories were assigned to each of the five levels of presentation frequency. The study list consisted of a total of 130 category-exemplar pairs. Category names were displayed in upper case letters in the middle of a computer screen and exemplar names were displayed two lines below the category name in lower case letters. A font size of 28 point was used for both items with each category-exemplar pair displayed on the screen for 4s.

Design

The experiment consisted of a 2 (exemplar typicality: typical or atypical) x 2 (exemplar presentation: single or multiple) x 5 (presentation frequency: 0, 2, 4, 8, and 12) mixed design. The between participant variables were exemplar typicality and exemplar presentation resulting in four experimental conditions. Presentation frequency was manipulated within participants.

Procedure

Twenty-six participants were randomly assigned to each of the four experimental conditions. Participants were seated in front of computer. All instructions and experimental stimuli were presented on the computer screen. Additionally, all responses and Participants were instructed that the study phase would consist of a series of word pairs based on a category and an exemplar (e.g., MONEY-dollar) and that their memory for these words would be tested later in the experiment. Instructions did not include information regarding the nature of the frequency judgment task in order to ensure learning of both words and to prevent participants from trying to keep a tally of frequencies per category. Each participant received a random presentation of the category-exemplar pairs.

In the typical-single-exemplar condition, one typical exemplar was always displayed each time it occurred during the study phase (e.g., OCCUPATION-doctor). In the atypical-single-exemplar condition, one atypical exemplar occurred each time the category was presented (e.g., OCCUPATION-butler). In the typical-multiple-exemplar condition, different typical exemplars occurred each time the category was presented on the screen (e.g., OCCUPATION-doctor, OCCUPATION-lawyer, and OCCUPATION-teacher). In the atypical-multiple-exemplar condition, different atypical exemplars occurred with each presentation of the category (e.g., OCCUPATION-butler, OCCUPATION-agent, and OCCUPATION-umpire). Furthermore, each category was counterbalanced such that it was equally likely to occur at each level of presentation frequency. Five categories were assigned to each level of presentation frequency. Thus, participants experienced 20 categories during the study phase: five unique categories were presented at frequencies of 2, 4, 8 and 12. Additionally, 5 categories were assigned to frequency 0 which the participant did not experience during the study phase.

Participants completed a brief filler task consisting of 10 multiple choice trivia questions between the study and test phases in order to ensure that most recent category-exemplar pairs were displaced from working memory. The test phase consisted of the 20 categories presented during the study phase as well as 5 categories assigned to frequency 0. In addition, 3 practice categories were included in the test phase in order for participants to acclimate to judgment task procedure.

Before the test phase began, participants were instructed that:

> In this part of the task, you will be presented with 28 category names. You must estimate as *accurately* as possible the number of times that each category name appeared in the list that you studied. Although we will keep track of how long you take to make your frequency estimation, we encourage you to be as accurate as possible.
>
> You will make your estimation in the following manner. You will see the category word appear on the screen. You will press the space bar as soon as a *single* numerical response comes to mind, but not before you have a number in mind. After pressing the space bar, you will use the blue numbered keys for your frequency estimation.

The initiation of a trial began by pressing any key on the keyboard to see the category name, and then the category name was displayed as well as a statement indicating that they should press the space bar when they have a number in mind. The next screen displayed the category label again as well as a statement to enter a numerical response. Blue keys on the keyboard were numbered 0 to 24. The top row of the keyboard from the 1 key to the + key were labeled from 1 to 12 consecutively. The second row of keys on the keyboard from the Q key to } key were labeled from 13 to 24. Numerical responses were entered by pressing the key corresponding to their frequency judgment. Participants estimated the frequency of each category in a random order. Between category frequency judgments, a statement on the screen indicated that to press any key to see the next category. A cued-recall task occurred immediately after all frequency judgments for the 25 categories were collected. Using the category label as the cue, participants were instructed to recall as many exemplars as they could remember from the study list.

Results

The data were analyzed in terms of RTs, frequency judgments, and exemplar recall. Because the recall data was used as a manipulation check for the exemplar typicality manipulation, it is presented first.

Recall

Table 1 presents the average number of exemplars recalled at each level of presentation frequency per condition. The 0 frequency condition was not included because participants could not recall correct exemplars for this frequency. Because the single-exemplar conditions could recall one exemplar, analyses were conducted within multiple and single-exemplars conditions but not between them.

In the multiple-exemplar conditions, typicality interacted with presentation frequency, F (3, 200) = 15.08, $p < .01$, $\hat{f}$ = .46.[6] As can be noted on Table 3, the typical condition retrieved more exemplars than the atypical condition and this difference was magnified as the presentation frequency increased. Using a Bonferroni alpha adjustment (alpha = .0125), all paired comparisons at each level of presentation frequency between typical and atypical conditions were significant. A main effect was present for exemplar typicality F (1, 200) = 26.28, $p < .01$, $\hat{f}$ = .37. The typical condition generated more exemplars than the atypical conditions. The typical

---

[6] $\hat{f}$ refers to sample estimate effect size (Kirk, 1995)

condition recalled 36% of total exemplars and the atypical condition recalled 20% of total exemplars. The recall for the typical condition is similar to 40 % recall reported by Brown (1995). The recall data also serves a manipulation check that typical exemplars are better recalled than atypical exemplars. This recall advantage is similar to normative word frequency effects observed in recall processes where high-frequency words are better recalled than low-frequency words. Additionally, there was a main effect of presentation frequency F (3, 200) = 165.18, $p < .01$, $\hat{f}$ = .87. The main effect reflects that fact that more exemplars are available for recall as presentation frequency increases and indeed more exemplars were recalled.

The above recall pattern was not observed in the single-exemplar condition. Only the main effect for presentation frequency was significant, F (3, 200) = 15.64, $p < .01$, $\hat{f}$ = .51. Additionally, comparisons within each level of frequency were not significant between typical and atypical conditions. The total number of exemplars recalled was similar for both conditions with atypical recalling 81% of exemplars and typical recalling 87% of exemplars.

Table 1: Mean Recall (Standard Error) for Multiple and Single-Exemplar Conditions in Experiment 1

Condition	2	4	8	12
Multiple-exemplar				
Typical	0.92 (.08)	1.73 (.15)	2.86 (.23)	4.02 (.30)
Atypical	0.45 (.06)	0.96 (.12)	1.77 (.17)	2.10 (.16)
Single-exemplar				
Typical	0.74 (.05)	0.89 (.03)	0.93 (.03)	0.93 (.03)
Atypical	0.66 (.08)	0.87 (.04)	0.85 (.04)	0.85 (.04)

RTs

In accordance with Brown, RTs were used to infer the type of strategy used. Support for enumeration is based on increased RTs as the presentation frequency increased. Relatively flat RTs across presentation frequencies support a familiarity strategy.

Analyses of RTs were based upon the aggregate of the RT for the time it took to press the space bar when a number came to mind and the RT for the time to press the key corresponding to their frequency judgment per category. Ratcliff (1993) outlined several methods that can be used to eliminate RT outliers. Outliers normally create additional nuisance variance; however, in the present set of studies, outliers are of interest because they potentially indicate the use of enumeration. Nonetheless, in order to reduce any potential skewed RT effects within any one level of frequency, analyses were based on each participant's median RT per category frequency rather than mean RTs. [7]

Group Median RTs. Figure 2a shows the average median RT across presentation frequencies for multiple-exemplar conditions and Figure 2b shows the average median RT for single-exemplar conditions. A main effect of presentation frequency occurred, F (4, 500) = 16.97, $p < .01$, $\hat{f} = .29$, indicating that RTs increased as presentation frequencies increased. This is reflected in the presence of three two-way interactions. A presentation frequency by exemplar typicality interaction was present, F (4, 500) = 3.27, $p = .01$, $\hat{f} = .11$. A presentation frequency by exemplar presentation interaction was also present. F (4, 500) = 3.35, $p = .01$, $\hat{f} = .13$. Also, an

[7] The use of median RTs was suggested by Tom Wallsten.

exemplar typicality by exemplar presentation interaction, F (4, 500) = 4.92, $p < .05$, $f^{\wedge}$= .09.

As can be noted from Figures 2a and 2b, the pattern differed in the multiple-exemplar between typical and atypical conditions with the typical condition producing larger RTs than the atypical condition; however, the pattern was similar for both typical and atypical conditions in the single-exemplar conditions. Simple main effects of this interaction revealed the typical-exemplar conditions differed from the atypical-exemplar conditions at median RT of frequency 12, F (1, 500) = 8.02, $p < .01$. The presentation frequency interactions are most likely due to the contribution of the typical-multiple-exemplar condition where RT increased across objective frequency whereas the other three conditions revealed similar patterns to one another with relatively flat RTs. However, the three-way interaction was not present.[8]

The RT results were not robust to transformations of RTs. Using a log transformation resulted in main effect of presentation frequency, F (4, 500) = 28.30, $p < .01$, $f^{\wedge}$= .42 and the interaction of exemplar typicality by exemplar presentation was close to significance, F (1, 500) = 3.48, $p = .07$, $f^{\wedge}$= .07. In contrast, using 1/RT to transform the RTs resulted in only a main effect of presentation frequency, F (4, 500) = 34.79, $p < .01$, $f^{\wedge}$= .45. Thus, transforming the RT data resulted in a loss in the number of significant effects which calls into question the ability to interpret the overall pattern of means.

[8]Similar results occurred when mean RTs were analyzed with the following exceptions: A significant three way interaction for presentation frequency by exemplar typicality by exemplar presentation, F (4, 400) =2.37 p =.05 was present. The p-value for the interaction of presentation frequency by exemplar typicality was slightly above .05 (p = .08).

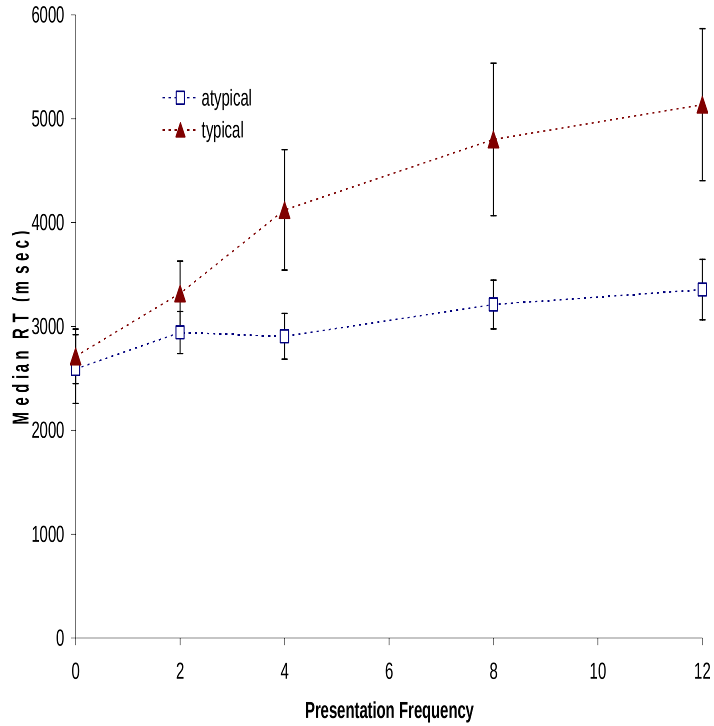

Figure 2a: Average Median RT per Presentation Frequency for Multiple-Exemplar Conditions in Experiment 1

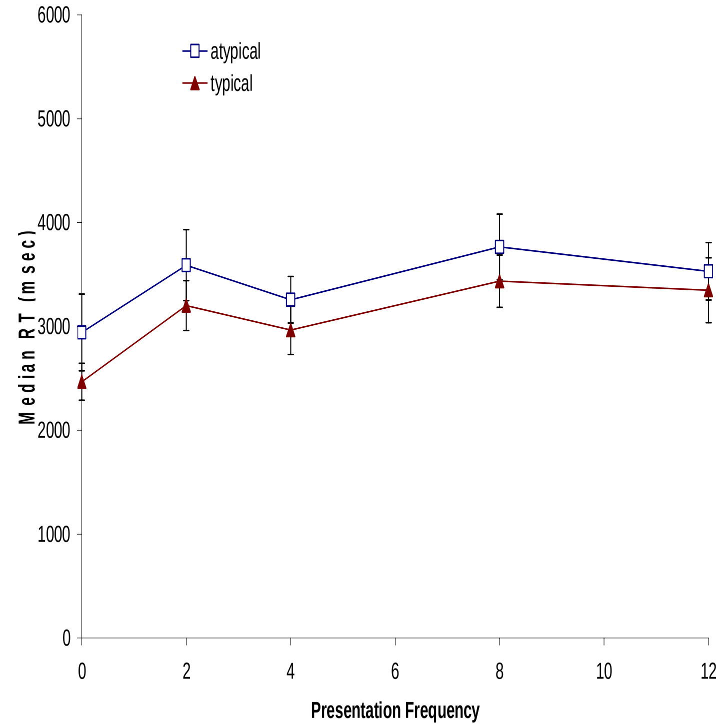

Figure 2b: Average Median RT per Presentation Frequency for Single-Exemplar Conditions in Experiment 1

Group Median RT Slopes. In addition to the median RTs, regression slopes were computed for each participant in each of the four experimental conditions. Individual regression slopes were computed by using each participant's median RT at each level of presentation frequency. Group regression slopes were based on the mean slope of all the individual slopes within an experimental condition. An ANOVA conducted on the group-level regression slopes revealed an exemplar presentation by exemplar typicality interaction, F (1, 100) = 4.06, $p < .01$, $\hat{f}$ = .17. Figure 3a presents the average median slope per condition. As can be noted from the figure, the slopes are similar for the two single-exemplars conditions, but this pattern was not replicated in the multiple-exemplar conditions. Additionally, there were two main effects: exemplar typicality F (1, 100) = 7.23, $p < .01$, $\hat{f}$ = .24, and exemplar presentation, F (1, 100) = 6.62, $p = .01$, $\hat{f}$ = .23. Typical exemplars resulted in larger slopes than atypical exemplars. Multiple-exemplar conditions also produced larger slopes than single-exemplar conditions. The group median RT slope pattern is similar to the results obtained using the median RTs as a repeated measure across presentation frequency.

