ABSTRACT

Title of Document: The Role of Part-Set Cuing and Retrieval Induced Forgetting in Subjective Probability Judgments

Tracy Darlene Tomlinson, Master of Science, 2007

Directed By: Professor Michael Dougherty, Psychology Department

A fundamental assumption of support theory is that unpacking an implicit disjunctive hypothesis into its component hypotheses increases its perceived likelihood compared to ratings of the implicit disjunction (Tversky & Koehler, 1994). However, recent work by Sloman et al. (2004) revealed that cuing participants with atypical exemplars from a category led to *decreases* in perceived likelihood. Three interpretations of this typicality effect are reviewed and three experiments are reported that examine these interpretations. Experiment 1 replicated the Sloman et al. (2004) findings but the generation data indicate that the judgment results may be due to a misinterpretation of the question. Experiment 2 adapted the retrieval-induced-forgetting paradigm and found that unpacking the implicit disjunction is affected by retrieval inducement processes, and the subjective probability judgments may be better accounted for by an averaging model. Experiment 3 indicates that these typicality effects are not observed within small judgment sets.

The Role of Part-Set Cuing and Retrieval Induced Forgetting in Subjective Probability Judgments

By

Tracy Darlene Tomlinson

Thesis 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
Master of Science
2007

Advisory Committee:
Professor Michael Dougherty, Chair
Professor Thomas Wallsten
Professor David Huber

Acknowledgements

I would like to thank Michael Dougherty for his countless hours of input and insight and for being a wonderful collaborator and advisor. I would also like to thank Thomas Wallsten and Dave Huber for their support and wonderful feedback.

Amber Sprenger also deserves my gratitude and acknowledgement for both her theoretical and methodological contributions, but also for her friendship and support throughout my graduate experience.

Harriet Tomlinson must be acknowledged as the reason my writing is not the grammatical nightmare that I know I am capable of producing and for being a wonderfully attentive and inquisitive friend and mother-in-law. Randy and Harriet Tomlinson both have my gratitude and love for their truly amazing emotional advice and support.

My parents, Doug and Becky Pickels are the reason I applied myself in school and it is to them that I owe my initial academic interest, but most importantly, my parents have never once faltered in giving me all of their love and support in every possible way.

I would like to especially thank my husband, Scott. Thank you for continually making me laugh, for providing me with every possible support, and for ultimately showing me a divinely inspired love.

Lastly, I am humbled to be a part of God's plan and am thankful that He has lead, provided, and loved me throughout my life and through this journey.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Subjective Probability Judgments and Unpacking

A normative assumption of subjective probability judgments is that of description invariance: The description of an event should not affect the judged probability of that event. However, there is empirical evidence that this normative assumption is often violated (Fischhoff, Slovic, & Lichtenstein, 1978; Tversky & Kahneman, 1983; Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). For instance, Rottenstreich & Tversky (1997) found that participants judged the probability that a randomly selected death was due to "homicide by an acquaintance or stranger" to be greater than "homicide" alone. When the implicit disjunction (homicide) is unpacked into an explicit disjunction by making exemplars of homicide explicit (homicide by an *acquaintance* or *stranger*), the judged probability of the event increased. The findings of Rottenstriech & Tversky (1997) are representative of a number of studies on judgments of disjunctive events (Brody, Coulter, & Daneshfar, 2003; Fischhoff, Slovic, & Lichtenstein, 1978; Tversky & Kahneman, 1983; Tversky & Koehler, 1994).

Recent research has explored the cognitive underpinnings of unpacking, and has begun to identify the boundary conditions of the general finding that unpacking an implicit disjunction hypothesis leads to an increased judged probability. These boundary conditions challenge the traditional interpretation of the unpacking effect. One particular set of studies relevant to such boundary conditions showed an explicit disjunction's judged probability to be *less* than the corresponding implicit disjunction (Sloman et al., 2004). When the implicit disjunction was explicitly unpacked with exemplars that were atypical of the event, the judged probability of the event decreased as compared to the corresponding implicit disjunction being unpacked with exemplars that were typical of the event and the implicit disjunction itself. For instance, participants judged the packed event that a randomly selected person in the United States would die from a disease to be equivalent to the typically unpacked event that a randomly selected person in the United States would die from "heart disease, cancer, stroke, or any other disease." Yet, the atypically unpacked event that a randomly selected person in the United States would die from "pneumonia, cirrhosis, diabetes, or any other disease," was judged to be significantly less probable than both the packed and the typically unpacked events. This empirical result from Sloman et al. (2004) is in clear opposition to the standard disjunction results.

The aim of the current paper is to test various theories that account for the effect of typicality on unpacking. I first provide a brief review of the dominant theory of subjective probability judgments, support theory, and discuss the failure of this model to account for typicality effects. I then review three alternative interpretations of the typicality effect on judgment: a modified version of support theory, a narrow interpretation of instances theory, and a misinterpretation hypothesis. Finally I present three experiments aimed at disentangling these various accounts.

Support Theory

Support theory proposes that subjective probabilities are based on descriptions of events, called hypotheses, rather than being based on events (Tversky & Koehler, 1994). Formally, support theory asserts that people make probability judgments by comparing the support for a focal hypothesis (A) with the support for an alternative hypothesis (B):

$$P(A,B) = \frac{s(A)}{s(A)+s(B)} \tag{1}$$

where s(A) and s(B) represent the summation of support for A and B respectively:

$$s(A) = \sum_{i=1}^{n} s(a_i) \tag{2}$$

$$s(B) = \sum_{i=1}^{n} s(b_i) \tag{3}$$

Events A and B each can be thought of as consisting of N elementary hypotheses (disjunctions) such that $\{a_1, a_2 \ldots a_N\} \in A$ and $\{b_1, b_2 \ldots b_N\} \in B$, where $N \geq 1$, and P(A, B) represents the probability of the focal hypothesis A occurring rather than the alternate hypothesis B occurring. Subjective probability is a function of the proportion of support that favors the focal hypothesis (A) rather than the alternative hypothesis (B). For example, suppose one is asked to estimate the likelihood that a person in the United States will die from a disease in the next year. According to support theory this judgment is made by comparing support for the focal hypothesis (e.g., death by disease) with the support for the alternate hypothesis (i.e., death by anything other than disease).[1]

[1] Support theory does not discuss retrieval processes used when retrieving items from the judgment set; it only addresses the comparison process. See Dougherty & Hunter (2003) for proposed retrieval processes.

A key distinction within support theory is the notion of packed (implicit disjunctions) versus unpacked (explicit disjunctions) hypotheses. A packed hypothesis is one in which the elements of the hypothesis are not explicitly stated. Conversely, an unpacked hypothesis is one in which some or all of the elements of the hypothesis set are explicitly stated. Tversky & Koehler (1994) argued that unpacking a packed hypothesis (e.g., death by disease) into its component hypotheses (e.g., death by cancer, heart disease, and other diseases) increases the perceived support of that hypothesis. Importantly, support theory anticipated that unpacking the numerator (A) in Equation 1 should lead to increases in judged probability, whereas unpacking the denominator (B) should lead to decreases in the judged probability. The perceived support for the unpacked focal hypothesis is expected to be equal to or greater than the packed hypothesis, and subsequently the judged probability of the unpacked hypothesis should be equal to or greater than the packed disjunction, which is what has been found empirically (Brody, Coulter, & Daneshfar, 2003; Mann, 1997; Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994).

Support theory explains violations of descriptive invariance, but it does not anticipate the typicality effects found by Sloman et al. (2004) without modification. At present support theory assumes an additive support function that is most clearly demonstrated by the denominator of Equation 1. Given this additive function, unpacking the numerator should lead to monotonic increases in judged probability, regardless of the support of the unpacked elements. However, in other domains, there is evidence that judgments often follow an averaging function rather than an additive function (Anderson, 1959; Anderson, 1991; Schlottman & Anderson, 1995; Shanteau, 1970; Shanteu, 1972). Thus, there is some basis for the hypothesis that perceived support is based on the average support of the unpacked hypotheses rather than an additive function.

Central Tendency Model

A straightforward modification of support theory would be to assume that the support values are a function of the central tendency of support rather than summation of support. The most commonly used measure of central tendency, the mean, will be the focus of this research as an average support function would predict the typicality results. An averaging model would have the support values being averaged instead of summed, as represented in equations 4 and 5:

$$\overline{s(A)} = \frac{\sum_{i=1}^{n} s(a_i)}{n} \qquad (4)$$

$$\overline{s(B)} = \frac{\sum_{i=1}^{n} s(b_i)}{n} \qquad (5)$$

where n is the number of unpacked hypotheses.

Assuming support is a function of the average, instead of the sum, enables support theory to account for the effect of typicality on judgment. If all generated items are of equivalent support, then mean support would be independent of the number generated. However, we assume that people generate items of varying support, and the more items that are generated the more likely it is that people will generate low-support items. The averaging model then leads to the counter-intuitive prediction that judgments should decrease as a function of the number of alternatives generated since the mean support should decrease as the number of items generated increases. This would then predict that the number of hypotheses generated should be negatively correlated with mean support. In contrast, the sum of the support for the hypotheses generated should be positively correlated with the number of hypotheses generated. Moreover, if judgments are based on an averaging model, we would expect that judgments of probability should be positively correlated with mean support of items generated, but if they are based on an additive model then judgments should be positively correlated with the sum of support of items generated.

Narrow Interpretation Theory

Another interpretation of the typicality results that builds upon support theory is offered by Sloman et al. (2004), who propose a modified theory of subjective probability referred to as the narrow interpretation theory. The narrow interpretation theory claims that "people interpret category descriptions narrowly, in terms of typical instances" (Sloman et al., 2004). This is similar to support theory, which claims that the category instances that are included in the judgment set and comprise the instances of perceived support are primarily based on the most representative category instances (Tversky & Koehler, 1994). Sloman et al. (2004) further posit that the most representative instances tend to be the most typical instances. Typical category instances may be the most representative of the category due to a common occurrence of these instances and a high similarity between these and other category instances (Hampton, 1998). The similarity with other category instances should aid in memory retrieval of typical instances, and the higher frequency of occurrence should serve to increase the assessment of support. Thus typicality is theorized to positively correlate with support, though exceptions do exist (Sloman et al. 2004).

The narrow interpretation theory and support theory appear to be in agreement that the category instances that are retrieved are the basis of support judgments, and that these retrieved instances are usually the most representative instances of the category being judged. Sloman et al. (2004) elaborate by claiming that the most representative instances are the typical instances, that typicality and support are correlated, and then they move beyond support theory by claiming that unpacking is not a sufficient condition to increase subjective probability judgments.

The narrow interpretation theory and support theory also agree that instances, or cues, serve to capture the judge's attention, but the narrow interpretation theory claims that this focus will cause the judgment to be closer to a judgment of the support of the unpacked cues rather than the entire category. The operating mechanism for this focusing effect on judgment is not made explicit within the narrow interpretation theory. However, part-set cuing effects, inhibition effects, and anchoring and adjustment are hypothesized as being the most likely mechanisms underlying the focusing effect assumption that leads probability judgments to be proportional to the support of the explicitly unpacked instances.

As a potential explanation of the narrow interpretation theory, the part-set cuing effect may be the most likely given the methodology of Sloman et al. (2004). Part-set cuing is the non-intuitive finding that providing participants with a few exemplars of a set of to-be-retrieved items can actually decrease retrieval (Slamecka, 1968). In the part-set cuing task participants memorize a set of items and later attempt to recall the items in the set, with some participants receiving several example items from the studied set at the time of recall. Participants who receive example item cues at recall often showed poorer recall for the rest of the (non-presented) items in the set.

