RICE UNIVERSITY

IMAGERY AND MEMORY: THE BIZARRENESS ISSUE REEXAMINED

by

PAMELA ANN KENNEDY

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

MASTER OF ARTS

APPROVED, THESIS COMMITTEE:

Sarah A. Burnett
Assistant Professor of Psychology
Chairperson

William C. Howell
Professor of Psychology

HOUSTON, TEXAS

MAY, 1979
Abstract

IMAGERY AND MEMORY: THE BIZARRENESS ISSUE REEXAMINED

by

PAMELA ANN KENNEDY

This research examined the effects of the bizarreness attribute of imagery on memory. While previous research has not generally supported the facilitory effects of bizarreness on recall performance, there are a number of conceptual and methodological problems with this body of research. The present research attempted to overcome these problems by developing a more rigorous conceptualization and operationalization of the construct of bizarreness, and utilizing a design which controls for past methodological contaminants.

Half of the 64 subjects were instructed to form images while the other half rehearsed phrases in rote fashion. Within each of these conditions, half of the subjects were tested by free recall and half by frequency estimation. For all subjects, half of the phrases were bizarre and half were common, as defined by pre-ratings made by independent subjects. Frequency level of phrase presentation was varied, with frequency levels 1, 2, 4 and 6 being represented. Finally, subjects were tested immediately after list presentation and again after one week.

The results indicated that for cued recall, bizarre phrases were superior to common phrases. In addition, the superiority of bizarre
over common imagery increased between the immediate and delayed tests. For frequency estimation, the data did not reveal any readily interpretable differences between common and bizarre phrases.

The results were discussed as they relate to previous research on bizarreness. It was concluded that bizarreness does have a facilitory effect in imagery mediation. Suggestions for future research to further clarify the effects of bizarreness were presented.
Acknowledgments

I would like to express appreciation to the members of my thesis committee: Sarah A. Burnett, who served as Chairperson, and William C. Howell. I am indebted to both of these for their guidance and assistance on all stages of my work.

I would also like to express sincere gratitude to Robert D. Pritchard, who was a constant source of moral support and encouragement during my hours of writing.
## Table of Contents

Title page

Abstract ........................................... i

Acknowledgments ................................. iii

Table of Contents ................................. iv

Introduction ................................... 1

Method .......................................... 25

Results ......................................... 30

Discussion .................................... 37

References .................................... 42

Appendix A ..................................... 45
Imagery has been recognized as a useful memory aid for centuries. Indeed the practical mnemonic application of imagery has existed in professional mnemotechniques since its formal inception in ancient Greece. According to several treatises (see Yates, 1966), a Greek poet named Simonodes was the first to develop and formally describe a method for using imagery to improve memory. This method, commonly known as the "method of loci," combined the use of images with rules for their orderly mental arrangement, which enabled the user to retain long speeches, lists, etc. (Yates, 1966). There are several steps involved in using this method. First, one would visualize a familiar, highly imageable set of locations (loci) and memorize them in order. After the set of loci are committed to memory, the items or passages to be learned are represented as images and stored sequentially in these locations. When recall of this material is desired one would progress mentally from one location to the next. By visualizing each location and its contents, the information stored there would be available for recall.

Many of the systems used by professional mnemonists employ the same principles outlined in the method of loci. For example, the peg word system is a commonly used mnemonic technique (see Pavio, 1971) which is based on imagery and rules to preserve order. In this system peg words which rhyme with the numbers one through ten (1 - bun, 2 - shoe, 3 - tree, etc.) are used in place of visual locations.
Specifically, a to-be-remembered word is associated with each peg word by forming an image between the two. Order is preserved by the numbers associated with each peg word. When retrieval is desired the peg word is used to elicit the image and the to-be-remembered word should be available for recall.