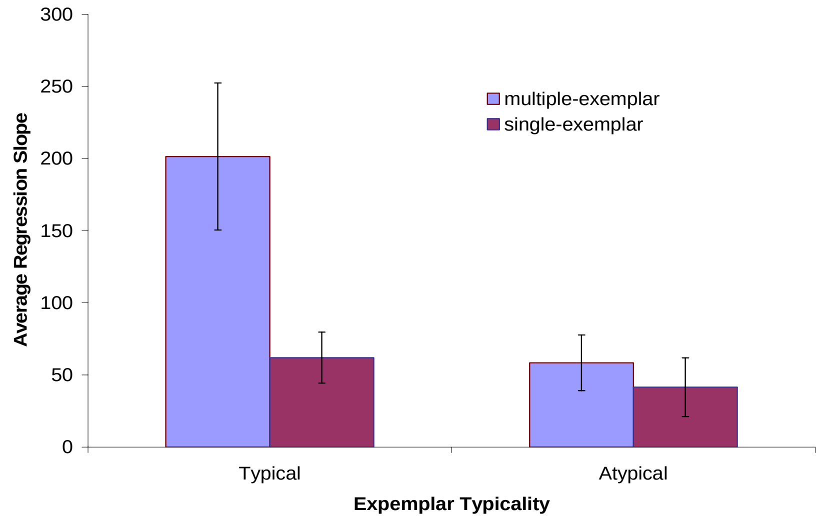

Figure 3a: Average RT Regression Slopes in Experiment 1

Individual Median Slopes. Although the group analyses indicated enumeration in the typical-multiple-exemplar condition, examining the distribution of slopes can determine to what degree the enumeration strategy occurred within and across conditions. Figure 3b presents distribution of slopes per condition. Each point on the graph represents an individual's slope with overlapping slopes indicated by darker areas. If one assumes that participants in the single-exemplar conditions use familiarity, then using the mean of all single-exemplar conditions provides a baseline for comparison of familiarity to enumeration strategies. Specifically, to the degree there are significant deviations from this average, these deviations would indicate individuals who are using an enumeration strategy. Consequently, of interest are the outliers who are presented two standard deviations (SD) above the mean of the single-exemplar conditions. The numbers to the far right of the figure represent the number of people whose RT slope were 2 SD above the mean.

In all, 12 individuals were 2 SD above the mean, with 9 of those participants in the typical-multiple-exemplar condition. As can be noted from the graph, the typical-multiple-exemplar condition had more individuals whose slope was 2 SD above the mean than any of the other conditions. These slopes indicate that RT increased as the presentation frequency increased. Hence, 35% of participants in the typical-multiple-exemplar condition engaged in an enumeration strategy. Fisher's exact test on the number of outliers in each condition revealed an effect of exemplar presentation ($p <$ .05), but not an affect of typicality (*ns*). Thus, some participants appear to enumerate and this occurs most often under multiple-exemplar conditions. Yet, the majority of participants appear to use non-enumeration strategies even in the multiple-exemplar conditions. This result suggests that the group-level analysis supporting the use of enumeration in the typical-multiple-condition was likely due to a small number of participants.

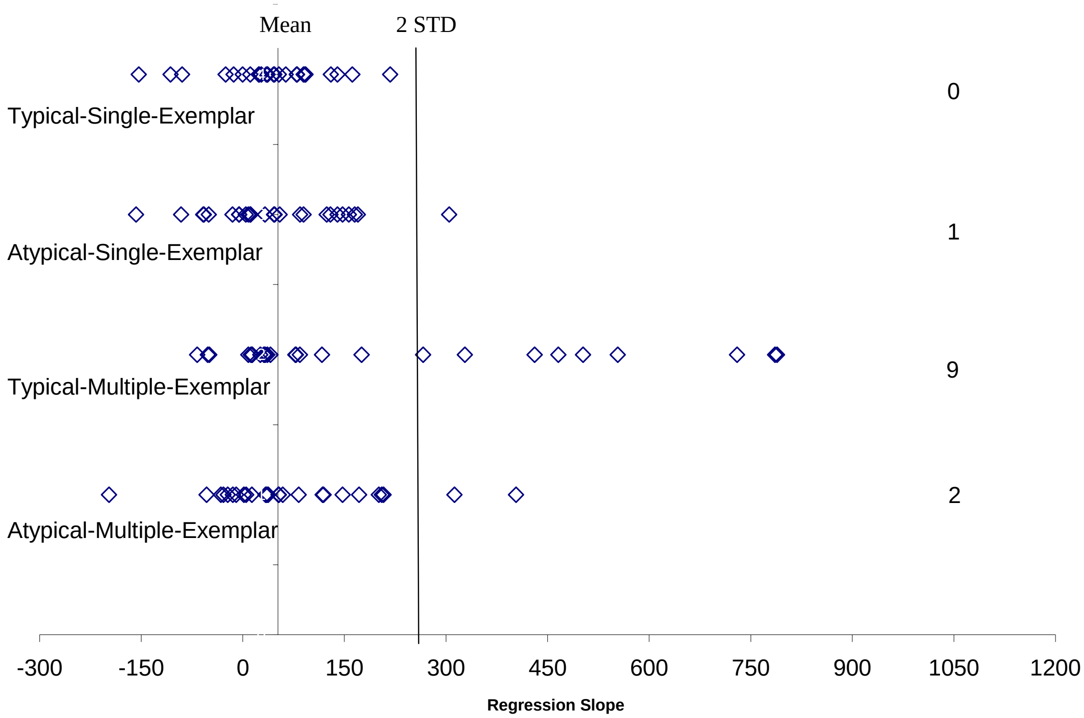

Figure 3b: Individual RT Regression Slopes per Condition in Experiment 1

Frequency Judgments

All frequency judgments were based on the average of 5 judgments per category frequency per participant.

Average Frequency Judgments. Consistent with Brown's (1995) findings, a presentation frequency by exemplar presentation interaction was present, F (4, 500) =12.16, $p < .01$, $\hat{f}$ = .28. Figure 4a depicts the frequency judgments for multiple-exemplar conditions and Figure 4b depicts the single-exemplar conditions. As can be noted in the figures, frequency judgments increased as a function of presentation frequency more steeply in the single-exemplar than the multiple-exemplar conditions. Typicality did not interact with exemplar presentation or presentation frequency and no other interactions were present. The single-exemplar conditions overestimated lower presentation frequencies, but underestimated the larger presentation frequencies. The multiple-exemplar conditions tended to underestimate most of the presentation frequencies. In the Brown studies, the single-exemplar condition tended to overestimate frequencies for all levels of presentation frequency. Perhaps the reason for the lack of overestimation at the larger presentation frequencies was due to the upper response scale bound of 24. Although this upper bound was twice the amount of the largest actual presentation frequency, it might have resulted in smaller frequency judgments.

Analyses also revealed a main effect of presentation frequency, F (4, 500) = 253.96, $p < .01$, $\hat{f}$ = .81 as well as a main effect of exemplar presentation, F (1, 500) = 40.98, $p < .01$, $\hat{f}$ = .58.These results are similar to those found by Brown (1995) demonstrating sensitivity to presentation frequency whereby frequency judgments increased as presentation frequency increased. Additionally, single-exemplar resulted in larger frequency estimations than the multiple-exemplar. Exemplar typicality did not affect frequency judgments. [9]

[9]In addition, the results were identical using median frequencies at each level of presentation frequency.

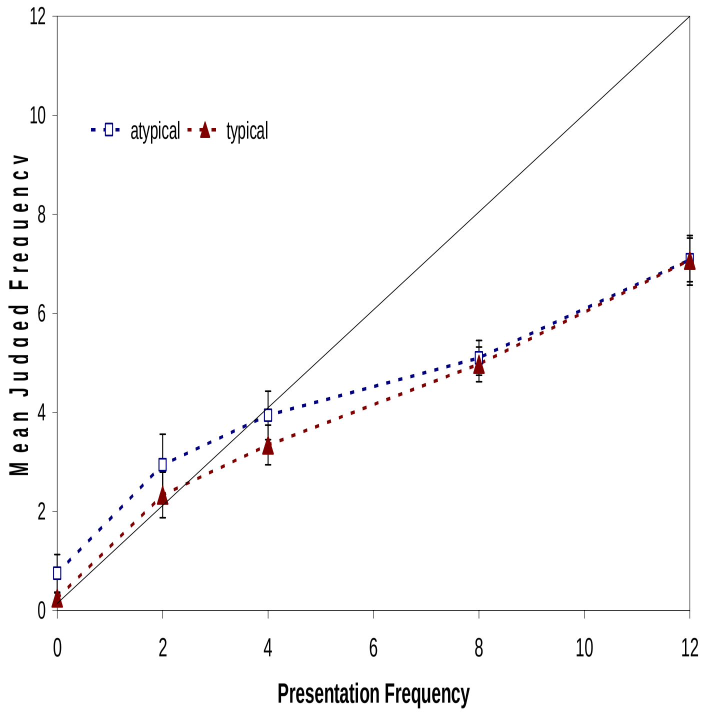

Figure 4a: Average Frequency Judgments per Presentation frequency in Multiple-Exemplar Conditions in Experiment 1

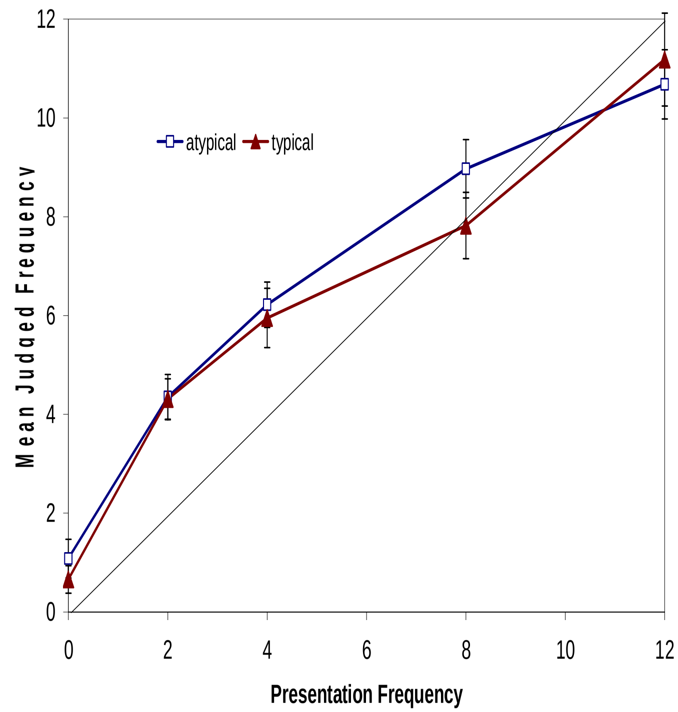

Figure 4b: Average Frequency Judgments per Presentation frequency in Single-Exemplar Conditions in Experiment 1

The range of frequency judgments provides an indication of distribution and use of the response scale. To illustrate the range of frequency judgments, Table 2 presents the upper and lower quartiles of frequency judgment at each level of presentation frequency by condition. Quartiles are presented as a description of range used within each condition.