This part-set cuing effect can be found in episodic as well as semantic memory, in recall as well as recognition, and with intralist and extralist cues (Brown, 1968; Todres & Watkins, 1981; Watkins, 1975). There are several theories of part-list cuing, including retrieval competition, inhibition, and strategy disruption (Anderson, M., Bjork, R.A., & Bjork, E.L., 1994; Rundus, 1973; Basden & Basden, 1995). Although there is no clear evidence for one theory over another, part-set cuing is a relatively robust and well documented effect (Brown, 1968; Foos & Clark, 2000; Slamecka, 1968; Todres & Watkins, 1981; Watkins, 1975). Since Sloman et al. (2004) employ a methodology very similar to the part-set cuing paradigm, it is likely that the observed effects of typicality on judgment are the result of part-set cuing.

If there is an effect of part-set cuing or inhibition, the prediction would be that the category cues lead to an impairment of the judge to retrieve and consider other instances in the category. The cues provided are therefore more integral to the support function and determining the magnitude of the judgment, with subsequent probability judgments being proportional to the cues provided. In this manner, the narrow interpretation theory prediction that judgments are proportional to the support of the unpacked instances may occur through a part-set cuing or inhibition process.

Rather than a part-set cuing or inhibition effect, it may be that probability judgments are mediated by an anchoring and adjustment process within the narrow interpretation theory. The explicitly unpacked instances may serve as an anchor from which judges adjust insufficiently for other category instances. When the implicit disjunction is presented or when typical instances are presented, a judge would have a high anchor point and would then insufficiently adjust the judgment down. Conversely, when atypical instances are made explicit there will be a low initial anchor point and then insufficient upward adjustment.

Misinterpretation Theory

A less glamorous, albeit plausible, explanation of the typicality effects observed by Sloman et al. (2004) is that participants misinterpreted the probability question as asking them to judge the probability of events similar to the exemplars. The judges may then interpret the question differently depending on the category instances that are explicitly unpacked. For instance, if atypical category events are provided in the judgment question then judges might interpret that question to be asking them to judge the likelihood that an atypical event, such as the events listed, will occur. Likewise, if a typical category event is provided then the judge may assume that they are to judge the likelihood that a typical event, such as the listed events, will occur. Since the most typical category instances are assumed to be the instances more likely to be retrieved and considered when no category instances are provided, the implicit disjunction and the explicit disjunction with typical instances may be interpreted and judged in a similar manner: judging the likelihood of a typical event. Under the misinterpretation hypothesis, there is no reason to expect that judgments would be related to the number, summed support, or mean support of hypotheses generated. Moreover, cue typicality should have no effect on the number of hypotheses generated, summed support, or mean support.

The above explanations are all plausible interpretations of the effect of typicality on judgment. However, the experiments completed by Sloman et al. (2004) do not allow one to distinguish amongst these various accounts. Sloman et al. (2004) employ a methodology that is very similar to the traditional part-set cuing paradigm, but they did not examine whether these cues had concomitant effects on retrieval. Such retrieval data is important for discriminating amongst the various interpretations outlined above. For example, a part-set cuing or inhibition account of the effect of typicality would anticipate that participants who are cued with typical or atypical exemplars would retrieve fewer hypotheses than participants in a no-cue condition. Additionally, a follow-up question that could be addressed by retrieval data is whether participants base their assessment of support on an additive model or an averaging model: An additive model would predict higher judgments with increased number of hypotheses while an averaging model would predict lower judgments with increased number of hypotheses retrieved.

Hypotheses and Predictions

Including a generation task in Sloman et al.'s (2004) modified part-set cuing paradigm should enable an assessment of what items are being included in the judgment set that is the basis of the subjective judgments, and subsequently which interpretation best captures the judgment data. In Experiment 1, participants are given a judgment task identical to that used in Sloman et al. (2004), following their judgment they are asked to generate items from the set they judged. Since the generation phase occurs immediately after the judgment phase it is plausible that the items generated are the items that were used in making their probability judgment.

Generation data enables an initial assessment of whether part-set cuing or inhibition effects are present by comparing how many items are generated when category cues are present versus not present. Part-set cuing and inhibition effects would manifest as fewer items being generated when a cue is provided as compared to a non-cued condition, regardless of the cue typicality. If cued and non-cued conditions have an equivalent number of items generated, or if one cue condition generates more items than the other cue condition, this would be evidence against the part-set cuing and inhibition.

Generation data, along with an independent measure of support values for the generation data, also enables an assessment of what type of support function best fits the data. This assessment of support function may occur through observing which function trends in the same manner as the judgment data. If an averaging function is being used then the average support of items generated when atypical cues are presented should be less than the average support of items generated when typical or no cues are presented. The average support of items generated when no cues or when typical cues are provided should be equivalent. Similarly, if an additive function is being used then the summation of items generated should follow the same trends as the judgment data.

It could be the case that both the average support and the summation of support of items generated have similar trends, in which case neither function could be differentiated as driving the judgment data. Conversely, the summation of support of items generated may not trend in the same manner and only one function may trend in the same manner as the judgment data. This would lead to a convincing case for either function if the summation and the average support of items generated diverged with only one function mapping onto the judgment data.

Another alternative is that the generation data does not correspond to the judgment data. If this is the case then it might be an indication that the generation data are not representative of what is being retrieved and judged, and that the judgment effects are due to a misinterpretation of the question.

Chapter 2: Experiment 1

Introduction

Sloman et al.'s (2004) results appear to support their narrow interpretation theory through a part-set cuing effect or inhibition, or an anchoring and adjustment process, but it could also indicate that people are using an averaging function to assess support for the focal hypothesis, or that they simply misinterpreted the probability question. Experiment 1 aimed to replicate the judgments results of Sloman et al. (2004) and extend their methodology by asking participants to explicitly unpack the focal hypothesis after making their subjective judgment. This generation task was implemented in order to assess if there is an effect of part set cuing or inhibition that may underlie the narrow interpretation theory, and to assess if an additive or averaging model best fits the judgment data.

Support values for the category items that were generated were obtained by an independent sample of participants who had not made any judgments on the category or generated any items from the category. These support values were then used to assess whether the obtained judgment data could be a function of the generation data through either an additive or averaging function.

Participants

Ninety-one University of Maryland undergraduate students participated and received extra credit for a psychology course.

Materials

The experiment was completely computerized and was programmed in Media Lab. The judgment question was taken verbatim from Sloman et al. (2004) and the generation question was revised from the judgment question to ask for generation rather than a judgment.

Design and Procedure

The methodology for Experiment 1 was taken from Sloman et al. (2004), whereby each participant was asked to make one judgment. The only modification to the methods was the addition of a generation task after the judgment that asked participants to explicitly unpack the focal hypothesis for the previous judgment.

The study is a between subjects design with three conditions: packed, atypical unpacked exemplars (atypical condition), and typical unpacked exemplars (typical condition). Participants were asked to provide a judgment to the following question, taken from Sloman et al. (2004):

> Consider all the people that will die in the U.S. next year. Suppose that we pick one of these people at random. Please estimate the probability that this person's death will be attributed to the following causes.

Participants in the packed condition were asked to judge the probability of the hypothesis "*disease*" (n = 30). Participants in the typical condition (n = 30) were asked to judge the probability of the same hypothesis, but were provided with three exemplars of the most common disease that people die from, along with the residual hypothesis: "*heart disease, cancer, stroke, or any other disease*." Likewise, participants in the atypical condition (n = 31) were provided with three exemplars that are fairly uncommon causes of diseases that people may die from, along with the residual hypothesis: "*pneumonia, diabetes, cirrhosis, or any other disease*."

Immediately following the judgment, participants were asked to unpack the focal hypothesis (i.e., diseases that a person in the United States may have died from last year). The judgment question was modified to ask for the participant to generate category instances rather than a judgment:

> Consider all of the people that will die from a disease in the U.S. next year. Suppose that we select one of these people at random. Please list the possible diseases this person could have died from.

The participants were limited to providing a maximum of twenty items for the generation task. The two tasks took approximately five minutes to complete.

Support Functions

An independent sample of thirty-five University of Maryland undergraduate students who had not taken part in Experiment 1, rated the support of items generated in Experiment 1. Participants received extra credit for a psychology course.

The items generated (*n*=141) from Experiment 1 were compiled into one list and were used to obtain independent support values. Each participant rated each of the 141 items that were generated on a 1-7 typicality scale with 1 being not at all typical and 7 being extremely typical. The participants were instructed:

> In this study you are going to be asked to rate possible diseases that may cause death... When rating the disease, please rate it in terms of how typical it is of a disease that people in the US died from last year. You will be asked to rate the disease on a 1 to 7 scale, with 1 being "Not at all Typical" and 7 being "Extremely Typical".

The ratings provided were then averaged across participants to create an average typicality score for each item that was generated. These average typicality scores were used as support values for each item.

Results and Discussion

Generation Data

There was no main effect of cue typicality on the number of items generated (see Figure 1): The three conditions did not significantly differ in the number of category items generated after providing a judgment.[2] This indicates that the category cues provided in the judgment question did not serve to decrease the number of items that were generated. Thus, based on the generation data, there is no evidence that providing participants with typical or atypical category cues led to part-set cuing or inhibition effects.

Support Values

There was also no main effect of cue typicality on the summation of support of the items generated (see Figure 2).[3] There were no significant differences between the atypically cued, typically cued, and packed conditions summation of support of the items generated.

[2] The generation data was compiled into three separate generation lists: One generation list consisted of the instances exactly as they were generated by participants, a second list had all the cues removed from the generation data, and a third list had all the cues included in the generation of instances, regardless of whether or not the participant had explicitly listed the cues in their generation data. All analyses yielded similar results and so only the cue inclusion list will be reported for Experiment 1.

[3] Each category instance that a person generated had a support value obtained from an independent sample of raters, and these values were summed for each participant so that each participant had a summation support strength score of items they generated.

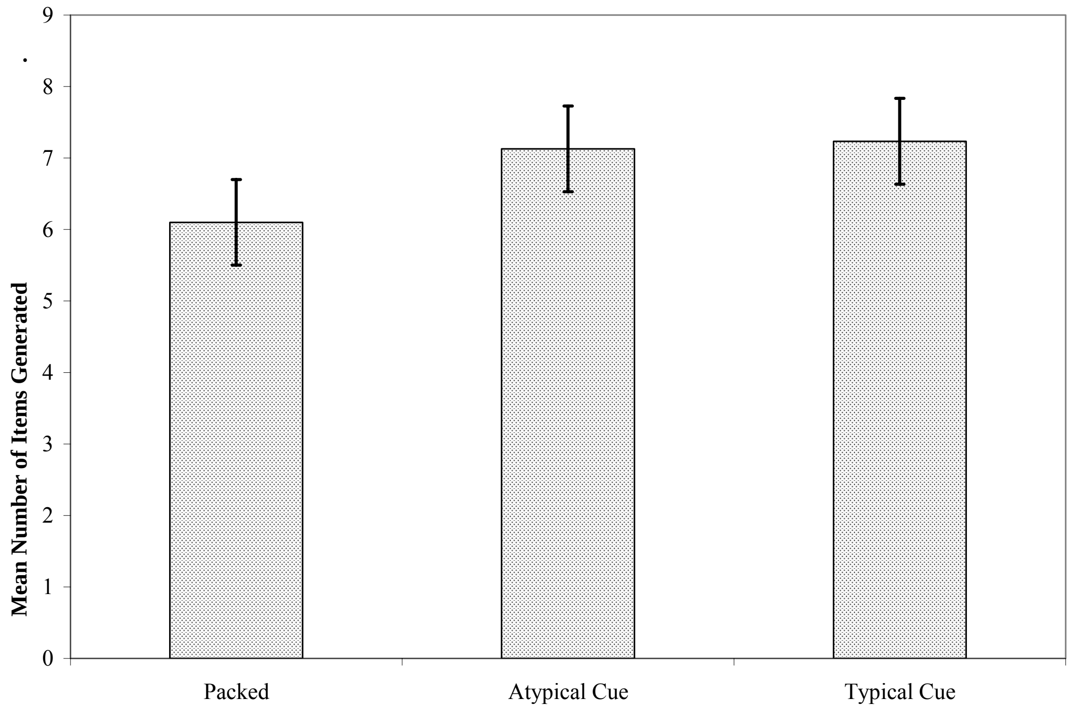

Figure 1. Experiment 1 Mean Number of Items Generated by Part-Set Cue Typicality with Standard Error Bars