Although many such ideas and schemes for improving memory have been suggested by stage mnemonists down through the years, only in the last twenty years or so have these techniques been tested in a systematic fashion by psychologists. According to psychologists during the behavioristic era, imagery could not be observed and was thus not a suitable subject for study. Thus the functional role of imagery in memory, the value of imagery mediation relative to other memory systems, and specific attributes of the image responsible for its effects had not been adequately understood. Recently however, by incorporating many of the practices and assumptions derived from mnemonic strategies into testable hypotheses, considerable research effort has been directed toward these issues. As a result, fairly consistent findings and generalizations have emerged supporting imagery mediation as a memory system and specifying qualities of the image which enhance its effectiveness as a mediational device.

If findings from imagery studies are examined, it is clear that the use of imagery has been shown to result in impressive memory performance in many situations. Although many early studies were conducted in the absence of control or comparison groups, the results have shown that by simply instructing subjects to form images using techniques such as the method of loci and peg word system, surprisingly excellent recall is produced (Bugelski, 1968; Pavio, 1970).
Furthermore, the use of imagery in traditional learning paradigms (with control groups) has yielded additional evidence supporting its value in the memory system. Take for example paired-associated learning, which is the paradigm most frequently employed to assess imagery effects. Several different instructional sets have been used in P-A paradigms. In these instructional sets subjects have been 1) instructed to form images connecting the stimulus and response terms of a paired-associate list, 2) told to use variations of the peg word system, 3) given pictures, sentences, or phrases and asked to generate the specified images. In all of these variations of the P-A task, memory facilitation has been demonstrated consistently (Pavio, 1971; Reese, 1977).

A third line of support for imagery comes from studies which have tested the value of imagery mediation relative to other memory systems (e.g. Delin, 1969). In general, instructing subjects to use imagery produces better memory performance than instructing them to rehearse in a rote fashion or giving them no instructions at all (Pavio, 1971; Bower, 1972; Reese, 1977; Kieras, 1978).

From this line of research a number of factors critical to the effect of imagery on memory performance have begun to be studied and identified. These factors include individual difference variables, the nature of the material to be learned, how it is coded, etc. From these studies, generalizations have emerged regarding the specific qualities of the image which increase its effectiveness as a mediational device. Two such well documented generalizations have to do with concreteness of the image and whether or not the image represents interaction between the members of the noun pair. The conclusion that
images formed from concrete nouns are easier to form and recall than those from abstract nouns is now widely accepted (Pavio, 1971; Bower, 1972). Furthermore, in a paired-associate task recall is facilitated if the image connecting the stimulus and response terms is one that depicts the referents of the two terms in interaction rather than in isolation (Wollen, Weber, and Lowry, 1972; Bower, 1972).

Another quality of the image which has received considerable research attention but has not produced clear results is bizarreness. The idea that bizarre images are remembered better than common images is found in recent mnemonic applications (Lorayne and Lucas, 1974; Bower, 1973), as well as in ancient treatises (Yates, 1966). In modern day theory the assumption that bizarreness produces superior retention seems to be based on the belief that making an image bizarre increases its distinctiveness and thereby reduces interference (see Pavio, 1971). Research on this particular attribute of images has been plagued with difficulties and consequently results have been inconclusive.

The purpose of this paper is to explore the contributions of bizarreness in the influence of imagery on cued recall. That is, while it is clear that imagery aids recall, it is less clear what role bizarreness plays in the facilitory effects of imagery. In the section that follows, the conclusions of the existing research dealing with this issue will first be summarized briefly. Next, the major problems with this body of research will be discussed. These problems involve: 1) the conceptualization of bizarreness, 2) the operational definition of bizarreness, and 3) a variety of methodological/design issues. In a second section, the ways the present study attempts to overcome the problems described in these three areas will be presented.
Evaluation and Review of Existing Research

The conclusions of the existing research on the relationship between bizarreness of the image and recall have been mixed. The majority of studies have concluded that bizarreness is not critical (Briggs, Hawkins, and Crovitz, 1970; Johnson, 1972; Nappe and Wollen, 1973; Hauck, Walsh, and Kroll, 1976; Wood, 1967; Collyer, Jonides, and Bevan, 1972; Wollen, Weber, and Lowry, 1972; Senter and Hoffman, 1976). Other studies have shown support for bizarreness (Crovitz, 1969; Delin, 1968; Perensky and Senter, 1970; Andreoff and Yarmey, 1976). Careful examination of these studies, however, suggests a variety of problems that make it impossible to draw any firm conclusions about the effects of bizarreness. Thus, the following discussion will consider the nature of these problems, and relevant aspects of each study will be reviewed as they relate to the specific problems being discussed.