Table 2: Lower and Upper Quartile Frequency Judgments for Presentation frequencies by Typicality and Exemplar Presentation in Experiment 1

Objective	Typical-Multiple		Atypical-Multiple		Typical-Single		Atypical-Single	
Frequency	Q1	Q3	Q1	Q3	Q1	Q3	Q1	Q3
0	0	0.2	0	0.4	0	0.6	0	1.4
2	1.4	2.6	1.2	2.8	2.4	5.4	2.8	5.8
4	2.2	3.8	2.4	4.2	3.6	8.0	4.2	8.0
8	3.8	6.4	3.6	6.6	5.6	11.2	6.2	11.2
12	5.0	9.0	5.2	8.6	7.8	15.4	8.2	13.2

Inspection of Table 2 indicates similar estimates for frequency 0 across conditions (Kruskal-Wallis, $p > .05$). However, the range of frequency judgments differed cross conditions for presentation frequencies 2, 4, 8, and 12, $\chi^2(3) = 30.96$, $\chi^2(3) = 30.00$, $\chi^2(3) = 32.42$, and $\chi^2(3) = 23.34$, *p's* < .01, respectively. Wilcox rank-sum analyses revealed the same pattern between single-exemplar and multiple-exemplar conditions for frequencies 2, 4, 8, and 12 (*p's* < .01). As can be noted from Table 2, multiple-exemplar conditions tended to use smaller estimates than the single-exemplar conditions. Typicality did not influence the range of frequency judgments.

Inspection of Table 2 reveals that ranges were similar across levels of frequency for typical and atypical conditions.

Group Median Frequency Slopes. Regression slopes predicting presentation frequency from frequency judgments were computed for each participant based on the median frequency judgment per presentation frequency. Using the average median slope per condition, ANOVA analyses revealed a main effect of exemplar presentation where multiple-exemplar slopes (*M* = .51, *STD* = .25) were smaller than single-exemplar slopes (*M* = .80, *STD* =. 40), *F* (1, 100) = 19.49, *p* < .01, $\hat{f}$ = .39. Exemplar typicality did not affect frequency slopes. Brown (1995) also demonstrated that frequency slopes were smaller in the multiple-exemplar condition compared to the single-exemplar condition.

Individual Median Frequency Slopes. Figure 5 presents the distribution of individual frequency slopes per participant in each of the conditions. The line on the graph represents a slope of 1.0 which would be perfect resolution between presentation frequency and frequency judgments. As can be noted from the graph, several participants in single-exemplar conditions had slopes above 1.0 whereas only one person had a slope greater than 1 in the multiple-exemplar conditions. Although the group level analyses indicate overestimation in the single-exemplar conditions, most slopes are below the slope of 1.0.

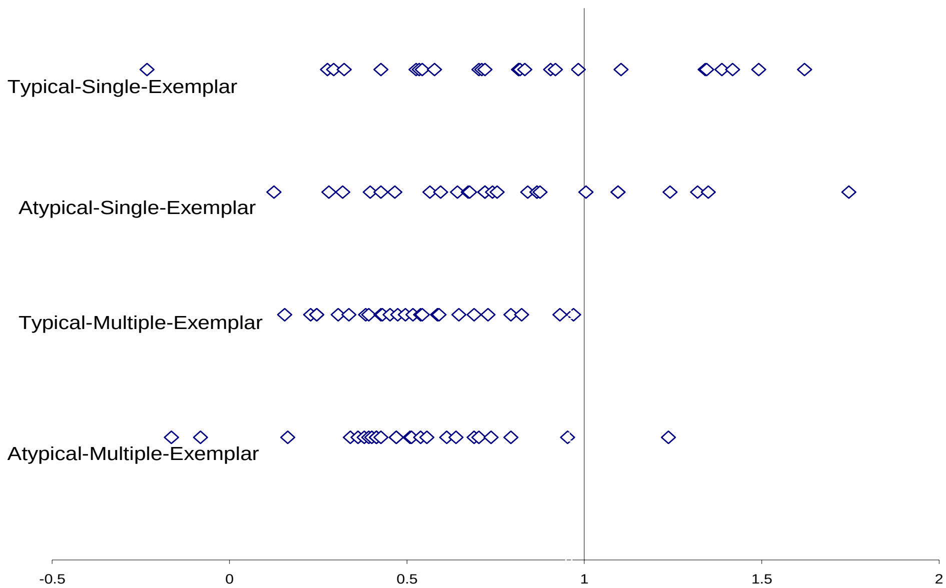

Figure 5: Individual Frequency Slopes in Experiment 1

Discussion

The group-level RT data indicated that an enumeration strategy was used in multiple-exemplar conditions. Exemplar typicality interacted with exemplar presentation. Further analyses revealed that multiple-exemplar conditions differed in terms of typicality with larger slopes in the typical than atypical-multiple-exemplar conditions.

However, the results at the individual-level were discordant with the group-level results. Although evidence of an enumeration strategy was observed in the multiple-exemplar conditions, very few participants engaged in enumeration. In fact, the shallower slope for atypical compared to typical-multiple-exemplar condition was most likely due a slightly larger number of participants using enumeration in the typical compared to atypical multiple-exemplar conditions.

The recall data supported the advantage of typicality in the multiple-exemplar condition where typical exemplars were recalled more often than atypical exemplars. Additionally, the number of exemplars recalled increased with presentation frequency.

Interestingly, similar frequency judgments were produced in the multiple-exemplar condition despite different RT slopes for typical and atypical conditions. Investigation of the individual frequency slopes revealed that these conditions provided similar frequency slopes.

In Experiment 1, typicality did not affect frequency judgments within each of the exemplar presentation conditions. In the single-exemplar conditions, we did not see the advantage for the atypical versus typical frequency judgments. The lower and upper quartiles of the frequency responses as well as rank sum analyses indicated no difference in scale use based on typicality. However, it is possible that the upper bound of the scale affected response ranges such that the atypical advantage was not observed. Specifically, because the response scale was identical, the mapping of familiarity to response scale could have restricted any advantage of atypical exemplars. This is assuming that there are small differences in familiarity for typical and atypical conditions.

Brown (1997) conceded that enumeration does not occur for all category trials in the multiple-exemplar condition. However, he assumed that people who enumerate for some trials should also tend to underestimate frequency judgments for non-enumeration trials. He posited that the range generated by enumeration trials is used to infer frequency in the non-enumeration trials. As a consequence, the response range should be smaller in conditions that use enumeration even if enumeration is used only some of the time. Indeed, the lower and upper quartiles of frequency responses were less in the multiple-exemplar than in the single-exemplar conditions.

Chapter IV: Experiment 2

Although the RT slopes at the individual level indicated that few participants engaged in enumeration, Experiment 2 was conducted to determine whether this pattern could be replicated. The group-level analyses indicated enumeration in the multiple-exemplar conditions (particularly the typical conditions). However, the individual-level analyses revealed that this pattern was due a smaller number of participants.

Additionally, Experiment 2 extends upon Experiment 1 to include a manipulation of response scale used in deriving frequency judgments. Specifically, the upper bound of the scale was set at 12 or no upper bound was imposed. In the unbounded condition, participant could use any number when making frequency judgments. The upper bound of 12 was selected to exacerbate any potential typicality differences that might have been attenuated in the previous experiment. This bound reflects the true upper bound of presentation frequency.

Experiment 2 consisted of 8 experimental conditions. Exemplar presentation (single or multiple), exemplar typicality (typical or atypical), and upper response scale (bounded or unbound) were manipulated between participants. As in Experiment 1, presentation frequency was manipulated within participants. The study phase of the experiment was identical to Experiment 1. The test phase differs from Experiment 1 in that the upper bound of the scale was manipulated. The experimental conditions reflect a three-way factorial design of exemplar typicality by exemplar presentation by scale. This resulted in eight conditions: typical-multiple-bounded, typical-multiple-unbounded, atypical-multiple-bounded, atypical-multiple-unbounded, typical-single-bounded, typical-single-unbounded, atypical-single-bounded, and atypical-single-unbounded.

Experimental Hypotheses

RT

As in Experiment 1, the group RT slopes in the multiple-exemplar condition should reveal enumeration, particularly for the typical condition. The response scale should not affect RTs. Individual RT slopes should demonstrate a similar pattern to Experiment 1 where enumeration occurs in the multiple-exemplar regardless of typicality, but overall, few participants should enumerate. For the single-exemplar conditions, group level RTs should be relatively flat across presentation frequencies as demonstrated in Experiment 1. The group and individual-level of analyses should reveal the same pattern for single-exemplar conditions.

Frequency Judgments

In experiments conducted by Brown (1995) participants in the single-exemplar condition tended to overestimate frequency compared to the presentation frequency. Experiment 1 did not provide support for this finding at the larger presentation frequencies. One concern is that the upper bound of the scale in the previous experiment resulted in attenuated frequency judgments.

According to Brown (1995), scale should not affect frequency judgments in the multiple-exemplar conditions. Assuming an enumeration strategy, the scale's upper bound limit should not influence frequency judgments. The assumption is that enumeration is based on a count of examples and the upper bound does not determine the count. Thus, the upper bound of the scale should not affect which exemplars are retrieved and the subsequent frequency judgment. Following Brown's (1997) logic, even if enumeration is not used for all of the multiple-exemplar trials, the scale range for non-enumeration trials should be based on range used for enumeration trials. As a result, the scale manipulation should not affect frequency judgments. Although unlikely, it is possible that people used the bound of 24 as a comparison for the number of exemplars retrieved and the degree to which they should extrapolate. For example, participants could judge the number of items retrieved with this bound and adjust their frequency judgment to reflect their belief about the upper bound. If this is the case, scale would affect frequency judgments in the multiple-exemplar condition such that larger judgments would be provided in the unbounded scale condition.

If a familiarity strategy is used in the single-exemplar conditions, then upper bound of the scale should influence frequency judgments. In particular, larger frequency judgments should be estimated in the unbounded scale compared to the bounded scale conditions.

Consistent with the results of Experiment 1, no difference in typicality should be observed in the single-exemplar condition with the bounded scale. The unbounded condition allows the opportunity to investigate how typicality affects familiarity and its frequency judgments. If differences emerge between typical and atypical single-exemplar conditions, it should be detected in the unbounded condition because there is no upper limit placed on the mapping process of familiarity to a numerical response.

Recall

Consistent with the previous experiment, recall should be greater in the typical than the atypical condition, but only for multiple-exemplar conditions. Scale should not interact with recall of exemplars for either the single or multiple-exemplar conditions. Differences should be small or nonexistent in recall for the single-exemplar conditions because only one exemplar was presented with each category.

Method

Participants

Two hundred and eight University of Maryland students participated in the experiment as partial fulfillment towards course credit. Participants from the previous experiment were excluded from participating in this experiment. Additionally, 12 participants in the bounded scale provided frequency judgments above the bound of 12. These participants were replaced with new participants who did not exceed the bound in their frequency judgments.

Materials

The materials used were identical to Experiment 1 including the 25 categories and 12 typical and 12 atypical exemplars per category (see Appendix A).

Design

The experimental design consisted of a 2 (exemplar typicality: typical or atypical) x 2 (exemplar presentation: same or different) x 2 (scale: bounded or unbounded) x 5 (presentation frequency: 0, 2, 4, 8, or 12) mixed factorial. Exemplar typicality, exemplar presentation, and scale bound were manipulated between participants resulting in 8 experimental conditions. Presentation frequency manipulated within participants.

Procedure

Participants were randomly assigned to one of eight conditions with 26 participants per condition. In the study phase, participants viewed the category-exemplar pairs in one of the following conditions: typical-single-exemplar, atypical-single-exemplar, typical-multiple-exemplar, or atypical-multiple-exemplar. The presentation and conditions were identical to the ones used in Experiment 1. As before, a brief filler task was implemented between the study and test phase. The test phase was similar to the previous experiment with the exception of the scale bound manipulation. As in Experiment 1, participants were presented with a category in a random order. Participants pressed any key to initialize the trial, pressed the space bar when they had a number in mind, and then entered their frequency judgment. For the frequency judgment task, a box appeared in the middle of the screen (with the category label still presented on the screen). Rather than pressing labeled keys on the keyboard, participants entered their frequency judgment using the number pad located on the right side of the keyboard and hit the enter key on the number pad to record their response. Half the participants received instructions that their frequency judgment should be between the numbers of 0 to 12. Thus, the upper bound of their judgment was identical to the largest presentation frequency of 12. No upper limit was placed on frequency judgment for the other half of participants. After the frequency judgments task, participants recalled exemplars from the study phase with the category label presented on the screen as a cue to recall exemplars from that category.