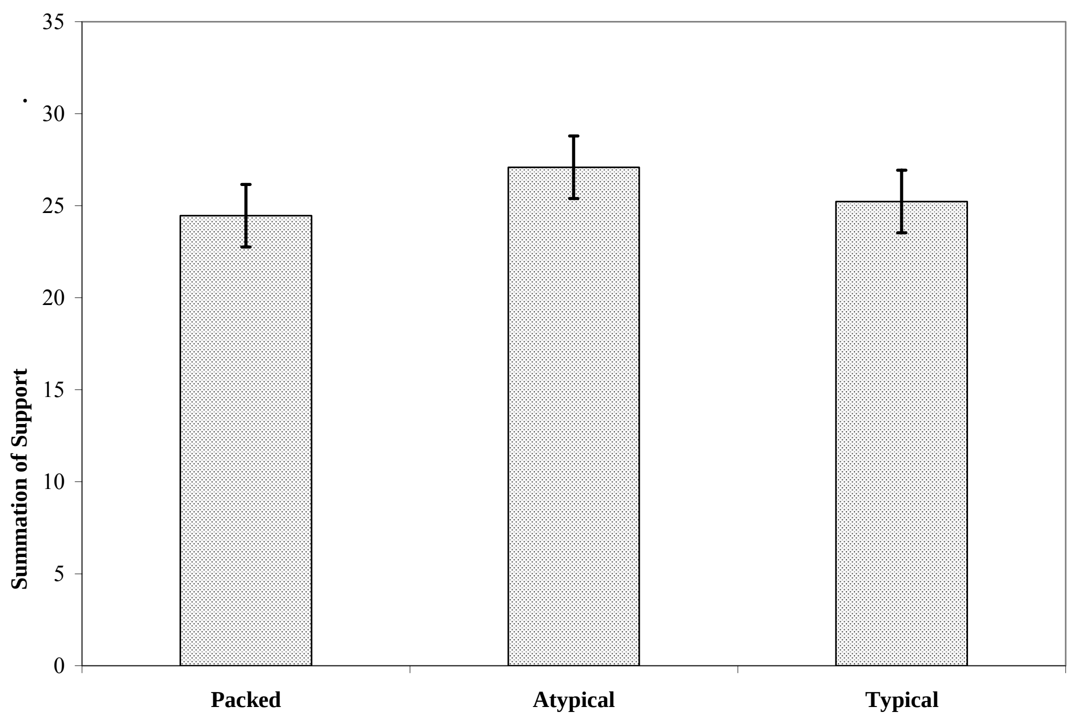

Figure 2. Experiment 1 Summation of Support of Items Generated by Part-Set Cue Typicality with Standard Error Bars

Additionally, an average strength of support score for items generated was obtained for each participant by averaging the support values for each item generated by each participant. There was a main effect of cue typicality on average strength of support of items, $F(2, 88) = 3.13$, $p<0.05$ (see Figure 3). However, there were no significant post-hoc differences between any of the cue conditions. At most, this result would indicate that an averaging model may predict an overall effect but no significant differences between conditions.

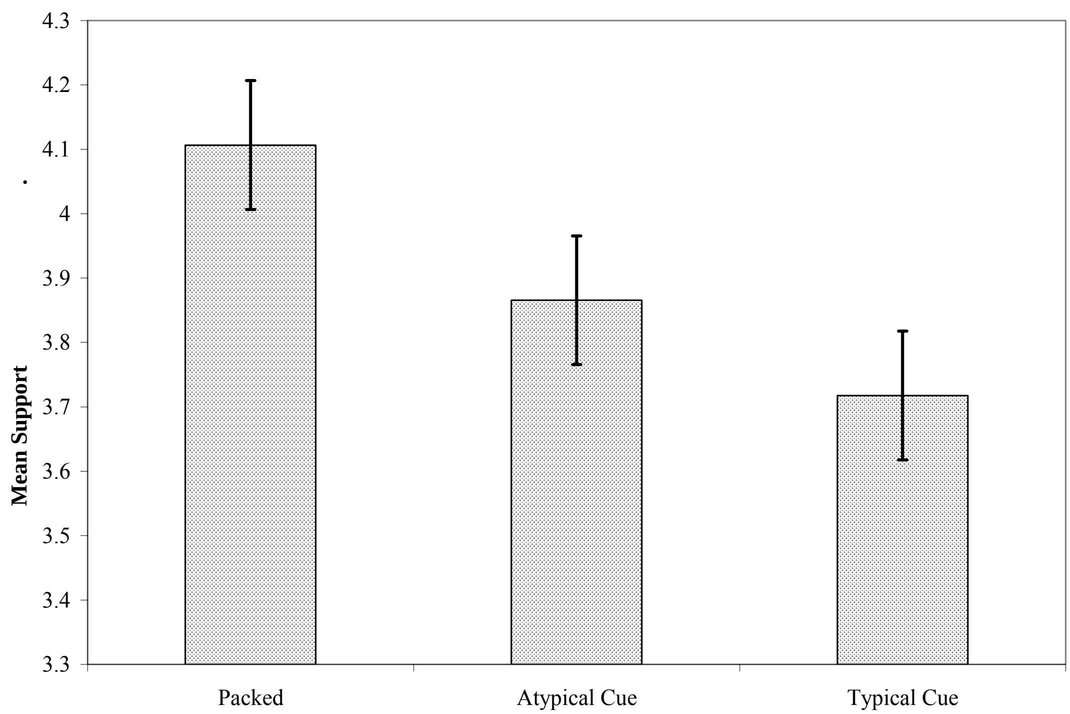

Figure 3. Experiment 1 Average Support of Items Generated by Part-Set Cue Typicality with Standard Error Bars

Taken together the summation and mean support of items generated do not give a clear picture of which, if either, of these two functions may underlie the participant's probability judgments. Since the manipulation of cue typicality had no effect on either the summation and a limited effect on the average support, both the additive and the averaging models would predict no effect of cue typicality on judgments.

Judgments

The judgment results replicated Sloman et al. (2004) and found a main effect of cue typicality on probability judgment $F(2, 88) = 4.26$, $p < 0.05$ (see Figure 4): The typical and packed conditions mean judgments were not significantly different, while the atypical conditions mean judgment was significantly less than both the packed and typical conditions judgments (see Table 1). Cue typicality had an effect on subjective probability judgments such that when atypical instances are explicitly provided the probability judgments are significantly less than when typical instances are provided or no instances are provided.

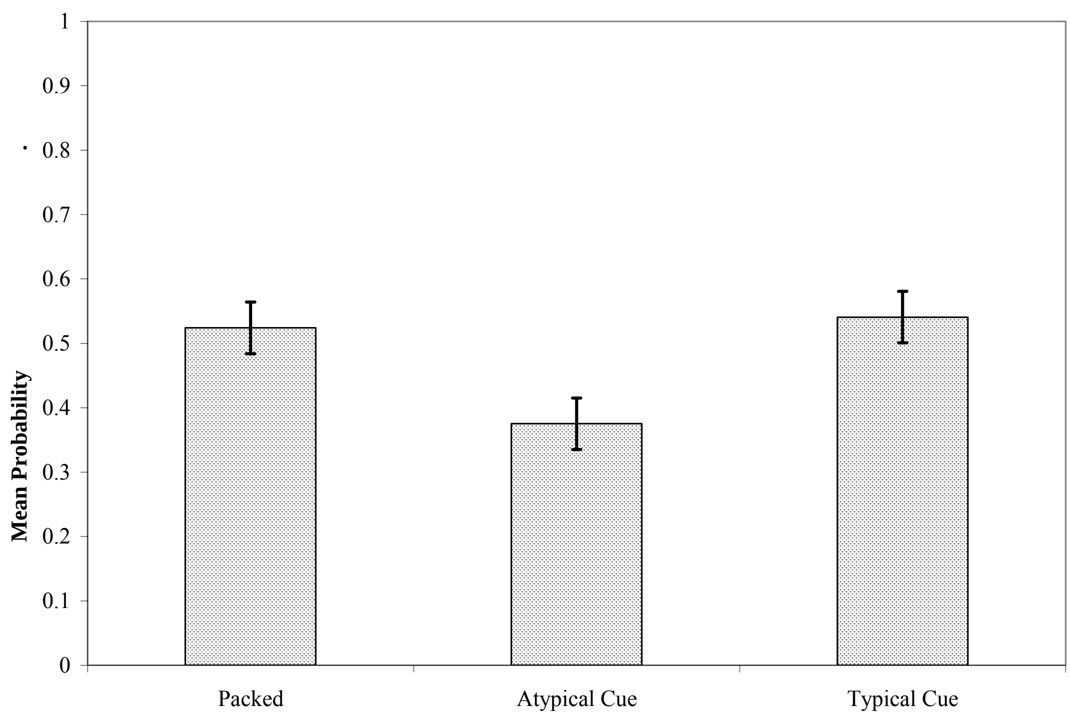

Figure 4. Experiment 1 Mean Probability Judgments by Part-set Cue Typicality with Standard Error Bars

	Packed	Atypical	Typical
Judgment Mean (Std. Deviation)	0.52 (0.22)	0.38 (0.26)	0.54 (0.25)
Number of Items Generated (SD)	6.10 (2.19)	6.94 (3.95)	7.03 (4.11)
Summation of Support of Items (SD)	24.60 (7.56)	26.11 (11.94)	35.93 (12.03)
Mean Support of Items (SD)	4.18 (0.63)	3.63 (1.05)	3.64 (1.16)

Table 1. Experiment 1 Means (and Standard Deviations)

There were no significant correlations between judgment and the number of items, summation, or average support of items generated (see Table 2). However, the correlations were based on 30 participants and thus may have low power for detecting potentially small but significant correlations between judgments and any of the generation data.

There was a significant positive correlation between the number of items generated and the summation of support of items generated for all typicality conditions (see Table 2). Consistent with our assumption that increased unpacking leads to an increased likelihood of unpacking atypical items, there was a significant negative correlation between the number of items generated and the average support of items generated for all typicality condition (see Table 2).

The correlation results indicate that as more items are generated the summation of support increases while the average support decreases. Additionally, the correlations between the summation and the average support of items generated were negative in all typicality conditions, but were only significant in the typical condition. These correlation results may be due to increased unpacking leading to more atypical items being included in the generation data, which is again consistent with our previous assumption of increased unpacking leading to more atypical items being retrieved. These results also indicate that the summation and average support of items generated may be diverging: Increasing the summation of support occurs while the average support of items generated decreases.

	No Category Cue Condition			
	Judgment	Number of Items	Summation of Support	Average Support
Judgment	1	-.06	-.10	.02
Number of Items	-.06	1	.91**	-.66**
Summation of Support	-.10	.91**	1	-.34
Average Support	.02	-.66**	-.34	1
	Atypical Category Cue Condition			
Judgment	1	.00	.09	.12
Number of Items	.00	1	.97**	-.44*
Summation of Support	.09	.97**	1	-.22
Average Support	.12	-.44*	-.22	1
	Typical Category Cue Condition			
Judgment	1	.03	-.08	-.02
Number of Items	.03	1	.92**	-.86**
Summation of Support	-.08	.92**	1	-.65**
Average Support	-.02	-.86**	-.65**	1

*p < .05. **p < .01

Table 2. Experiment 1 Correlations by Typicality: Judgment, Number of items Generated, Summation of Support of Items Generated, and Average Support of Items Generated

To further test whether the effect of cue typicality on judgment was due to the generation of hypotheses, a series of covariate analyses were conducted. The covariate analyses were performed to test whether the effect of cue typicality remained significant after controlling for variance due to the number, the summation of the support, and the mean support of the items generated. None of these factors were significant covariates and in all cases the effect of cue typicality remained significant. This indicates that the number, the summation of support, and the mean support of items generated are not predictive of judgments. Taken together these results are inconsistent with an additive and averaging support theory, but the results still are consistent with the anchoring and adjustment and misinterpretation hypotheses.