The first problem area deals with the conceptualization of bizarreness. As Bower (1970) has indicated, bizarreness is a poorly defined construct. This is clearly evident in the literature. For example, it has been defined as unnatural (Nappe and Wollen, 1973), strikingly unusual (Andreoff and Yarmey, 1976; Wollen, Weber, and Lowry, 1972), and implausible, incongruous, grotesque, ludicrous, odd, strange (Collyer, Jonides, and Bevan, 1972). The real issue here is that authors have not used a common or even clearly defined conceptualization of the construct of bizarreness. In some studies a definition is presented, but as exemplified in the four studies cited above, the definitions vary. In other studies, no actual
definition is given, and the conceptualization of the construct must be deduced from the methods of operationalization.

This lack of a common conceptual base for bizarreness presents two basic problems. First, it affects the design of the studies. Different conceptualizations lead to different operationalizations and different methodologies. Without a clear conceptualization, operational definition issues can, and have been approached by different researchers as if they were methodological issues rather than issues associated with the conceptualization. For example, as will be shown later, the manner of generating the stimulus material used in a study depends heavily on the way the construct of bizarreness is conceptualized. Thus, stimulus material generation is not simply a methodological issue, but should be directly related to the conceptualization.

The second problem with the lack of a clear conceptual base is related to comparability across studies. With different studies using varying or even undefined conceptualizations of bizarreness it is impossible to build a body of results from which firm conclusions can be derived. Not only is it impossible to compare the results of studies using different conceptualizations/operationalizations with each other, but it also makes generalization extremely difficult.

The importance of having a clear conceptual base is apparent from a study by Collyer, Jonides, and Bevan (1972). In this study one group of subjects was given common noun-verb-noun triplets and a second group was given bizarre triplets. The authors defined bizarre triplets as those involving implausible relationships, while common triplets involved plausible relationships. However, as the authors
point out, some of the common triplets were, in fact, quite unusual. For example, one triplet in the common set was elephant-trample-garden. While the image formed by this triplet is plausible, it is hardly common. Thus, the authors may well have been comparing bizarre images as defined by being implausible with images that were some combination of plausible—common and plausible—uncommon. This sort of manipulation not only limits the definition of bizarre to being implausible, but also confounds commonness with unusualness. A clearly defined conceptualization of bizarreness and commonness would avoid this problem.

The second major area where problems exist in the literature deals with the operational definition of bizarreness. This does not refer to the general issue of the lack of clear conceptual base, but rather, to two specific operationalization issues. The first is whether to use subject-generated versus experimenter-generated images, and the second is whether to evaluate the degree of bizarreness of the resulting images subjectively or objectively.

The question of experimenter versus subject generated images is essentially one of who generates the elaborated relationship between the noun pair. One way for this to be done is for the experimenter to give subjects a noun pair imbedded in a phrase or sentence which suggests the image. The subject is then asked to form this image. This method shall be referred to as imposed imagery since the experimenter is suggesting the image for the subjects to use. The other way the image can be formed is by giving subjects a noun pair and asking him/her to generate an image that relates the two nouns. This type of manipulation shall be referred to as induced imagery since the
elaborated relationship between the two nouns is actually generated by the subject, not the experimenter.

In comparing these two methods, ancient mnemonic strategies (Yates, 1966) as well as current mnemotechnic systems (Lorayne and Lucas, 1974) have proscribed the induced approach because idiosyncratic image generation is believed to make the images stronger and the encoding more unique. On the other hand, several studies have shown no better recall for induced than for imposed imagery (e.g., Johnson, 1972; Briggs, Hawkins, and Crovitz, 1970; Kemler and Jusczyk, 1975).