Results

Identical to Experiment 1, experimental results for Experiment 2 are presented in the next sections in terms of recall, RTs and frequency judgments.

Recall

The recall data were separated by exemplar presentation because the single-exemplar conditions could only recall one exemplar whereas the multiple-exemplar conditions could recall multiple exemplars. Additionally, the zero frequency was eliminated from recall data. Table 3 presents the mean recall data for multiple and single-exemplar conditions by typicality and scale.

Table 3: Mean Recall (Standard Error) for Multiple and Single-Exemplar Conditions in Experiment 2

Condition	2	4	8	12	sum
Multiple-exemplar					
Bounded					
Typical	0.81 (.06)	1.41 (.13)	1.97 (.21)	2.61 (.35)	34.00 (3.42)
Atypical	0.54 (.06)	0.75 (.10)	1.51 (.17)	1.78 (.23)	22.88 (2.40)
Unbounded					
Typical	0.65 (.08)	1.30 (.13)	2.43 (.23)	3.25 (.33)	38.16 (3.41)
Atypical	0.56 (.08)	0.84 (.09)	1.44 (.14)	1.99 (.17)	24.08 (1.93)
Single-exemplar					
Bounded					
Typical	0.85 (.04)	0.91 (.03)	0.95 (.02)	0.96 (.01)	18.42 (0.42)
Atypical	0.62 (.06)	0.82 (.04)	0.92 (.03)	0.88 (.17)	16.15 (0.61)
Unbounded					
Typical	0.85 (.05)	0.86 (.04)	0.95 (.02)	0.95 (.03)	18.04 (0.57)
Atypical	0.70 (.06)	0.84 (.04)	0.87 (.04)	0.84 (.05)	16.23 (0.75)

The recall data for four multiple-exemplar participants (2 atypical-multiple-exemplar-unbounded 1 typical-multiple-exemplar-bounded and 1 typical-multiple-exemplar-unbounded) were not included in the subsequent analyses due to missing recall data. In the multiple-exemplar conditions, two interactions were observed: frequency by typicality, F (3, 384) = 7.72, p <.01, $\hat{f}$= .41 and frequency by scale F (3, 384) = 2.95, p = .03, $\hat{f}$=.24. The interaction of presentation frequency by typicality reflects more recalled items as frequency increased for typical compared to atypical condition participants. Although the interaction of frequency by scale was not expected, the unbounded scale resulted in slightly more recalled exemplars as frequency increased compared to the bounded condition.

Additionally, two main effects were present: presentation frequency, F (3, 384) = 146.49, p < .01, $\hat{f}$=.90, and exemplar typicality F (1, 384) = 39.40, p < .01, $\hat{f}$=.39. These main effects merely represent that recall increased as the presentation frequency increased and typical conditions resulted in more recalled exemplars than atypical conditions. Although the typical advantage in recall occurred, the overall number of items recalled was less than in Experiment 1. The typical conditions produced 26-29% and atypical conditions produced 18-19% of total exemplars recalled.

Scale bound did not affect recall in the single-exemplar conditions as it did in the multiple-exemplar conditions. Analyses of recall for the single-exemplar conditions revealed an interaction of presentation frequency by typicality, F (3, 400) = 4.87, p < .01, $\hat{f}$=.32. As can be noted from Table 3, recall was consistent across frequency for the typical condition but not in the atypical condition. Furthermore, two main effects were present: exemplar typicality, *F* (1, 400) = 11.60, $p < .01$, $\hat{f}$=.31 and presentation frequency *F* (3, 400) = 29.26, $p < .01$, $\hat{f}$=.67. The main effect of presentation frequency was observed in Experiment 1, but the typicality main effect was not observed. The number of participants in this experiment increased the power to detect the main effect of typicality whereby typical exemplars were better recalled than atypical exemplars. The atypical conditions retrieved approximately 81% of exemplars whereas the typical conditions retrieved approximately 91% of exemplars.

RTs

As in Experiment 1, RT data was based on the aggregate of the RT to press the space bar when they had a number in mind and the RT to enter the frequency judgment. All analyses were based on the median RTs per category frequency.

Group Median RTs. Figures 6a-6d presents the mean median RT at each level of presentation frequency for each of the experimental conditions. Figure 6a and Figure 6b represent the multiple-exemplar for bounded and unbounded scale conditions, respectively. Figure 6c and Figure 6d represent the single-exemplar for bounded and unbounded scale conditions, respectively.

Unlike Experiment 1, typicality did not interact with exemplar presentation or presentation frequency. As before, a presentation frequency by exemplar presentation resulted in a significant interaction, *F* (4, 1000) = 3.96, $p < .01$, $\hat{f}$=.11. As can be noted from Figures 6a-6d, the multiple-exemplar conditions yielded larger median RTs as presentation frequency increased whereas the slope for the single-exemplar RTs were shallower (despite the slope increasing in the single-unbounded conditions).

A main effect of presentation frequency also occurred, F (4, 1000) = 36.59, $p <$ .01, $\hat{f}$ = .35. This main effect reflects that RT tended to increase across presentation frequency when all RT data were collapsed across conditions. [10] As with Experiment 1, the RT data were converted to log and 1/RT transformations. This time, the log transformation yielded the same results as the untransformed data, producing a main

[10]Analyses of the RT data using the mean rather than the median RT resulted in identical results in the multiple-exemplar condition, but unbounded condition resulted in larger slopes than the bounded condition in the single-exemplar condition.

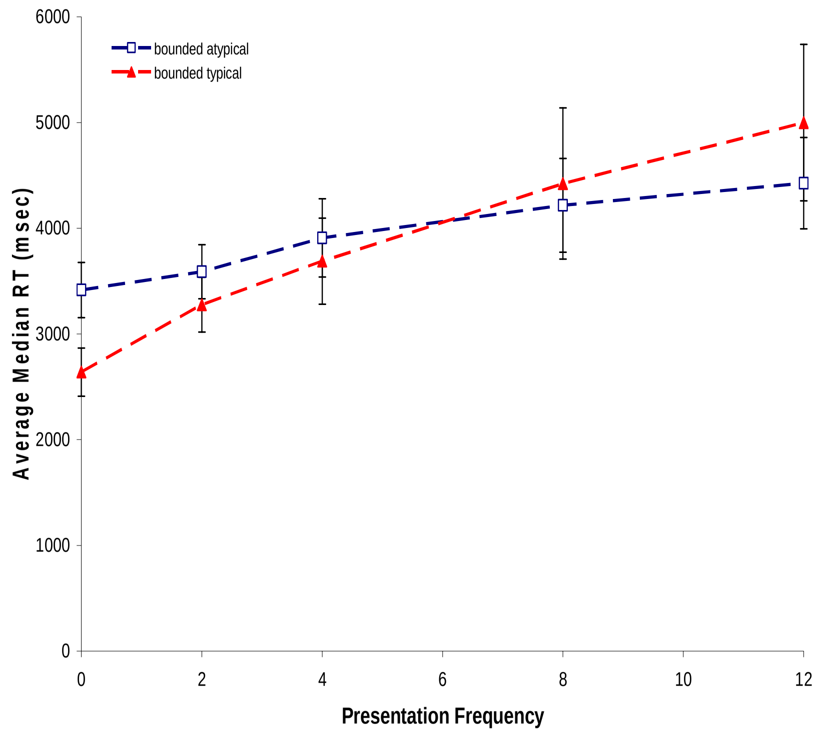

Figure 6a: Average Median RT for Multiple-Exemplar-Bounded Condition in Experiment 2

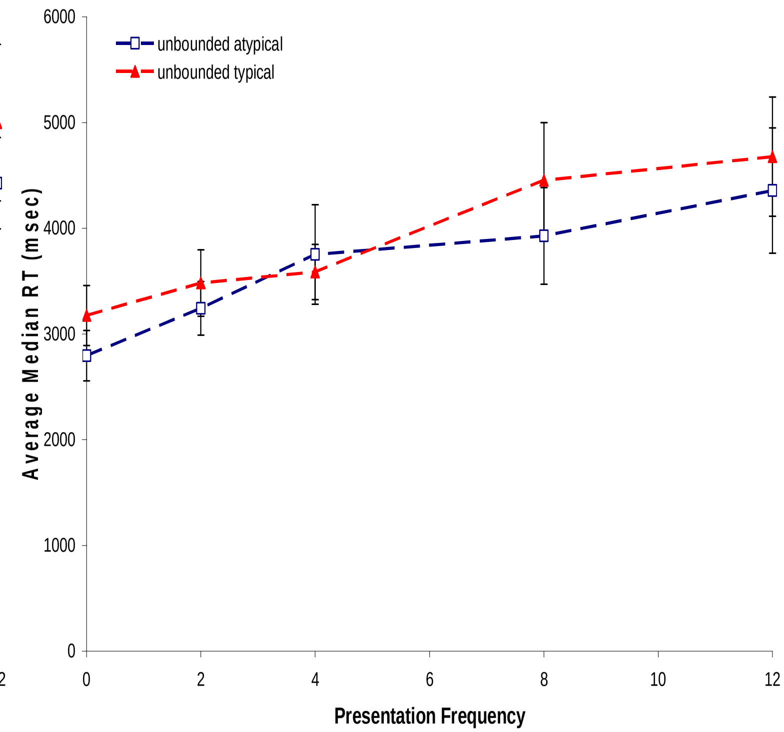

Figure 6b: Average Median RT for Multiple-Exemplar-Unbounded Condition in Experiment. 2

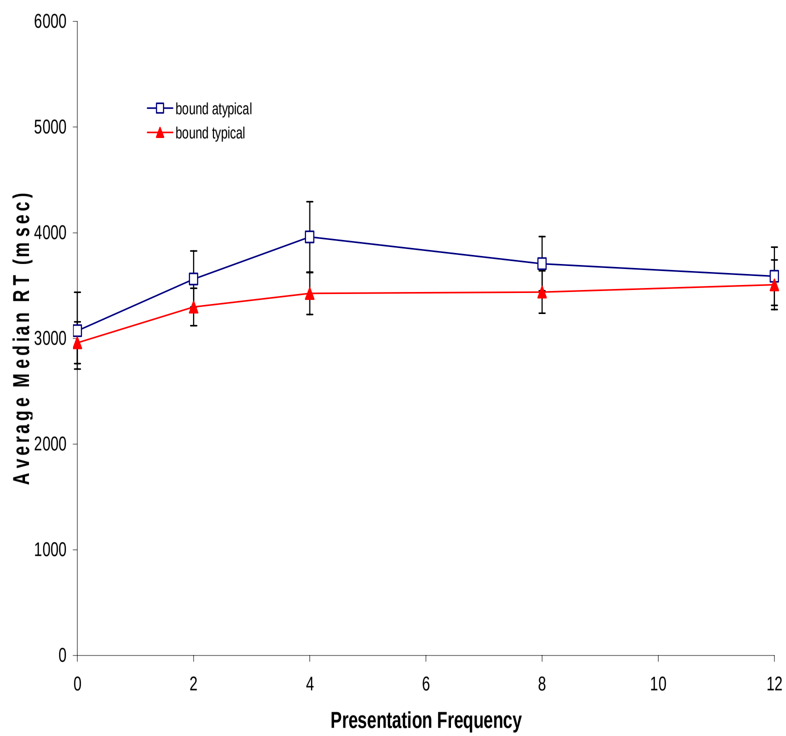

Figure 6c: Average Median RT for Single-Exemplar-Bounded Condition in Experiment 2