In sum, the results from Experiment 1 replicated Sloman et al.'s (2004) judgment results with atypically cued judgments being significantly lower than the typically cued and non-cued packed conditions. However, the generation data from Experiment 1 do not show any effects of cue typicality on the quantity of items generated after making the category judgment. Since the atypical and typically cued conditions do not show a decrease in the number of items generated as compared to the non-cued condition, this suggests that there is no part-set cuing or inhibition affecting the judgment or generation processes. It would appear that neither part-set cuing nor inhibition are the underlying mechanisms of the narrow interpretation theory. Furthermore, the summation and average support of items generated do not differ by cue typicality, which indicates that neither an additive nor an averaging support function is being used in making the probability judgments. Alternately, it may be indicative that a misinterpretation of the question or an anchoring and adjustment process is driving the judgment results.

Chapter 3: Experiment 2

Introduction

Experiment 2 sought to incorporate another memory paradigm into the judgment literature to enhance our understanding of the interplay between memory and judgment, while simultaneously minimizing potential misinterpretations of the question or anchoring and adjustment processes. In the modified part-set cuing paradigm used in Experiment 1 and by Sloman et al. (2004), the category cues and the judgment question are intertwined. This may increase the possibility of a misinterpretation of the question since the cues are present in the judgment question. Experiment 2 uses a modified retrieval-induced-forgetting paradigm that allows the category cues to precede the question so that all typicality conditions can receive the same judgment question.

Retrieval induced forgetting is similar in effect and paradigm to part-set cuing: The process of retrieving some items appears to cause forgetting of other items. The retrieved item will be facilitated and more likely to be recalled again later, but other items that are associated with the same retrieval cue as the retrieved item are more likely to be forgotten (Anderson, 2006; Anderson, 2005; Anderson, Bjork, & Bjork, 200; Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995). Much like part-set cuing there are various theories for the retrieval induced forgetting effect. Two of the dominant theories are the competitor strengthening theory whereby the recalled item results in a strengthened association with the retrieval cue that weakens the other items related to that cue (Blaxton & Neely, 1993; Brown, 1981), or the inhibition theory whereby a suppression of all non-recalled related words occurs during selective retrieval (Anderson 2006; Anderson, 2005; Anderson 2003; Anderson, Bjork, & Bjork, 1994; Anderson, 2004; Levy & Anderson, 2002). Though the exact causal mechanism is under debate, the effect is well documented and the paradigm may be well suited to assess judgment as well as memory (Anderson, 2001; Anderson, Bjork, & Bjork, 2000; Anderson, Bjork, & Bjork, 1994; Anderson, 2004; Anderson, Green, & McCulloch, 2000; Anderson & McCulloch, 1999; Anderson & Spellman, 1995).

In the typical retrieval-induced-forgetting paradigm, participants are presented with several items from a category and instructed to memorize these items. They subsequently are presented with the category and word-stem cues for half of the category items as retrieval cues. This retrieval practice leads to improved recall for the cued items and impairs recall of the non-recalled items in a cued-recall test as compared with a category condition where there is no retrieval practice (Anderson, Bjork, & Bjork, 1994; Anderson & Spellman, 1995).

This memory paradigm can be modified in the same manner as the part-set cuing paradigm was modified to become a judgment paradigm: Assume a pre-experimental semantic set exists so that there is no experimental learning phase, and proceed immediately to "cue" the participants with items from the set by having them perform selective retrieval. Baüml (2002) demonstrated that retrieval inducement can occur even for pre-experimental knowledge, or knowledge of items not explicitly studied (see also, Anderson & Bell, 200; Johnson & Anderson, 2004).

If items from the pre-experimental category set are selectively retrieved then the cues would *precede* the judgment question, allowing the judgment question to be the same implicit disjunction for all judges. With all judges receiving the same judgment question the likelihood of a misinterpretation of the question is decreased as well as anchoring and adjustment process since there are no cues in the judgment question.

As in Experiment 1, a generation task followed the judgment question. This allowed us to test if there were part-set cuing or inhibition effects on memory, and whether an additive or averaging model best explained the judgment data.

Participants

Forty-nine University of Maryland undergraduate students participated in the study. Participants received extra credit for a psychology course. Two participants did not correctly retrieve any of the category cues and their data were subsequently not used in any judgment or generation analyses.

Materials

The experiment was completely computerized and was programmed in Revolution. There were two inducement cues for each condition: atypical inducement condition (i.e., diabetes and tuberculosis) and typical inducement condition (i.e., heart disease and cancer). These cues were taken from the atypical and typical instances in the cued judgment question from Sloman et al. (2004), with the exception of one of the atypical cues. One atypical cue was replaced from the original cues since a small pilot study found that cue to be too difficult for people to retrieve. The cue used (tuberculosis) was more likely to be retrieved and still seemed to be an atypical item representing a disease that causes death in the United States. The judgment question was taken verbatim from Sloman et al.'s (2004) packed judgment question. The generation question was revised from the judgment question to ask for generation rather than a judgment, and was the same generation question from Experiment 1.

Design and Procedure

The study was a between subjects design with two conditions: atypical cue inducement (atypical condition), and typical cue inducement (typical condition). Two cues were used in both the atypical (diabetes and tuberculosis) and the typical (heart disease and cancer) condition. There were three phases to the experiment: a retrieval inducement, a judgment and a generation phase.

Since there was no learning phase the inducement phase relied on an assumed pre-experimental semantic set of diseases. In the inducement phase the category name (disease) was presented along with the first two letters of the cue that was to be retrieved and participants were instructed to fill in the blank with an appropriate word. For example, in the typical condition a participant might see "Disease – Ca________" and would be asked to fill in the blank (i.e., cancer).

For the judgment phase participants were asked the packed probability judgment question from Sloman et al. (2004) and Experiment 1. Immediately following the judgment, participants were asked to unpack the focal hypothesis. The generation instructions were the same instructions used in Experiment 1. There was no maximum number of items that the participants could generate and the three phases took approximately five minutes to complete.

Support Functions

Support values were obtained for all of the generation data from the same sample, procedures, and items used in Experiment 1. The raters had not participated in either Experiment 1 or Experiment 2.

Results and Discussion

Generation Data

A main effect of cue typicality on the number of items generated was obtained $F(1, 45) = 14.07$, $p < 0.001$ (see Figure 5).[4] The atypically cued condition generated significantly more items than the typically cued condition (see Table 3). The atypical inducement lead participants to generate more items from the category. This result appears inconsistent with a part-set cuing effect: There were no cues present in the judgment question and so there would be no expectation that the generation data for the two conditions would differ in the number of items generated.

Support Values

There was a main effect of cue typicality on the summation of the support of the items generated $F(1, 45) = 5.71$, $p<0.05$ (see Figure 6). The participants who retrieved atypical category instances had a significantly *higher* summation of support than the participants who retrieved typical category instances (see Table 3).

[4] The exact generation list was used in all analyses since the cues were not present in either the judgment or the generation question.

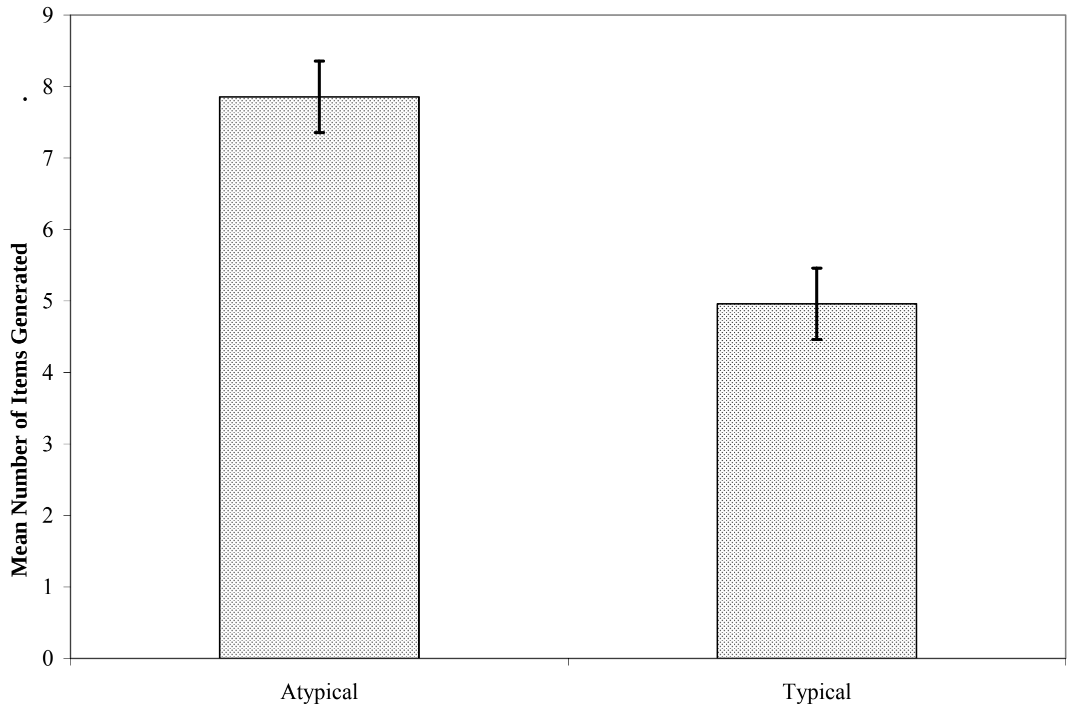

Figure 5. Experiment 2 Mean Number of Items Generated by Retrieval Inducement Cue Typicality with Standard Error Bars

	Atypical	**Typical**
Judgment Mean (Std. Deviation)	0.41 (0.29)	0.56 (0.22)
Number of Items Generated (SD)	7.90 (3.10)	4.85 (2.43)
Summation of Support of Items (SD)	27.82 (10.81)	20.84 (9.2)
Mean Support of Items (SD)	3.58 (0.65)	4.30 (1.07)

Table 3. Experiment 2 Means (and Standard Deviations)

There was also a main effect of cue typicality on the average support of the items generated $F(1, 45) = 7.18$, $p < 0.05$ (see Figure 7). The atypical-cue condition had a *lower* average support of items generated than the typical-cue condition (see Table 2). Thus there are main effects for both the summation of support and mean support of items generated by cue typicality. The atypical-cue condition generated significantly more items than the typical-cue condition and so they have a significantly higher summation of support of items generated: The more items that are generated the greater the summed support of all the items. However, the atypical-cue condition has a significantly lower mean support of items generated than the typical-cue condition. This indicates that the atypical cue did not simply lead to more items being generated, but lead specifically to more items with lower support values being generated than the typical condition.