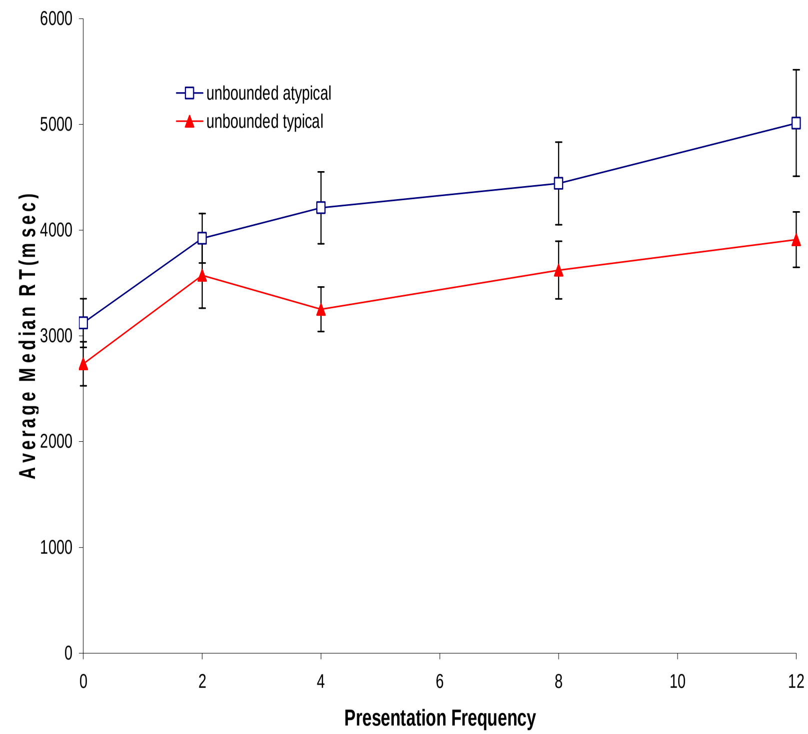

Figure 6d: Average Median RT for Single-Exemplar-Unbounded Condition in Experiment. 2

effect of frequency, *F* (4, 1000) = 55.95, *p* < .01, $\hat{f}$ = .42 and an interaction of presentation frequency by exemplar presentation, *F* (4, 1000) = 2.46, *p* = .05, $\hat{f}$ = .07. However, the 1/RT transformation yielded only a main effect of presentation frequency, *F* (4, 1000) = 59.54, *p* < .01, $\hat{f}$ = .43. Once again, the RT data were not robust to various transformations.

Group Median RT slopes. Regression slopes were computed for each participant based on median RT at each level of presentation frequency. Analyses examining the group-level mean regression slopes (based on median RTs) produced a main effect of exemplar presentation, *F* (1, 200) = 6.39, *p* = .01, $\hat{f}$ =.16. This is consistent with the interaction exemplar presentation by presentation frequency for the median RT presented above. Overall, typical conditions resulted in larger RT slopes than the atypical conditions. Additionally, two interactions were near significance: Exemplar typicality by exemplar presentation, *F* (1, 200) = 2.87, *p* =.09 $\hat{f}$ =.09, and exemplar presentation by response scale, *F* (1, 200) = 2.98, *p* =.09, $\hat{f}$ = .10. Figure 7 presents the mean regression slopes per condition. Note that in the multiple-exemplar conditions, the regression slopes are slightly lower for the atypical than typical condition, but the reverse pattern occurs in the single-exemplar conditions. In particular, the atypical slopes are larger than the typical slopes in the unbounded scale conditions.

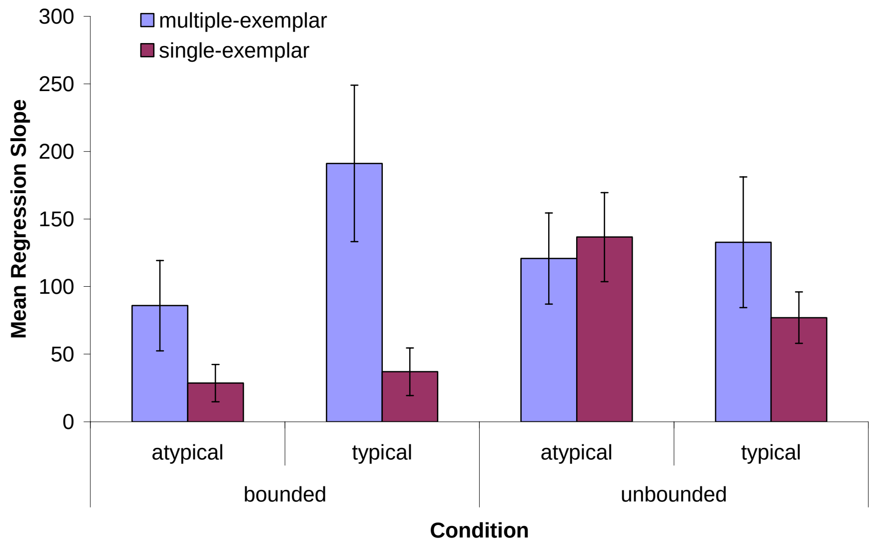

Figure 7: Average RT Regression Slope per Condition in Experiment 2

Individual Median RT Slopes. Individual regression slopes are of special interest because they indicate how many participants engage in enumeration in each condition. Figure 8a presents the individual regression RT slopes for each participant in the bounded condition. Figure 8b presents the regression slopes for each participant in the unbounded conditions. Overlapping individual slopes within a condition are marked by darker points on the figures. The mean per figure is computed by using a separate mean for single-exemplar conditions in the bounded condition and unbounded conditions. The mean slope for the bounded-single-exemplar (collapsed across typicality) conditions was 32.77 (*STD* = 80.08) compared to a mean slope of 106.84 (*STD* = 138.88) for the unbounded-single-exemplar (collapsed across typicality) conditions.

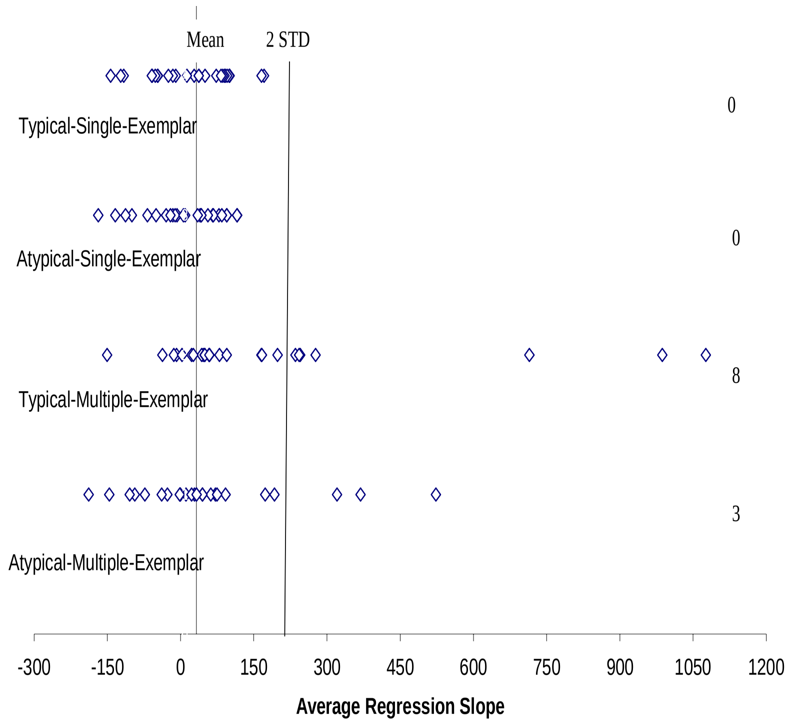

Figure 8a: Individual RT Regression Slopes in Bounded Conditions in Experiment 2

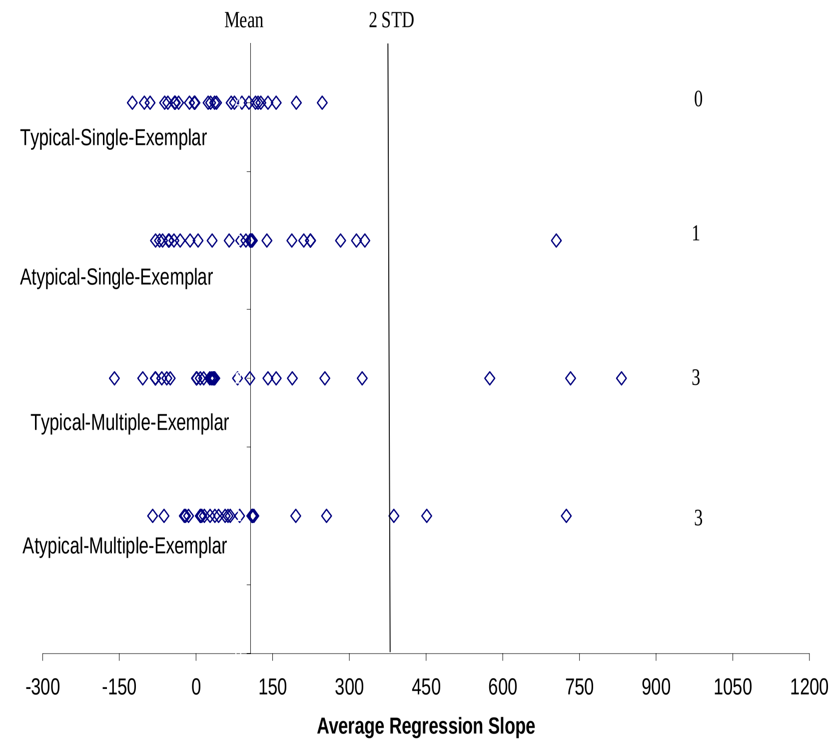

Figure 8b: Individual RT Regression Slopes in Unbounded Conditions in Experiment 2

Of interest are the outliers who are presented two SD above their respective mean. The numbers on the far right of the figures represent the number of participants per condition who were 2 SD above the mean slope.

Overall, few (n = 18) individuals were 2 SD above the mean. Outliers were observed most often in the multiple-exemplar conditions. In the typical-multiple-bounded condition, 31% of the participants were considered outliers for this condition. In the other three multiple-exemplar conditions (i.e., atypical-multiple-bounded, typical-multiple-unbounded, and atypical-multiple-unbounded), only 12% of the participants for each of the conditions were considered outliers for their respective conditions. Also, the atypical-single-unbounded condition produced one outlier. Fisher's exact test revealed a significant difference on the number of outliers based on exemplar presentation ($p < .05$), but typicality and scale did not affect outlier distributions (*ns*). Again, slopes of several individual participants indicated the use of enumeration strategies, but most of the slopes indicate that people relied on non-enumeration strategies. The pattern observed in the group-level of analysis is not observed at the individual-level of analysis. Differences in the group-level RTs are reflected by a few outliers. The number of outliers (31% of participants) in the typical-multiple-bounded condition is similar to the number observed in Experiment 1(34% of participants) for the typical-multiple-exemplar condition. However, the typical-multiple-unbounded resulted in only 12% of participants who used enumeration.

Frequency Judgments

All frequency judgments were based on the average of 5 judgments per category frequency per participant.

Group Frequency Judgments. Figures 9a through 9d present the mean frequency judgments. Figure 9a and Figure 9b represent the multiple-exemplar conditions with the bounded scale condition in Figure 9a and the unbounded scale condition in Figure 9b. Figure 9c and Figure 9d represent the single-exemplar conditions with the bounded scale in Figure 9c and the unbounded scale in Figure 9d. An ANOVA conducted on estimated frequency revealed a three-way interaction for presentation frequency by exemplar presentation by scale, F (4, 1000) = 2.69, p =.03, $\hat{f}$=.08. The single-exemplar condition, unbounded (Figure 9d) resulted in larger frequency estimations across presentation frequency in comparison to the other three figures. This is further illustrated by three significant interactions: exemplar presentation by scale, F (1, 1000) = 8.68, $p < .01$, $\hat{f}$=.09, presentation frequency by exemplar presentation, F (4, 1000) =14.19, $p < .01$, $\hat{f}$=.22 and presentation frequency by scale, F (4, 1000) =4.18, $p < .01$, $\hat{f}$=.11. The exemplar presentation by scale interaction reflects the single-exemplar was more influenced by scale than the multiple-exemplar. The frequency by exemplar presentation interaction is reflected in the steeper frequency slopes for the single-exemplar than the different-exemplar conditions. As presentation frequency increased the unbounded scale estimates increased to a greater degree than the bounded scale estimates resulting in the presentation frequency by scale interaction. Typicality did not interact with exemplar presentation, scale or presentation frequency. As can be noted in the figures, frequency judgments were virtually identical irrespective of exemplar typicality.

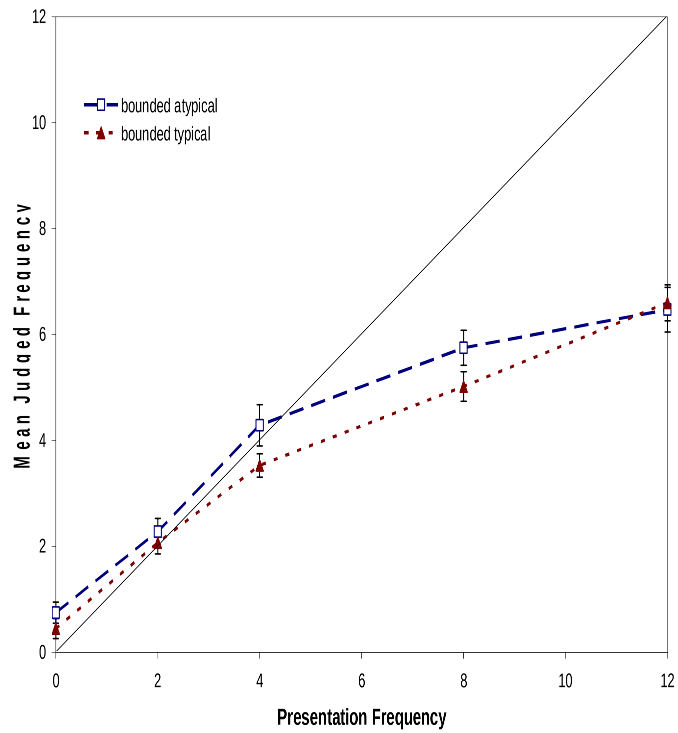

Figure 9a: Average Frequency Judgments for Multiple-Exemplar-Bounded Condition in Experiment 2

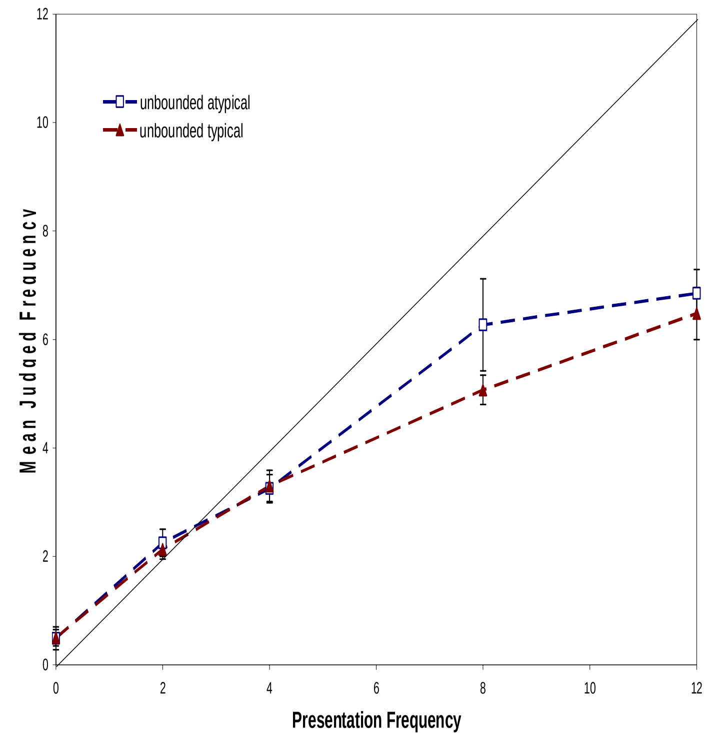

Figure 9b: Average Frequency Judgments for Multiple-Exemplar-Unbounded Condition in Experiment 2

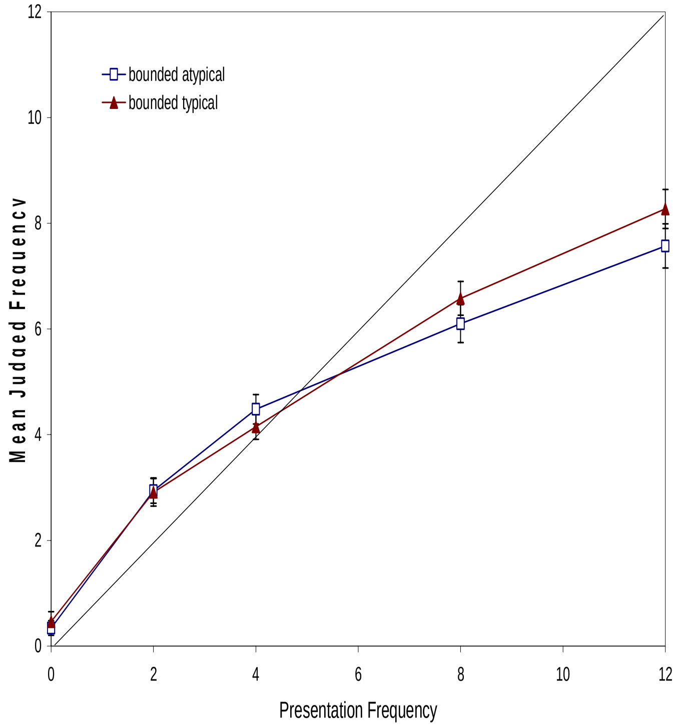

Figure 9c: Average Frequency Judgments for Single-exemplar-Bounded Condition in Experiment 2

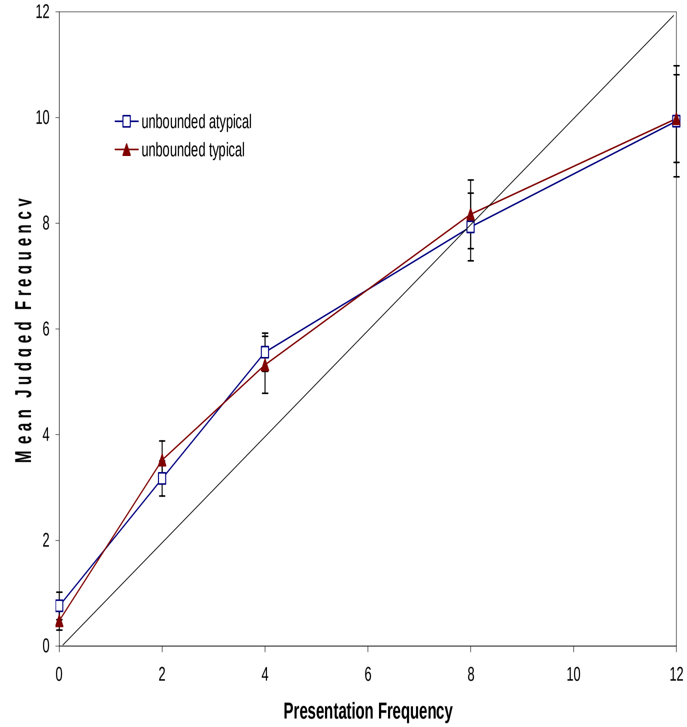

Figure 9d: Average Frequency Judgments for Single-exemplar-Unbounded Condition in Experiment 2

Furthermore, the main effects of all the independent variables (with the exception of typicality) were significant. Single-exemplar resulted in larger frequency estimations than the multiple-exemplar resulting in a main effect of exemplar presentation, F (1, 1000) =39.32, p < .01, $\hat{f}$=.19. The unbounded scale resulted in larger frequency estimates than the bounded scale resulting in a main effect of scale, F (1, 1000) =6.90, p < .01, $\hat{f}$=.08. The frequency main effect reflects general sensitivity to presentation frequency, F (4, 1000) =616.05, p < .01, $\hat{f}$=.84. [11]

Table 4 presents the upper and lower quartiles of frequency judgments at each level of presentation frequency by condition. Rank-sum analyses were conducted for each level of presentation frequency across conditions and between conditions. Similar to Experiment 1, the analyses revealed no difference in 0 frequency (*ns*) and differences across conditions in frequencies 2, 4, 6, and 12, $\chi^2(7) = 29.84$, $\chi^2(7) = 45.31$, $\chi^2(7) = 48.69$, and$\chi^2(7) = 33.40$, p < .01, respectively. Similar to frequency judgments, the scale range illustrates that multiple-exemplar conditions produced smaller ranges than the single-exemplar conditions, (all Wilcox rank-sum analyses produced significant differences for 2, 4, 8, and 12 frequencies). Again, as in Experiment 1, typicality did not affect scale range. Interestingly, scale (bounded versus unbounded) also did not produce an effect. This is most likely due to the fact that scale only affects the single-exemplar conditions and not the multiple-exemplar conditions as well as the imposed bound of 12. As a consequence, averaging across

[11]Analyses using the median rather than the mean frequencies revealed the same results except that the presentation frequency by scale interaction was near significance (p = .07).

scale conditions produces a null effect because most of the conditions resulted in underestimation of frequency.

Table 4: Lower and Upper Quartile Frequency Judgments for Presentation frequencies Per Condition for Experiment 2

Objective	Typical-Multiple		Atypical-Multiple		Typical-Single		Atypical-Single	
Frequency	Q1	Q3	Q1	Q3	Q1	Q3	Q1	Q3
Bounded Scale								
0	0	0.2	0	1.0	0	0.6	0	0.2
2	1.7	2.4	1.4	3.0	2.0	3.4	2.2	3.6
4	3.5	4.2	2.6	5.2	3.4	5.2	3.4	6.0
8	5.0	5.6	4.6	6.6	5.8	7.2	4.6	7.0
12	6.7	7.6	5.0	8.0	7.2	9.4	5.6	9.2
Unbounded Scale								
0	0	0.8	0	0.6	0	0.2	0	1.2
2	2.1	2.8	1.4	3.2	2.4	3.6	2.2	3.8
4	3.3	4.0	2.4	4.0	3.8	6.0	4.2	6.6
8	5.1	5.8	4.0	6.2	6.2	9.8	6.6	8.6
12	6.5	7.6	5.6	8.2	7.6	12.4	6.4	11.6

Group Median Frequency Slopes. Regression slopes were computed for each participant based on the median frequency estimate per level of presentation frequency. An ANOVA on the group regression slopes revealed a main effect of exemplar presentation, F (1, 200) = 20.88, $p < .01$, $\hat{f}$=.30. Consistent with Brown (1995) and Experiment 1, multiple-exemplar (*M* =.50, *STD* = .19) slopes were smaller than the single-exemplar (*M* =.67, *STD* = .35) conditions. The exemplar presentation by scale interaction was significant, *F* (1, 200) = 4.08, $p < .05$, $f^{\wedge}$=.12. This is reflected by the single-exemplar slopes for bounded (*M* =.60, *STD* = .19) and unbounded conditions (*M* =.75, *STD* = .44), *t* (51) = 2.16, *p* =.03, are more spread apart than the multiple-exemplar slopes for bounded (*M* =.50, *STD* = .19) and unbounded (*M* =.50, *STD* = .20), *t* (51) = 0.19, *ns*. Typicality did not affect slopes. Overall, all conditions produced underestimation of estimated frequency compared to presentation frequency. This is a function of our bounded scale of 12 which cannot lead to overestimation. The single-exemplar unbounded condition did not overestimate presentation frequencies of 12. In Brown's (1995) experiments, all presentation frequencies were overestimated in the single-exemplar conditions (particularly when unbounded limit was imposed on the response scale).

Individual Median Frequency Slopes. Figures 10a and 10b present the individual frequency slopes for each of the 8 conditions. Figure 10a represents the bounded conditions and Figure 10b represents the unbounded conditions. The line intersecting each figure represents the slope of 1.0 which would indicate perfect correspondence between frequency judgments and presentation frequency. As can be noted from the figures, few slopes indicated that participants overestimated the presentation frequencies. Part of the reason for the lack of overestimation is due to the restricted bound of 12 which would deem impossible for anyone to overestimate the presentation frequency. In the unbounded condition, slopes above 1.0 were found in the single-exemplar conditions. The group-level data indicated underestimation of presentation frequencies and the individual-level data revealed the same pattern.

However, the differences in the single-exemplar conditions were due to a few participants who had a slope larger than 1.0 in single-unbounded condition.