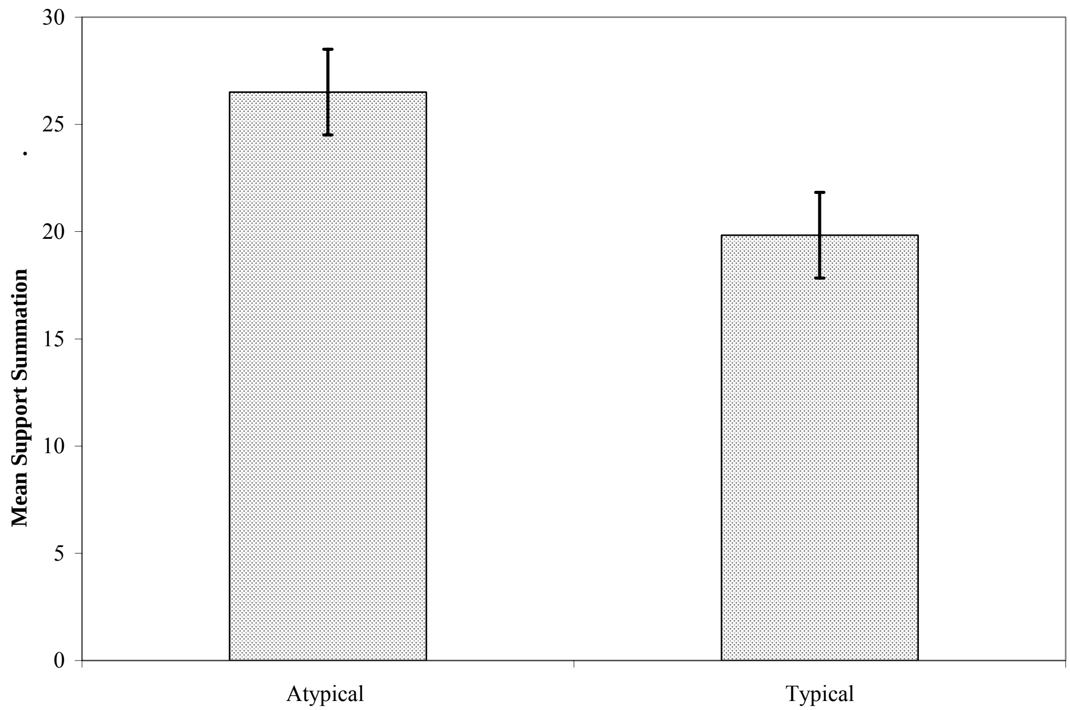

Figure 6. Experiment 2 Summation Support of Items Generated by Retrieval Inducement Typicality with Standard Error Bars

In sum, the atypical-cue condition has a higher summation but a lower average support of items generated than the typical-cue condition. These generation results provide differential predictions for the judgment data for the additive and averaging functions. If an additive function is being used to make judgments then the atypical-cue condition should have a higher judgment than the typically cued since the atypically cued condition has a higher summation of support. Conversely, if an averaging function is being used then the atypical-cue condition should have a lower judgment than the typical condition since the atypically cued condition has a lower average support. The generation data thus provide two opposing predictions for the judgment data based on the two different support functions.

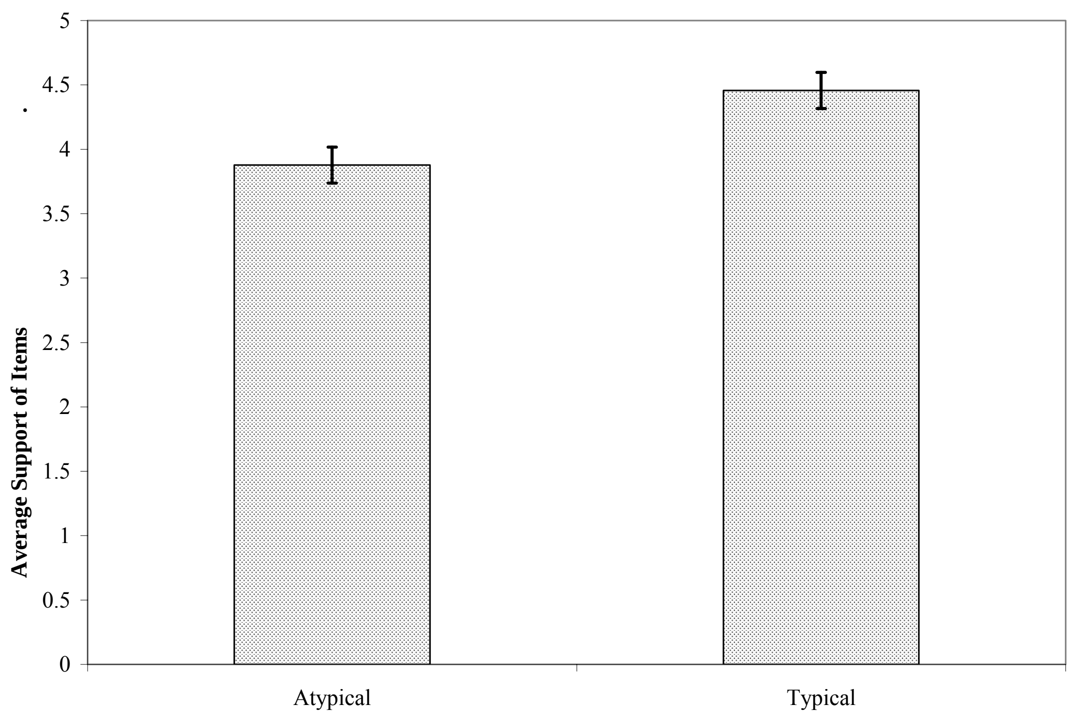

Figure 7. Experiment 2 Average Support of Items Generated by Retrieval Inducement Typicality with Standard Error Bars

Judgments

There is a main effect of cue typicality on judgments $F(1, 45) = 4.27$, $p < 0.05$ (see Figure 8), with judgments following atypical retrieval inducement being significantly lower than judgments following typical retrieval inducement (see Table 3). Even though both conditions received the exact same judgment question, they have significantly different judgments. If participants retrieve atypical instances from the category to be judged, this appears to decrease their subsequent probability judgment as compared to participants who retrieve typical category instances. This provides some support against a misinterpretation hypothesis or an anchoring and adjustment process since all participants judged the same question and still had significantly different judgments.

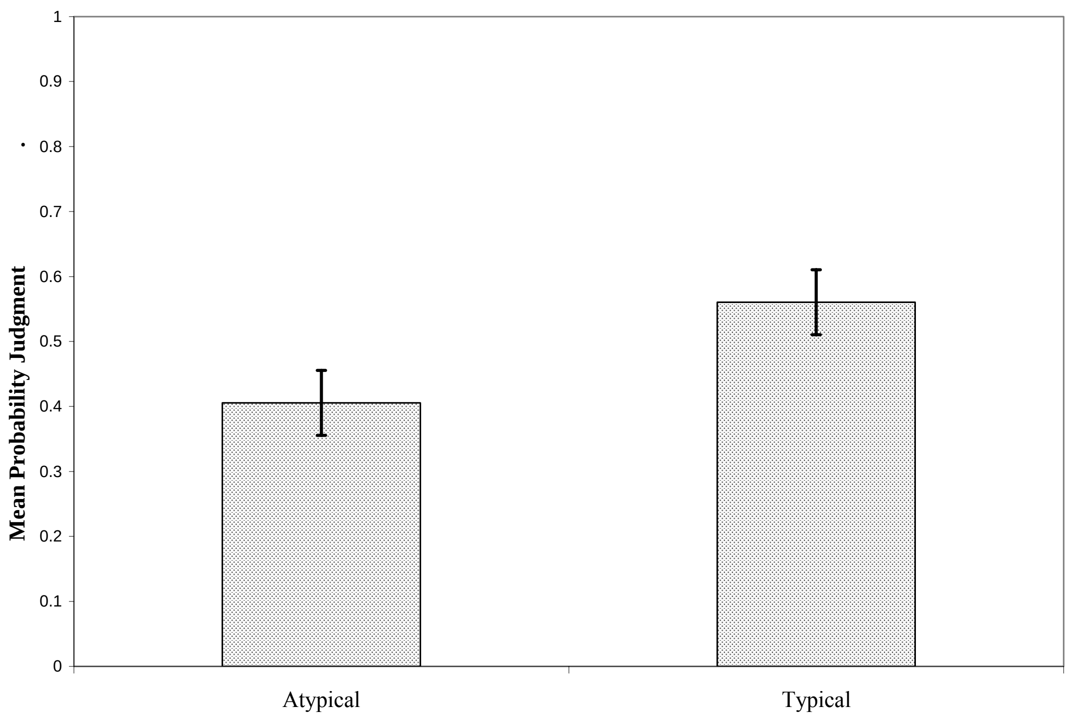

Figure 8. Experiment 2 Mean Probability Judgments by Retrieval Inducement Cue Typicality with Standard Error Bars

The atypical inducement led to a decrease in probability judgments compared to typical inducement, which is what the averaging model would predict from the generation data. Since the atypical-cue condition had a significantly lower average support of items generated the averaging model predicts the observed judgment data: a lower probability judgment for the atypical condition relative to the typical condition. The additive model predicts just the opposite. Since the atypical-cue condition had a significantly higher summation of support of items generated, the additive model would predict that the atypical condition should have a higher probability judgment than the typical condition, which is not what occurred. Taken together the generation and judgment data appear to support an averaging rather than additive function underlying judgments.

However, there were no significant correlations between judgment and the number of items, summation, or average support of items generated (see Table 4). These correlations were based on 21-26 participants for each correlation and thus may have low power for detecting potentially small but significant correlations between judgments and any of the generation data.

	Atypical Cue Inducement			
	Judgment	Number of Items	Summation of Support	Average Support
Judgment	1	-.38	-.39	-.11
Number of Items	-.38	1	.97**	.01
Summation of Support	-.39	.97**	1	.16
Average Support	-.11	.01	.16	1
	Typical Cue Inducement			
Judgment	1	-.23	-.10	.28
Number of Items	-.23	1	.93**	-.25
Summation of Support	-.10	.93**	1	.11
Average Support	.28	-.25	.11	1

*p < .05. **p < .01

Table 4. Experiment 2 Correlations by Typicality: Judgment, Number of items Generated, Summation of Support of Items Generated, and Average Support of Items Generated

Consistent with Experiment 1 there was a significant positive correlation between the number of items generated and the summation of support of items generated for both typicality conditions (see Table 4). However, there was not a significant correlation between the number of items generated and the average support of items generated for either typicality condition as there was in Experiment 1.

Retrieval of different cue typicality had an effect on subjective probability judgments such that when atypical instances are retrieved the probability judgments are significantly less than when typical instances are retrieved. In order to test whether the effect of cue typicality on judgment was due to the generation of hypotheses, a series of covariate analyses were conducted: Three separate analyses were performed to test whether the effect of cue typicality remained significant after controlling for variance due to the number of items generated, the summation of the support values of the generated items, and the mean support of the generated items.

Only the number of items generated was a significant covariate of judgments $F(1, 43) = 4.30$, $p<0.05$, and when the number of items generated was controlled for as a covariate the effect of cue typicality was no longer significant. As suggested by the correlations presented in Table 4, the number of items generated was negatively related to judgment. This finding is clearly inconsistent with the additive model, which predicts that increases in the number of alternatives generated should lead to increases in judged probability. However, the results are consistent with the averaging model, which predicts that mean support should decrease as a function of the number of items generated, which should in turn lead to decreases in judged probability.

Chapter 4: Experiment 3

Introduction

The judgment results from Experiment 1 were replicated in Experiment 2: Cue typicality affected judgment magnitude with atypical cues leading to lower judgment than typical cues. Experiment 2 also demonstrated that cue typicality affected the number of items generated, the summation of support of items generated and the average support of items generated. However, one limitation of the previous Experiments is that the support functions are post-hoc subjective ratings and the terms 'typical' and 'atypical' have been used extensively, but have not been well defined or operationalized. For this reason, in Experiment 3 we operationalized 'typicality' in terms of experienced frequency. Experience with items may lead to a classification of typicality such that the more often an item is viewed the more typical it may appear to be of its representative category. It is plausible, therefore, that typicality may reflect frequency of occurrence and it is this conceptualization of typicality that will be examined.

Frequency learning tasks are often used in memory paradigms to examine judgment accuracy (Dougherty & Hunter, 2003; Sprenger & Dougherty, 2006; Windschitl et al., 2002). For example, Dougherty & Hunter (2003) had participants learn what items were ordered by particular individuals at a diner by presenting individual food items on the screen each time the individuals ordered the item. The participants learned the complete items in the set (i.e., the actual items ordered) and the frequency with which the items appeared (i.e., how often the items were ordered). This frequency learning is useful in that the set of items is experimentally determined and is thus a well defined and exhaustive set, and the frequency of each item is also experimentally controlled.

Since typicality and frequency are theorized to be closely linked, one may reasonably determine to operationalize, and test the definition of typicality as high presentation frequency (Sloman et al., 2004). With this definition, items with the highest presentation frequency would be considered the most typical items while items with the lowest presentation frequency would be considered the least typical items. Typicality can now be tested as frequency presentation and may be well defined and manipulated experimentally to more directly assess the potential affects of typicality on category judgments and generation.