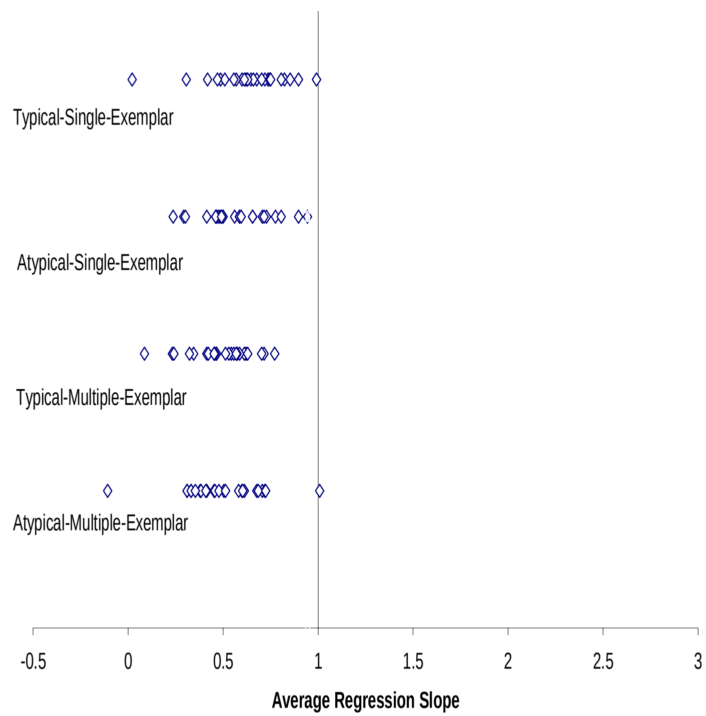

Figure 10a: Individual Frequency Slopes per Bounded Conditions in Experiment 2

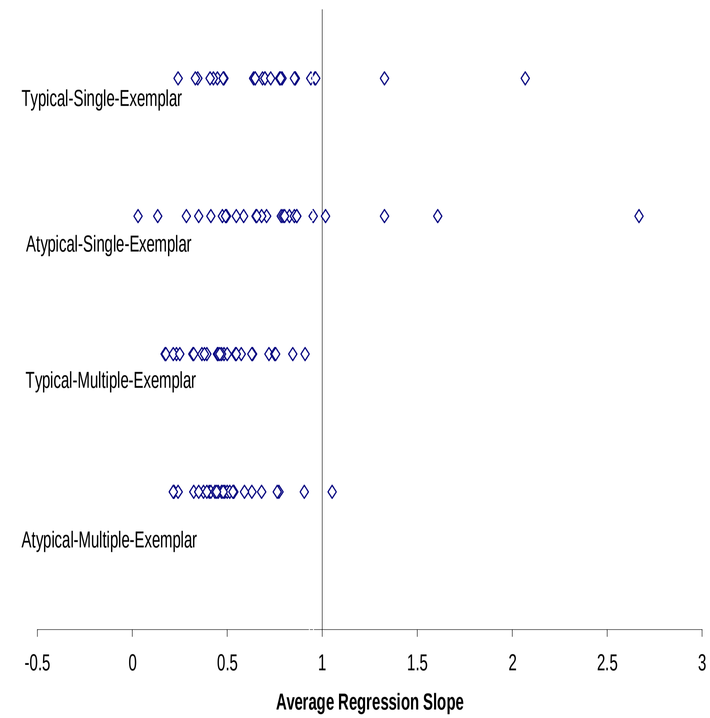

Figure 10b: Individual Frequency Sloes per Unbounded Conditions in Experiment 2

Discussion

Unlike Experiment 1, typicality did not interact with exemplar presentation at the group-level RTs. However, exemplar presentation did interact with presentation frequency such that the multiple-exemplar conditions demonstrated increased RTs as presentation frequency increased. Corresponding with the results of Experiment 1, the group-level and individual-level analyses revealed a different pattern in the multiple-exemplar conditions based on the unit of analysis. Enumeration strategies appeared at the group-level analyses, but the individual-level analyses indicated only a small number of participants engaged in enumeration. The group-level and individual-level analyses were similar in the single-exemplar conditions.

In terms of frequency judgments, there was no effect of typicality. Scale and exemplar presentation did affect frequency judgment. As expected, the unbounded scale affected frequency judgments in the single-exemplar conditions but not in the multiple-exemplar conditions. One point of interest is that the single-exemplar condition did not result in overestimation as demonstrated in Brown (1995) when the scale was unbounded. The difference in under/overestimation between these experiments might be the result of the levels of presentation frequency used. Specifically, in the Brown experiments, frequency 16 represented the highest level of presentation frequency whereas frequency 12 represented the highest level of frequency in the experiments outlined above. Because categories were counterbalanced across levels of frequency in the present experiments, several categories did not contain 16 atypical exemplars resulting in only 12 exemplars per category. Thus, mapping familiarity of frequency of 12 might not require a large scale range compared to frequency 16. The group and individual levels approaches to frequency judgments revealed that the single-exemplar condition unbounded frequency judgments were due to a few participants whose slope was above 1.0.

As in Experiment 1, typicality affected recall in the different condition with typical exemplars being recalled more often than atypical exemplars. This effect was also present in the single-exemplar conditions. Thus, recall data support the typicality effects are similar to normative word frequency effects in recall, but typicality did not affect frequency estimates and RT in Experiment 2.

Chapter V: General Discussion

The group and individual approaches allowed for the examination of strategy use and frequency judgments within and across groups. RT slopes suggest different interpretations depending on whether the group or individual-level of analysis was implemented. The RT slopes at the group-level indicated that enumeration was used in the multiple-exemplar conditions. If enumeration is used for all trials, then outliers should be prevalent in the individual slope data, especially in the multiple-exemplar conditions. Although the typical-multiple-exemplar conditions in Experiments 1 and 2 produced several outliers, 12-34% of participants engaged in enumeration (as indicated by the RT slopes). Few atypical-multiple-exemplar participants (approximately 8-12 %) had RT slopes consistent with the use of enumeration. As a result, RT data can yield different conclusions about which strategy is evoked based on the level of analysis. The results appear to be conflicting when the individual versus group analyses are used to infer strategy use based on RTs. Using the group to infer strategy use within a judgment task can lead to overestimation of the prevalence of a specific strategy.

Although Brown (1997) conceded that people might not use enumeration for all trials. A concern with using RTs to infer strategy use is that RTs cannot adequately reflect which participants engage in a mixed strategy across trials. Using the average or median RT provides some indication of which strategy might be used most often, but one cannot draw conclusive inferences.

In category-exemplar pairs, typicality does not play an integral role in which strategy people use and in the actual judged frequencies. Is the lack of effect due to the fact that typicality does not operate like normative word frequency effects in recall and recognition? The recall of exemplars provided evidence that the typicality advantage operated similarly to normative word frequency effects in recall. However, this recall effect did not yield differences in frequency judgments in either experiment. The typicality advantage was not observed in the single-exemplar conditions. There are two possible explanations for the lack of a typicality effect in the single-exemplar conditions.

The first explanation is based on what information is used to probe memory. Because the test phase used only the category name and not exemplars, the probe would not necessarily contain any exemplar elements, which would result in a null typicality effect on familiarity processes.

The second explanation focuses on the effect of repetition on frequency judgments. Specifically, it is still possible that participants use the exemplar alongside with the category to probe memory, but the presentation frequency might wash out any advantages due to typicality. One of the reasons provided for the normative word frequency effect in recognition is that low-frequency words receive more attention at encoding than high-frequency words. In the experiments, participants studied the categories with the same exemplar at repeated frequencies. Thus, the presence of repetitions might allow all exemplars to be encoded at the same rate irrespective of typicality.

Can the global memory models using familiarity account for the data observed in the experiments for both single and multiple-exemplar conditions? If one assumes that familiarity occurs in both exemplar presentations, then global memory models would account for the differences in frequency judgments between single and multiple-exemplar conditions. Familiarity can be used to account for underestimation of presentation frequencies in the multiple-exemplar conditions based on the match between the probe and the traces stored in memory (cf. Hintzman, 1988). In the multiple-exemplar conditions, the traces are less similar to one another. The result of less similar traces is a decrease in familiarity which would yield smaller frequency judgments. However, it would be difficult to incorporate the scale effects for the single-exemplar and multiple-exemplar conditions without making ad hoc assumptions.

Future Directions

One concern is that the RTs do not allow the distinction between participants who engage in a mixed strategy consisting of enumeration and non-enumeration strategies and those who engage exclusively in non-enumeration strategies. The premise of using a pure exemplar presentation where all exemplars vary or remain constant at each level of presentation frequency was to ensure that one memorial process did not determine which strategy was used. Specifically, recall processes (enumeration strategies) would not be disrupted or overwhelmed by recognition processes (non-enumeration strategies). On the surface, the results might appear to be cleaner using a pure exemplar presentation experiment. On the other hand, in reality, a combination of strategies is likely to be used when estimating the frequency of an event. For example, estimating the number of times you have dined at restaurants can be based on going to a small number of restaurants many times as well as many restaurants less frequently. Thus, it seems necessary to investigate whether mixed exemplar presentations produce the same effects on RT, frequency estimates, and scale as do pure exemplar presentations.

Furthermore, assuming enumeration is effortful because it generally takes more time to for sequential retrieval, why do people use enumeration rather than rely solely on a familiarity process? Perhaps the key to understanding when enumeration will be used depends on the ability to retrieve examples from memory.

For example, working memory capacity might determine the number of items retrieved and held in working memory which should affect which strategy is likely to be used and its affect on frequency judgments. Engle and colleagues have proposed that tasks that measure working memory (WM) capacity actually reflect general controlled-attention capacity (Engle, 1996; Engle, Kane, & Tuholski, 1999; Kane Bleckley, Conway, & Engle, 2001).

WM plays a role in memory and retrieval. Rosen & Engle (1997) demonstrated that WM is positively correlated with verbal fluency in that people with high WM span generated more examples of a category than people with low WM span. Additionally, Dougherty and Hunter (2003) demonstrated that individual differences in WM capacity correlated with the number of alternatives generated in that individuals with higher WM generated more alternatives than individuals with low WM. Consequently, differences might emerge especially in multiple-exemplar conditions based working memory capacity.

Implications

It is important to define the conditions that might give rise to specific strategies and its subsequent effect on frequency judgments. Frequency estimates play a role in many facets of everyday life. For example, people's assessment of frequency can be used to determine which consumer products are developed or advertised, government allocation of resources (e.g., mass transit development), and even employee performance ratings (how often someone was late, secured clients etc.). The present data indicate that judging the frequency of a typical event will not necessarily result in a different judgment than an atypical event.

Although there was little evidence of enumeration strategy is the experiments, this does not preclude the possibility that an enumeration strategy is used to assess the frequency of an event. Evidence for different strategies used in frequency judgments was established in behavioral reports. Brown and Sinclair (1999) found that women used enumeration more often than men in estimating lifetime sexual partners. However, enumeration might not be the default strategy when multiple-exemplar conditions. Participants need not to rely on counting in order to make meaningful frequency judgments. Watkins and Le Compte (1991) found that number of examples retrieved were not relevant to the frequency estimate. As Brown (1995) stated that multiple-exemplar conditions are necessary but not sufficient for enumeration to occur.

In recent years, there has been increased interest in examining individual differences in psychological research. For example, Stanovich (1999) used an individual difference approach to examine those individuals who were able to excel at reasoning and decision-making scenarios. Additionally, individual differences in working memory have been used to examine memorial processes (cf. Engle, 1996). However, an individual versus group level approach to examine patterns of data is rarely used. Typically, outliers are eliminated from the data in order to achieve homogeneity of the group. However, the presence of outliers can also detect a different strategy or perhaps a meaningful interpretation to the experiment. A problem ensues when the few outliers who are behaving in a meaningful way guide the interpretation at the group-level of analysis. Examining group and individual-levels of the data analysis is useful to detect whether a consistent pattern of results can be replicated at both levels of analyses. Thus, outliers can be informative especially if they occur in the conditions that are expected, but results need to be examined at both the individual-level and group-level of analysis to determine the consistency with which a specific pattern occurs and its implications on the theory that is being tested.