Experiment 3 further assessed a potential limitation of the previous typicality findings: Does typicality affect judgments and generation in a small, well defined set? The previous experiments relied upon very large and fairly ill-defined categories in order to assess the effects of typicality on subjective probability judgments. It may be that a large and fuzzy category is needed in order for typicality of items presented to affect judgments. Once the judgment set is small and well defined the effects of typicality may be minimized or eliminated since the judge can more easily retrieve and consider all the items in the judgment set.

Additionally the current research more directly assess the differences between the modified part-set cuing and modified retrieval induced forgetting to see if there might be a greater affect on judgments or generation from one memory paradigm.

Participants

One hundred sixty two University of Maryland undergraduate students participated in Experiment 3. Participants received extra credit for a psychology course.

Materials

The experiment was completely computerized and was programmed in Revolution. There were two lists of stimuli: a list of ten fruits and a list of ten animals that was taken from the VanOverschelde, Rawson, and Dunlosky (2004) category norms from the fruit and animal category lists. The top two most common and the bottom two least common fruits and animals were excluded from the lists and within each list no item started with the same letter as any other item in that list (i.e., apple and apricot would not both appear on the fruit list since they both start with the letter “a”). This was to ensure that in the retrieval inducement phase the participant could be presented with the category title and a first letter of an item from the list and have that letter refer to a unique item in the list. For each participant the ten fruit and animal items in each list were randomly assigned to one of the possible presentation frequency (14-14-11-11-8-8-5-5-2-2).

Pictures for each category item from the two lists were obtained from Microsoft word clip art and from Google pictures such that each category item had a corresponding cartoon-type picture that was presented in conjunction with the item name.

The judgment question was modified from Sloman et al.’s (2004) packed judgment question; it was modified to reflect the current task and category to be judged. The generation question was revised from the judgment question to ask for generation rather than a judgment and was similar to the generation question from Experiment 1.

Design and Procedure

The study was a two (inducement type: part-set cuing versus retrieval inducement) by three (cue typicality: atypical, typical, or no-cue) between subjects design with both inducement and cue typicality being manipulated between participants. The entire experiment took approximately half an hour to complete.

Participants all received the same initial training with the two category lists. Participants were told that Steve the sloppy stamp collector loves to collect stamps, but Steve is very particular about the stamps that he collects; he only collects stamps that are exactly the same size, weight, texture, and are sold as a single stamp. However, whenever Steve collects a new stamp he simply throws it into a brown paper bag rather than organizing his stamp collection. The participants were told that Steve is starting a new collection of stamps and they are going to observe the stamps that he collects, and throws into his stamp bag, by observing each stamp picture and label that appear on the screen. Each presentation of a picture and label represents a new stamp that Steve has acquired, and that they will be asked questions of the stamps Steve collects at a later time.

Within each category list the items were randomly assigned to one of the possible frequency presentations (14-14-11-11-8-8-5-5-2-2). The animal and fruit category lists were then combined and randomized to create one master learning list. Participants were then shown each item sequentially in the master learning list with one item label and picture on screen for 4 seconds followed by .33 seconds where the screen is cleared and the next item label and picture is presented.

The part-set cuing and retrieval inducement conditions received the same initial training, the same judgment and generation tasks, but differed on type of inducement following training. After the learning phase the part-set cuing condition went immediately to the category judgment and generation phases while the retrieval inducement condition proceeded to the retrieval inducement phase which was then followed by the judgment and generation phases.

The category judgment phase followed the initial training in the part-set cuing condition participants. Each participant had one of the two categories (fruit or animals) that was randomly selected to be the target category such that the judgment category cues were taken from that category and the judgment question was about that category. For the category judgment phase participants were asked to provide a judgment to the following question:

> If Steve the sloppy stamp collector were to shake the sack of stamps up, and while blindfolded randomly select a stamp from the bag with ALL the stamps that you saw in it, and only the stamps that you saw, what is the probability that he would select a stamp with a [FRUIT or ANIMAL] on it?

For the no-cue condition the participants were asked the above category judgment question verbatim. For the atypical cue condition the probability question contained four cues in the above judgment question that were the category items that had the lowest presentation frequency (2-2-5-5) in the initial learning phase. The conclusion of the category judgment question in the atypical cue condition would ask “what is the probability that he would select a stamp with a(n) [FRUIT or ANIMAL] on it, such as a ___, ___, ___ or ____ stamp?" The blanks would be filled in with the appropriate low frequency category items for that participant. Likewise the typical cue condition judgment question was exactly the same format as the atypical cue condition except that the four cues that were used in the question were the category items that had the highest presentation frequency (11-11-15-15) in the initial learning phase. Since the items were randomly assigned to presentation frequency for each subject, these atypical and typical judgment cues differed by participant.

Immediately following the judgment phase participants in the part-set cuing condition proceeded to the generation phase where they were asked to unpack the focal hypothesis from the previous category judgment. This task was essentially a free recall of items learned in the initial learning phase from the category they just judged:

> Please try to recall all of the different types of [fruit or animal] stamps that you saw Steve put in his sack (i.e., what type of [fruit or animal] was on the stamp).
>
> Please type in a [fruit or animal] that you saw on a stamp that Steve put in his sack in the below box and when you are finished typing the [fruit or animal], press "Enter" AFTER EACH [FRUIT or ANIMAL] that you type in the box. When you can not recall any more [fruit or animal] stamps, or if you think you have recalled all the [fruit or animal] stamps, type "done" and then press "Enter" to continue.

There was no maximum number of items that the participants could generate and the items.

After the generation phase the participants immediately proceeded to a category item judgment phase where they were asked to judge the likelihood that the individual items they generated in the preceding category generation phase would be randomly selected from that category:

> Steve the sloppy stamp collector decided to sort through his special stamp sack and separate the fruit and animal stamps into 2 different brown sacks. If Steve shook up the sack with the [FRUIT or ANIMAL] stamps in it, and then while blindfolded he randomly picked a stamp from that fruit sack, what is the probability that Steve would pick a _____ stamp?

The above blank was filled in with an item from the category generation task that the participant had generated. Participants judged the probability of each item they generated in the generation phase. The items generated in the generation phase were randomized and then presented sequentially in the above judgment with one judgment immediately following another until all the generation items had been judged.

The category item judgment phase was followed by a final residual judgment phase where participants were asked to judge the likelihood that they forgot any category items in the generation phase, and what the probability would be of randomly selecting the non-recalled items from the category:

> Steve the sloppy stamp collector decided to sort through his special stamp sack and separate the fruit and animal stamps into 2 different brown sacks. If Steve shook up the sack with the [FRUIT or ANIMAL] stamps in it, and then while blindfolded he randomly picked a stamp from the [FRUIT or ANIMAL] sack, what is the probability that Steve would pick a [fruit or animal] stamp other than the [fruit or animal] stamps that you recalled that are listed below?

All of the items that the participant had recalled in the generation phase were listed below the residual judgment question.

After completing all of these phases the participants were debriefed. In summary, the part-set cuing condition had five phases: a learning, category judgment, generation phase, item generation judgment phase, and residual judgment phase.

In contrast to the part-set cuing condition, after the initial learning phase in the retrieval inducement condition an inducement phase was interposed between the judgment and generation phase. The inducement task used the same methodology from Experiment 2. In summary, the retrieval inducement condition had six phases: a learning, retrieval inducement, category judgment, generation, item generation judgment, and residual judgment phase.

In the inducement phase the category name of the category that was randomly selected to be judged (fruit or animal) was presented along with the first letter of the cue that was to be retrieved and participants are instructed to fill in the blank with the appropriate word from the stamp items previously learned. For example, in the typical condition a participant might see “Fruit – P________” and would be asked to fill in the blank with a fruit stamp item that they learned previously that started with the letter ‘P’. There were four cues that were used and each cue word stem was presented three times for a total of twelve retrieval inducement trials. In the atypical condition the four cues used were the cues that were randomly assigned to the presentation frequencies of two and four, which were the two lowest presentation frequencies. In the typical condition the four cues used were the cues that were randomly assigned to the presentation frequencies of eleven and fourteen, which were the two highest frequency presentations. The inducement cues were randomly presented for this retrieval inducement phase.

The retrieval inducement phase in the retrieval inducement condition was then followed by the category judgment phase whereby the participants were asked a packed category probability question for both the typical and atypical cue conditions that was the same packed question as the part-set cuing condition’s no-cue condition judgment question.

Immediately following the category judgment phase participants proceeded to a generation phase and were asked to unpack the focal hypothesis from the preceding judgment phase. This was followed by a category item judgment phase where the items the participants generated were judged. Finally there was a residual judgment phase followed by a debriefing. The generation, category item judgment and residual judgment questions and procedures were the same as those used in the part-set cuing condition.

Results and Discussion

Generation Data

There was no main effect of cue frequency or inducement type on the number of items generated (see Figure 9): The number of category items generated was equivalent in all typicality and inducement conditions (see Table 5). Since generation of items from a small set appears to be unaffected by cue frequency and inducement type, this may be consistent with Experiment 1 and Experiment 2 and argues against a part-set cuing or inhibition account.

Support Values

Support values were based on the presentation frequency of each item for each participant and were isomorphically translated into an estimate of support ranging from 2 – 15. Individual values of summation of support and a mean support of items generated for each participant were then calculated and averaged across conditions.

There was no main effect of cue frequency or inducement type on summation of support (see Figure 10): All frequency and inducement conditions generated items that had equivalent summation of support (see Table 5). There was also no interaction between cue frequency and inducement type.

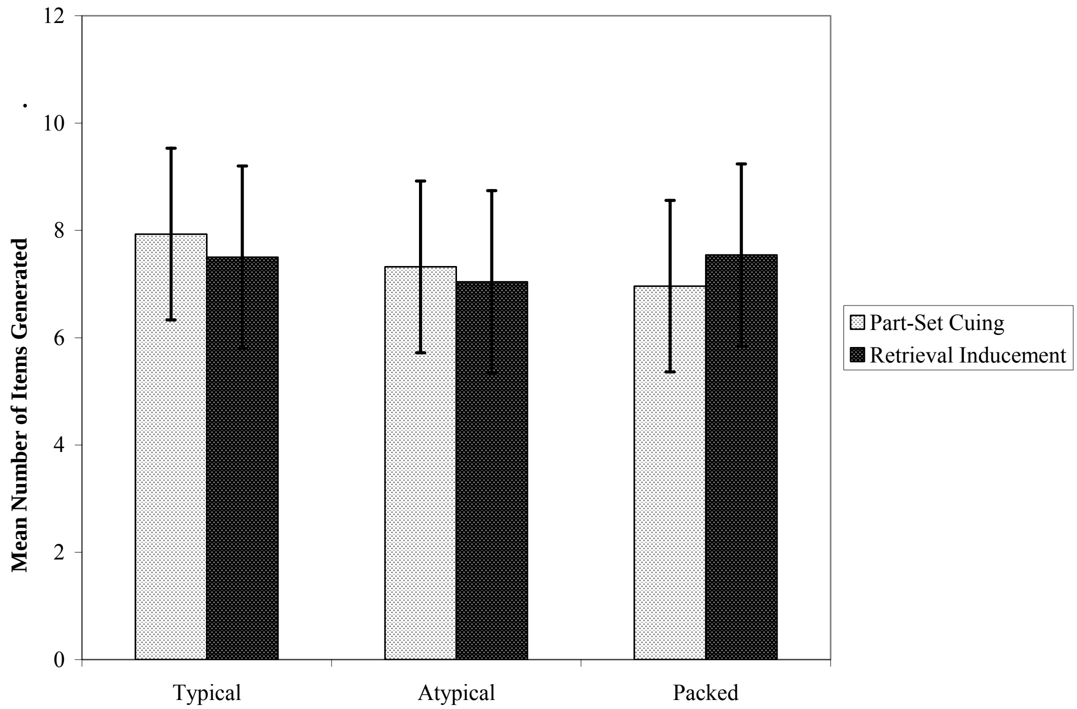

Figure 9. Experiment 3 Mean Number of Items Generated by Cue Typicality and Inducement with Standard Error Bars