Appendix A

ANIMAL

Typical	#	Atypical	#
dog	426	burro	1
cat	412	colt	1
horse	348	cub	1
cow	284	fawn	1
lion	225	iguana	1
tiger	203	jackal	1
elephant	182	mole	1
pig	142	tortoise	1
bear	129	walrus	1
mouse	118	boar	2
deer	95	gopher	2
sheep	85	bison	4

BIRD

Typical	#	Atypical	#
robin	377	quail	9
sparrow	237	rooster	7
cardinal	208	egret	5
bluejay	180	dodo	2
eagle	161	partridge	1
crow	149	heron	4
bluebird	138	condor	4
canary	134	martin	6
parakeet	115	meadowlark	7
hawk	111	toucan	2
blackbird	89	cockatoo	1
wren	83	emu	4

BODY

Typical	#	Atypical	#
legs	402	skull	4
arm	398	thorax	4
head	308	bicep	3
eye	303	femur	3
foot	295	backbone	1
nose	281	tendons	1
finger	279	tonsil	1
ear	20	uterus	1
hand	228	face	1
toe	223	eyelid	1
mouth	184	arteries	5
stomach	128	spleen	6

BUILDING

Typical	#	Atypical	#
window	338	gutters	7
door	322	shutters	7
roof	266	shingle	6
wall	242	pipe	6
floor	238	cornice	4
ceiling	167	pillar	2
room	161	carpet	2
basement	108	staircase	2
brick	101	dormer	2
hall	65	columns	1
stair	64	portico	1
elevator	61	tile	5

CITY

Typical	#	Atypical	#
New York	361	Buffalo	9
Chicago	270	Hong Kong	9
Los Angeles	194	Sacramento	8
Washington	179	Calcutta	2
San Francisco	173	Charlotte	2
Baltimore	165	Helsinki	1
London	84	Burbank	1
Paris	84	Anchorage	1
Miami	79	San Antonio	2
Boston	74	Norfolk	3
Philadelphia	63	Louisville	4
Detroit	52	Nashville	1

CLOTHES

Typical	#	Atypical	#
shirt	352	swimsuit	4
socks	330	pajamas	8
pants	318	cap	9
shoes	274	robe	7
blouse	261	apron	2
skirt	261	cape	2
coat	260	jeans	2
dress	240	tuxedo	2
hat	201	parka	2
sweater	163	gown	1
tie	136	jersey	2
slip	127	blazer	1

COLLEGE

Typical	#	Atypical	#
Harvard	160	Brigham Young	
Yale	133	Citadel	1
UCLA	92	Emory	3
Princeton	84	Howard	5
Duke	57	Brandeis	9
Penn State	56	Colgate	6
Northwestern	48	Temple	5
MIT	42	Morgan State	3
Purdue	38	Villanova	3
Cornell	37	Galludet	1
Indiana	36	Frostburg	9
American	33	Marymount	2

COLOR

Typical	#	Atypical	#
blue	438	salmon	3
red	435	lime	1
green	431	plum	1
yellow	387	amber	3
orange	382	ivory	3
black	314	lilac	2
purple	282	taupe	1
pink	224	auburn	2
white	272	bronze	1
brown	217	copper	1
violet	153	crimson	2
gray	94	azure	6

COUNTRY

Typical	#	Atypical	#
France	339	Borneo	1
Russia	303	Croatia	1
England	300	Honduras	1
Germany	228	Kenya	1
Canada	163	Libya	1
Italy	159	Morocco	2
Spain	159	Sudan	2
Mexico	136	Syria	2
China	115	Ghana	3
Japan	103	Korea	5
Sweden	78	Ukraine	1
Brazil	77	Albania	5

FISH

Typical	#	Atypical	#
trout	216	sturgeon	9
bass	195	squid	8
shark	176	starfish	8
herring	161	octopus	5
catfish	143	grouper	5
perch	143	bluefin	4
salmon	142	sole	4
tuna	139	snapper	3
goldfish	127	jellyfish	8
swordfish	87	pickerel	7
sunfish	80	ray	4
rock	72	tiger	5

FLOWER

Typical	#	Atypical	#
rose	421	begonia	9
tulip	209	poinsettia	9
carnation	183	crocus	8
daisy	176	posy	8
violet	147	hibiscus	2
orchid	135	jasmine	1
chrysanthemum	124	blossom	1
lily	108	bud	1
pansy	108	honeysuckle	2
gardenia	89	dahlia	4
daffodil	87	wild	3
dandelion	75	dogwood	6

FRUIT

Typical	#	Atypical	#
apple	429	boysenberry	1
orange	30	cranberry	1
pear	32	citrus	1
banana	283	guava	1
peach	249	papaya	8
grape	247	honeydew	4
cherry	183	nectarine	12
plum	167	kumquat	10
grapefruit	154	melon	11
lemon	134	avocado	17
tangerine	110	mango	18
apricot	102	blackberry	9

FURNITURE			
Typical	#	Atypical	#
chair	440	armoire	1
table	408	console	1
bed	328	cradle	1
sofa	232	chaise	2
desk	230	vanity	7
dresser	143	settee	3
stool	72	ottoman	5
bureau	64	hutch	7
lamp	227	bookshelf	1
couch	168	cupboard	5
television	105	armchair	3
rug	51	cot	2

INSECT			
Typical	#	Atypical	#
fly	337	millipede	1
ant	258	mite	8
bee	227	scorpion	4
mosquito	227	cicada	5
spider	177	aphid	9
beetle	161	weevil	1
roach	123	fruitfly	8
wasp	108	bedbug	5
grasshopper	100	stinkbug	5
ladybug	99	katydid	3
gnat	77	chigger	4
moth	62	bumblebee	4

INSTRUMENT			
Typical	#	Atypical	#
piano	329	bongo	8
drum	322	tambourine	9
trumpet	279	mandolin	6
violin	271	recorder	5
clarinet	259	lute	4
flute	246	chimes	3
guitar	231	lyre	1
saxophone	176	kazoo	1
trombone	173	tympani	6
oboe	144	autoharp	2
tuba	119	ketlledrum	2
harp	105	woodwind	1

KITCHEN			
Typical	#	Atypical	#
spoon	369	baster	1
fork	349	canister	2
pan	242	spatula	1
pot	205	cleaver	2
spatula	155	cutlery	1
stove	74	dicer	1
bowl	69	drainer	1
cup	49	funnel	1
knife	382	griddle	1
mixer	74	mold	1
dish	42	server	1
plate	34	skewer	1

MATERIAL			
Typical	#	Atypical	#
cotton	404	chamois	3
wool	347	chintz	3
silk	292	damask	2
rayon	225	sateen	1
nylon	210	velour	1
dacron	153	tapestry	1
linen	142	organza	2
satin	78	twill	2
denim	29	cashmere	6
madras	32	polyester	6
velvet	32	suede	7
burlap	29	knit	8

METAL			
Typical	#	Atypical	#
iron	353	barium	4
copper	309	chrome	4
steel	281	pewter	4
gold	270	boron	3
aluminum	259	cesium	3
silver	253	argon	2
tin	173	plutonium	2
zinc	130	alloy	6
brass	97	francium	4
lead	96	radium	3
bronze	78	palladium	2
magnesium	75	bismuth	3

OCCUPATION

Typical	#	Atypical	#
doctor	362	agent	1
lawyer	269	astronomer	1
teacher	155	butler	1
dentist	112	miner	1
engineer	109	fisherman	1
professor	94	gardener	1
carpenter	83	reporter	1
nurse	47	seamstress	1
plumber	41	umpire	1
accountant	37	analyst	1
scientist	30	optician	1
clerk	29	teller	1

SPORT

Typical	#	Atypical	#
football	396	curling	1
baseball	376	karate	1
basketball	360	rowing	5
tennis	329	skating	4
swimming	277	running	8
soccer	160	croquet	7
golf	153	canoeing	6
hockey	130	cricket	6
track	111	waterpolo	1
lacrosse	107	kickball	1
badminton	96	climbing	1
bowling	96	hiking	6

STONE

Typical	#	Atypical	#
diamond	435	granite	2
ruby	419	marble	2
emerald	329	amherst	2
sapphire	245	limestone	1
pearl	177	sardonyx	1
opal	100	tigereye	1
jade	84	zephyr	1
topaz	72	coral	2
amethyst	62	peridot	2
onyx	59	ebony	2
garnet	44	crystal	5
turquoise	33	krypton	1

TOOL

Typical	#	Atypical	#
hammer	431	bezel	1
saw	394	ratchet	1
nails	248	wheelbarrow	1
screwdriver	214	axle	1
level	168	backsaw	1
plane	147	stapler	3
chisel	103	scissors	4
ruler	76	tape	4
wrench	61	triangle	4
pliers	58	router	1
drill	52	jigsaw	5
sawhorse	29	mallet	2

TOY

Typical	#	Atypical	#
doll	285	scooter	9
ball	212	checkers	7
car	146	kite	
gun	125	puppet	5
truck	125	sled	5
train	113	baton	1
game	106	robot	5
block	98	trike	1
bat	84	basketball	5
top	60	pistol	2
boat	56	seesaw	2
dollhouse	52	paint	2

TREE

Typical	#	Atypical	#
oak	394	fig	9
maple	314	mahogany	9
pine	214	acorn	6
elm	210	aspen	6
apple	163	balsam	5
birch	134	teak	5
cherry	119	rubber	5
dogwood	84	coconut	4
spruce	74	beechwood	2
redwood	71	coniferous	2
peach	67	olive	2
walnut	64	bamboo	1

VEGETABLE

Typical	#	Atypical	#
carrot	316	artichokes	5
pea	308	eggplant	8
corn	247	okra	5
bean	237	collards	4
potato	224	rhubarb	4
tomato	215	mushroom	2
lettuce	189	parsnip	2
spinach	163	pumpkin	2
asparagus	138	watercress	2
broccoli	126	leak	1
celery	96	yams	1
cabbage	94	rutabaga	2

References

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

Brown, N.R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1539-1553.

Brown, N. R. (1997). Exemplar presentation memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 898-914.

Brown, N. R. (2002). Encoding, representing, and estimating event frequencies: A multiple strategy perspective. In P. Sedlmeier and T. Betsch (Eds.), *Etc. frequency processing and cognition,* (pp. 37-66). New York: Oxford University Press.

Brown, N. R., & Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *The Journal of Sexual Research, 36,* 292-297.

Conrad, F. G., Brown, N. R., & Dashen, M. (2003). Estimating the frequency of events from unnatural categories. *Memory & Cognition, 31*, 552-562.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37-60.

Dougherty, M.R.P, & Franco-Watkins, A.M. (2003). Reducing bias in frequency judgment by improving source monitoring. *Acta Psychologica, 113,* 23-44.

Dougherty, M. R. P., & Hunter, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity, *Acta Psychologica.*

Engle, R. W. (1996). Working memory and retrieval: An inhibition-resource model (pp. 89-119). In J. T. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, R. T. Zacks (Eds.), *Working memory and human cognition*. New York: Oxford University Press.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex (pp. 102-134). In A. Miyake and P. Shah (Eds.), *Models of working memory: Mechanism of active maintenance and executive control.* Cambridge, UK: Cambridge University Press.

Flexser, A. J., & Bower, G. H. (1975). Further evidence regarding instructional effects on frequency judgments. *Bulletin of the Psychonomic Society, 6,* 321-324.

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

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 3,* 546-567.

Greene, R.L. (1986). Effects of intentionality and strategy on memory for frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 489-495.

Greene, R.L. (1988). Generation effects in frequency judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 298-304.

Hanson, C. & Hirst, W. (1988). Frequency encoding of token and type information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 289-297.

Hart, J. T. (1965). Memory and feeling-of-knowing experience. *Journal of Educational Psychology, 56,* 208-216.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356-388.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information. *American Psychologist, 39,* 1372-1388.

Hintzman, D.L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers, 16,* 96-101.

Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple trace model. *Psychological Review, 95,* 528-551.

Hintzman, D.L., & Stern, L. D. (1978). Exemplar presentationual variability and memory for frequency. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 539-549.

Howell, W. C. (1973). Storage of events and event frequency. *Journal of Experimental Psychology, 98,* 260-263.

Kane, M. J., Bleckley, M. K., Conway, A. R. A, & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169-183.

Kirk, R. E. (1995). Experimental design: Procedures for the behavioral sciences (3^{rd} ed.). Pacific Grove, CA: Brooks/Cole.

Kucera, H., & Francis, W. N. (1967). Computational Analysis of Present-day American English. Providence, RI : Brown University Press.

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.

Naveh-Benjamin, M., & Jonides, J. (1986). On the automaticity of frequency coding: Effects of competing task load, encoding strategy, and intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 378-386.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Pavio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology, 80,* 279-285.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510-532.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126,* 211-227.

Sanders, R.E., Gonzalez, E.G., Murphy, M.D., Liddle, C.L., & Vitina, J.R. (1987). Frequency of occurrence and the criteria for automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 241-250.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Lawrence Erlbaum.

Tversky, A,, & Kahneman, D. (1974, September). Judgment under uncertainty: Heuristics and biases. *Science, 165*, 1124-1130.

Watkins, M. J., & Le Compte, D. C. (1991). Inadequacy of recall as the basis for frequency knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1161-1176.

Williams. K.W., & Durso, F.T. (1986). Judging category frequency: automaticity or availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 387-396.

Zacks, R. T., Hasher, L., & Sanft, H. (1982). Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 106-116.