	Part-Set Cuing		
	Packed	*Atypical*	*Typical*
Judgment Mean (Std. Deviation)	0.52 (0.02)	0.51 (0.02)	0.55 (0.02)
Number of Items Generated (SD)	6.96 (1.43)	7.32 (1.66)	7.93 (1.72)
Summation of Support of Items (SD)	56.96 (14.54)	61.18 (13.57)	63.50 (20.65)
Mean Support of Items (SD)	8.08 (1.30)	8.27 (1.29)	8.07 (1.26)
	Retrieval Inducement		
	Packed	*Atypical*	*Typical*
Judgment Mean (Std. Deviation)	0.49 (0.02)	0.52 (0.02)	0.54 (0.02)
Number of Items Generated (SD)	7.54 (1.66)	7.04 (1.67)	7.50 (1.73)
Summation of Support of Items (SD)	57.58 (17.98)	56.44 (20.43)	62.69 (12.32)
Mean Support of Items (SD)	7.50 (1.81)	7.23 (1.76)	8.51 (1.21)

Table 5. Experiment 3 Means (and Standard Deviations)

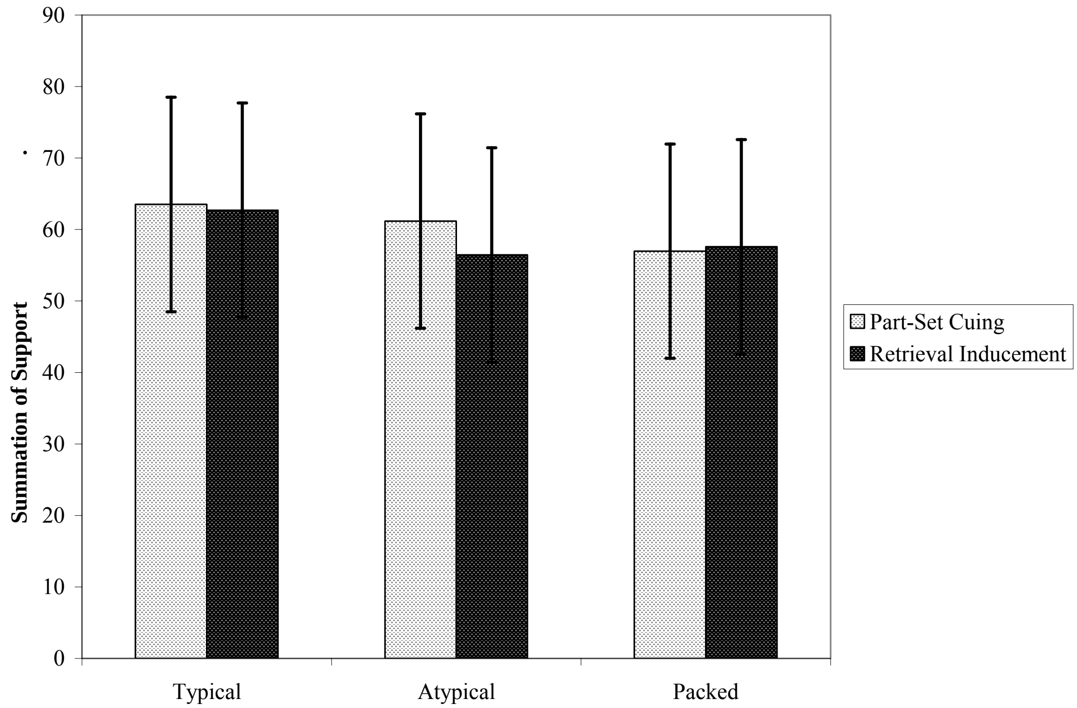

Figure 10. Experiment 3 Summation of Support of Items Generated by Cue Typicality and Inducement with Standard Error Bars

However, there was an interaction effect of cue frequency and inducement type on the mean support of items generated $F(2, 156) = 3.74$, $p < 0.05$ (see Figure 11). There was a main effect of cue frequency on mean support of items generated for the retrieval inducement $F(2, 74) = 4.51$, $p < 0.05$, but not for the part set cuing inducement. Further analyses indicate that the high frequency-cue condition had a significantly greater mean support of items generated than the low frequency-cue condition in the retrieval inducement conditions, but neither of the frequency conditions differed significantly from the packed conditions mean support of items generated (see Table 5). Taken together these results indicate that when a small set is being judged the support of items generated is largely unaffected by part-set cuing inducement and cue frequency, while retrieval inducement appears to affect the mean support of items generated differentially depending on retrieved cue frequency.

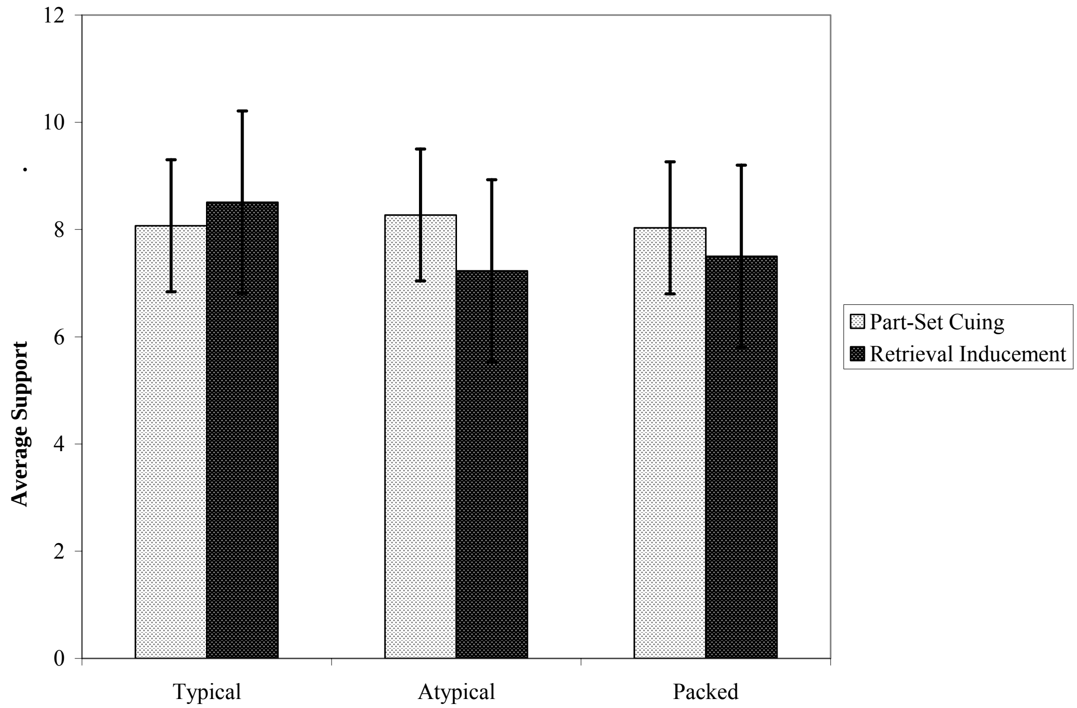

Figure 11. Experiment 3 Average Support of Items Generated by Cue Typicality and Inducement with Standard Error Bars

Judgments

There was no main effect of cue frequency or inducement type on judgment, nor was there an interaction between cue frequency and inducement type (see Figure 12): The subjective judgments were equivalent in all frequency and inducement conditions (see Table 5). This would indicate that judgments of small sets are not affected by unpacked or retrieved cue frequency and may operate by different processes than judgments of large sets.

Consistent with Experiment 1 and Experiment 2 there were no significant correlations between judgment and the number of items, summation, or average support of items generated for any of the frequency conditions collapsed across inducement type (see Table 6). There was also a significant positive correlation between the number of items generated and the summation of support of items generated for all frequency conditions (see Table 6). The correlations with the average support of items generated differ slightly among the frequency conditions. The low frequency-cue and packed conditions do not have a significant correlation between the number of items and the average support of items generated, but there is a significant positive correlation between the summation and the average support of items generated (see Table 6). The high frequency-cued condition has a near reversal of the other conditions correlation results: There is a significant negative correlation between the number of items and the average support of items generated and no significant correlation between the summation and the average support of items generated (see Table 6).

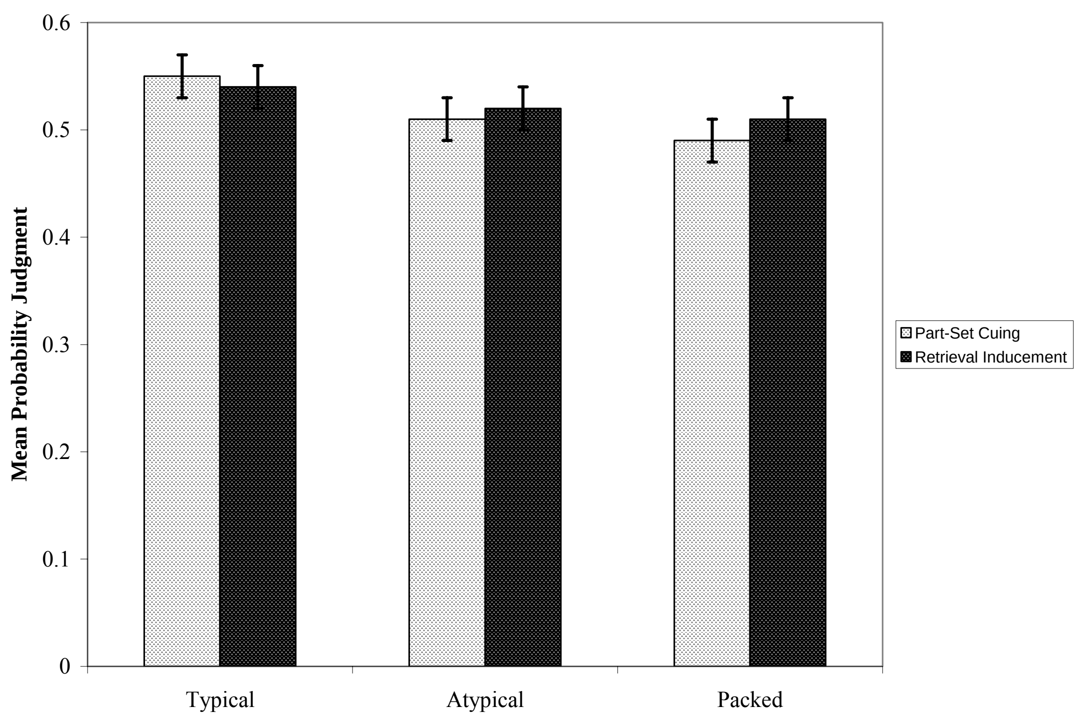

Figure 12. Experiment 3 Mean Probability Judgments by Cue Typicality and Inducement with Standard Error Bars

	No Category Cue Condition			
	Judgment	Number of Items	Summation of Support	Average Support
Judgment	1	-.12	.03	.10
Number of Items	-.12	1	.68**	.13
Summation of Support	.03	.68**	1	.65**
Average Support	.10	.13	.65**	1
	Atypical Category Cue Condition			
Judgment	1	-.01	.02	.20
Number of Items	-.01	1	.60**	-.05
Summation of Support	.02	.60**	1	.51**
Average Support	.20	-.05	.51**	1
	Typical Category Cue Condition			
Judgment	1	.125	-.05	-.13
Number of Items	.125	1	.65**	-.38**
Summation of Support	-.05	.65**	1	.14
Average Support	-.13	-.38**	.14	1

Table 6. Experiment 3 Means (and Standard Deviations)

The correlation data indicate that in all frequency conditions the more items that are generated the greater the summation of support, but when high frequency cues are provided, the more instances that are generated the less the average support of the generated items. This makes sense inasmuch as the items left to unpack after the high frequency cues are provided are the low frequency cues, so if judges continue to unpack after the high frequency cues are provided they are by definition unpacking low frequency items that drive their summation of support up and their average support down. Since the low frequency cues are provided, the judges start out with the highest average support possible (all of the highest support items and none of the low support items). Unpacking will then increase the summation of support, but it can only serve to decrease the average support for the judges.[5]

Conversely, when the low frequency cues are provided that leaves the high frequency cues to be unpacked, and if judges continue to unpack this would lead to an increase in the summation of support as well as an increase in the average support as higher frequency items with high support are included. Since the low frequency cues are provided, the judges start out with the lowest average support possible (all of the lowest support items and none of the high support items). Unpacking can only increase both the summation and the average support since only higher support items will be generated.

In order to test whether the effect of cue frequency on judgment was due to the generation of hypotheses, a series of covariate analyses were conducted to test whether the effect of cue frequency was significant after controlling for variance due to the number of items generated, the summation of the support values of the generated items, and the mean support of the generated items. None of these were significant as covariates and all effects of cue frequency and inducement remained not significant: The number, summation, and mean support of items generated are not predictive of judgments. Taken together it appears that none of the generation data are predictive of judgments.

[5] This is assuming that the judge unpacks relevant instances from the learned category set.

Chapter 5: General Discussion

The purpose of this paper was to assess three theories that account for the effect of typicality on unpacking: the central tendency of support theory, the narrow interpretation theory, and a misinterpretation theory. The paper also contrasted the additive and averaging support models for subjective probability judgments.

The generation data from all three Experiments indicate that there are no part-set cuing effects: Experiments 1 and 3 show that the cued conditions do not generate significantly less items than the non-cued conditions, while Experiment 2 shows that the atypical-cue condition generated more items than the typical-cue condition. Part-set cuing effects, which were hypothesized as a potential mechanism for the narrow interpretation theory, would not predict these generation results: If part-set cuing effects were present it would be expected that the non-cued conditions would generate more items than the cued conditions.

Experiments 1 and 2 replicated the judgment results found by Sloman et al. (2004): The typicality of category item cues affected subsequent subjective judgments. Participants who were cued with atypical category instances provided significantly lower mean category judgments than those participants who received no cues or typical category cues. The no-cue and typically cued conditions provided mean judgments that were not significantly different from each other. In contrast, Experiment 3 did not show any effects of cue typicality on judgments of small sets and indicates that the typicality effect may be specific to large judgment sets.

Though the judgments differed significantly by cue typicality, the generation data from Experiment 1 showed no effect of cue typicality. Additionally, none of the generation data were found to be significant predictors or mediators of judgment, and the main effect of cue typicality on judgment remained significant when controlling for number, summation, or mean support of items generated. Thus, the generation data are not able to predict the judgment data and do not allow for discrimination between an additive and averaging model of judgment since neither model would predict the judgment results. The contrasting effect of typicality on judgments and lack of typicality effects on the generation data indicate that a misinterpretation of the question may be causing the judgment data.

Experiment 2 used a modified retrieval induced forgetting paradigm to reduce the likelihood of bias and misinterpretation by enabling all participants to be asked the same packed category judgment question. Since the judgment results from Experiment 2 also corroborate the judgment results of Sloman et al. (2004) and Experiment 1, this indicates that neither a misinterpretation of the question nor an anchoring and adjustment process underlies the differences in judgments.

Furthermore, in the retrieval-induced forgetting paradigm in Experiment 2, both the judgment and generation data show an effect of cue typicality. The atypical-cue condition had a significantly higher number of items generated and subsequently a higher summation of support, but a lower average support of items than the typical-cue condition. This diverging result offers differential predictions for the additive and averaging models with the averaging model accurately predicting the judgment results and the additive model failing to accurately predict.

Moreover, the number of items generated is a significant *negative* predictor of judgments, and when controlled for the effects of cue typicality become non-significant. This indicates that the number of items generated is mediating judgments such that judgments decrease as more items are generated. This result is also inconsistent with an additive model, which predicts a positive relationship between number-of-items generated and judgment. In contrast, the result is not inconsistent with an averaging model.

However, the judgment results from Experiment 3 indicate that frequency may not be equivalent to typicality. Frequency of presentation, or experience with a set of items, appears to be a different construct than typicality inasmuch as we have diverging effects on judgment for frequency and typicality. It may be that typicality is more of a prototype comparison than a frequency learning construct, and is something that future research may want to explore in order to clearly assess what the construct of typicality represents.

Experiment 3 also shows that frequency of presentation does not affect judgments and indicates that the typicality effect may be specific to large judgment sets. There were no significant effects of cue frequency or inducement type on category judgments, number of items generated, or the summation of support of the items that were generated, nor were any of these measures significant predictors of judgment.

A potential reason that there are no frequency effects on judgment may be due in part to these judgment results being biased towards the null hypothesis since the objective probability was 50%. The 50% response appears to the probability most often chosen when people do not know the objective probability (Fischhoff & Bruine De Bruin, 1999). Additionally, in frequency learning it may be more difficult to learn a ratio of 70-30 than a ratio of 50-50, and this difficulty in learning may provide more variability between participants that could increase differences in judgments. Future studies may want to manipulate the objective category frequencies to assess if such a bias is a concern in interpreting the data.

Interestingly, the results of Experiments 1 and 2, both of which revealed main effects of typicality on judgment, might be due to different mechanisms. For example, the most plausible account of the results of Experiment 1 is the misinterpretation hypothesis. That is, because participants were given slightly different versions of the judgment question, this may have enticed them to interpret the judgment question differently (even though they were normatively equivalent). As Stanovitch and West (2000) point out, participants sometimes interpret judgment questions differently from what is intended by the experimenter:

> It is the argument that although the experimenter may well be applying the correct normative model to the problem as set, the subject might be construing the problem differently and be providing the normatively appropriate answer to a different problem – in short, that subjects have a different interpretation of the task.

Our replication of Sloman et al (2004) in Experiment 1 suggests that this might just be the case here where it appears reasonable to assume that participants are misinterpreting the question due to the different exemplars provided.

However, the results of Experiment 2 are open to a more interesting psychological account, namely that retrieval-induced forgetting appears to have differential effects on retrieval depending on cue typicality. Cue typicality inducement may either be preventing participants from, or facilitating generation of alternative hypotheses for including in the judgment process. Participants induced with typical exemplars fail to retrieve (atypical) items that participants induced with atypical exemplars retrieve. Conversely, participants induced with atypical exemplars show a benefit in retrieval inasmuch as they generate more items than the typical-cue condition.

Moreover, the generation data in conjunction with the judgment data from Experiment 2 are consistent with an averaging model of support theory, rather than an additive model. The differential effects of cue typicality on retrieval provided opposing predictions for the judgment data based on the additive and averaging models, with the averaging model correctly predicting the judgment results.

Finally, the results of experiment 3 suggest that the effect of typicality on judgment may be limited to large sets of items or natural categories. In fact, this finding is in keeping with recent research showing that part-set cueing is not effective when implemented within small sets (Oswald, 1999).

Alternately, it may be the case that the generation data do not correspond to the items considered when making the preceding judgments. Future studies should manipulate the timing of the generation task, such that participants generate items either prior to, or after, making their judgment. In this way, one can more closely examine the impact of generation on judgment.

Bibliography

Anderson, M.C. (2006). Repression: A cognitive neuroscience approach. In M.Mancia (Eds.), *Neuroscience and Psychoanalysis* (pp327-350). Milan: Springer.

Anderson, M.C. (2004). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. In C. MacLeod & B. Uttl (Eds.), *Dynamic Cognitive Processes* (pp.159-190). Tokyo: Springer-Verlag.

Anderson, M.C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49,* 415-445.

Anderson, M.C. (2001) Active forgetting: Evidence for functional inhibition as a source of memory failure. *Journal of Aggression, Maltreatment and Trauma, 4*, 185-210.

Anderson, M.C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130*, 544-570.

Anderson, M.C., Bjork, E.L., & Bjork, R.A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin and Review, 7*, 522-530.

Anderson, M.C., Bjork, E.L., & Bjork, R.A. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063-1087.

Anderson, M.C., Green, C., & McCulloch, K.C. (2000). Similarity and inhibition in long term memory: Evidence for a two-factor model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1141-1159.

Anderson, M.C., & McCulloch, K.C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 608-629.

Anderson, M.C., & Spellman, B.A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102,* 68-100.

Anderson, N.H. (1959). Test of a model for opinion change. *Journal Abnormal Social Psychology, 59,* 371-381.

Anderson, N.H. (1991). A cognitive theory of judgment and decision. In N. H. Anderson (Eds.), *Contributions to information integration theory: Vol. I. Cognitive* (pp. 105-142). Hillsdale, NJ: Erlbaum.

Basden, D.R., & Basden, B.H. (1995). Some tests of the strategy disruption hypothesis of part-list cuing inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1656-1669.

Baüml, K. (2002). Semantic generation can cause episodic forgetting. *Psychological Science, 13,* 356-360.

Blaxton, T.A., & Neely, J.H. (1983). Inhibition from semantically related primes: Evidence of a category-specific inhibition. *Memory and Cognition, 11,* 500-510.

Brody, R.G., Coulter, J.M., Daneshfar, A. (2003). Auditor probability judgments: Discounting unspecified possibilities. *Theory and Decision, 54,* 85-104.

Brown, A.A. (1981). Inhibition in cued recall. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 204-215.

Brown, J. (1968). Reciprocal facilitation and impairment of free recall. *Psychonomic Science, 10,* 41-42.

Dougherty, M.R. & Hunter, J.E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica, 113,* 263-282.

Fischhoff, B. & Bruine De Bruin, W. (1999). 50% = Fifty-fifty. *Journal of Behavioral Decision Making, 12,* 149 – 163.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Human Perception and Performance, 4(2),* 330-344.

Foos, P.W., & Clark, M.C. (2000). Old age, inhibition, and the part-set cuing effect. *Educational Gerontology, 26,* 155 - 160.

Johnson, S.K., & Anderson, M.C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science, 15,* 448-453.

Levy, B.J., & Anderson, M.C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences, 6,* 299-305.

Mann, D. (1997). Tests of the unpacking principle in support theory. *Dissertation Abstracts International, 58,* 9B.

Oswald, K.M. (1999). The role of implicit interference on recall: Evidence from part-set cuing inhibition. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 60* (3-B) 1325. (UMI No. 9922144)

Rottenstriech, Y. & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 104,* 406-415.

Rundus, D. (1973). Negative effects of using lit items as recall cues. *Journal of Verbal Learning and Verbal Behavior, 12,* 43-50.

Schlottman, A., & Anderson, N.H. (1995). Belief revision in children: Serial judgment in social cognition and decision-making domains. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1349-1364.

Shanteau, J.C. (1970). Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology, 93,* 63-68.

Shanteau, J.C. (1972). An additive model for sequential decision making. *Journal of Experimental Psychology, 85,* 181-191.

Slamecka, N.J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology, 76,* 504-513.

Sloman, S., Rottenstreich, Y., & Wisniewski, E. (2004). Typical versus atypical unpacking and superadditive probability judgment. *Learning, Memory, and Cognition, 30,* 573-582.

Sprenger, A. & Dougherty, M.R. (2006). Differences between probability and frequency judgments: The role of individual differences in working memory capacity. *Organizational Behavior and Human Decision Processes, 99*, 202-211.

Stanovich, K. E., & West, R. F. (2003). The rationality debate as a progressive research program. Behavioral and Brain Sciences, 26, 531-534.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences, 23, 645-665.

Todres, A.K., & Watkins, M.J. (1981). A part-set cuing effect in recognition memory. *Journal of Experimental Psychology: Human Learning & Memory, 2,* 91-99.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 292-315.

Tversky, A. & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547-567.

Van Overschelde, J.P., Rawson, K.A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289-335.

Watkins, M.J. (1975). Inhibition in recall with extralist "cues." *Journal of Verbal Learning and Verbal Behavior, 14,* 294-303.