However, several other points are pertinent to deciding which of these two procedures to use. First, imposed imagery allows for a much tighter definition of the manipulation. That is, the images can be formed consistent with the conceptualization. In turn, the experimenter can have more confidence that the images are truly common or bizarre as defined by the conceptualization. In contrast, induced imagery allows the subject to develop the image, and this image may or may not be consistent with the conceptualization. Second, imposed imagery allows for more rigorous control over the degree of bizarreness across subjects. That is, when all subjects receive the same elaborated relationship between the two nouns a known and constant degree of bizarreness is present. When subjects generate their own images, the degree of bizarreness will vary depending on the idiosyncratic relationships the subjects develop. Thus, from a methodological standpoint it seems clear that the imposed method has clear advantages over the induced method.

The second operational definition issue deals with whether the degree of bizarreness in the resulting images is evaluated subjectively
or objectively. It is obviously necessary to determine the degree of bizarreness of the images used. One way this has been done is for the experimenter to make an intuitive judgment as to the degree of bizarreness. The other approach is to use empirical rating techniques. The first approach will be referred to as subjective evaluation of bizarreness, and the second as objective evaluation of bizarreness.

Comparison of the subjective with the objective technique is fairly straightforward. When the bizarreness of the images is evaluated subjectively it may not be clear exactly what conceptualization is being used. When the conceptualization is defined and independent raters agree on the degree of bizarreness, the researcher has greater confidence in the manipulation.

The discussion so far has introduced the induced-imposed and subjective-objective issues in general terms, without reference to the specific studies involved. Although a few early studies did not attempt to evaluate the degree of bizarreness of the images used (Wood, 1967; Crovitz, 1969; Perensky and Senter, 1970) most studies have used one of the four combinations of induced-subjective, induced-objective, imposed-subjective, or imposed-objective strategies. Since each of these four strategies has different advantages and disadvantages, it is more informative to consider the specific studies in terms of the particular combination used.

The first combination to be addressed is induced imagery and subjective evaluation. In the one study using this procedure (Andreoff and Yarmey, 1976), the experimenter subjectively evaluated the common and bizarre image arousing capability of the noun pairs before they were presented to the subjects. The subjects were then given
instructions to form either bizarre or common images relating each noun pair during the experiment.

It should be evident that the problem with this method is the absence of both control over the type of image formed by the subjects, as well as any reliable method for determining the effectiveness of the bizarreness instructional manipulation. In other words, there is no a priori way to know whether or not the images formed will correspond to the desired type (common or bizarre), and no objective assessment of the extent to which the images did in fact conform to these two types.

The second combination, which has been used frequently in the literature, is imposed imagery with subjective evaluation (Johnson, 1972; Senter and Hoffman, 1976; Briggs, Hawkins, and Crovitz, 1970; Wollen, Weber, and Lowry, 1972). In this combination the experimenter forms the phrases/sentences according to his/her definition of common and bizarre images. These phrases/sentences are then given to the subjects with instructions to form images conforming to the relationship suggested by the phrases/sentences. The experimenter would evaluate the common or bizarre arousing capability of the stimulus phrases/sentences in much the same way as with the noun pairs in the induced-subjective combination discussed above.

By definition, it is clear that pure subjective evaluation used in conjunction with either induced or imposed imagery yields no empirical assessment of the degree to which bizarreness was effectively manipulated. Thus, the problems discussed above under subjective evaluation clearly apply to both induced-subjective and imposed-subjective combinations.
The third combination, induced imagery with objective evaluation has also been a common strategy in bizarreness studies (Nappe and Wollen, 1973; Hauck, Walsh, and Kroll, 1976; Delin, 1968). In this strategy subjects are instructed to form their own images relating each noun pair and to record these images during the experimental session. Multiple judges then formally rate the resulting images in terms of the degree of bizarreness represented. The major difference between this and the induced-subjective combination is that the resulting images are rated by a number of judges on some dimension of bizarreness. The major problem with this method of objective evaluation is the absence of any consistent conceptualization of the construct of bizarreness. In addition, the experimenter is left with only those images the subjects generate, regardless of how poorly they may approximate the definition of common and bizarre imagery groups.

The final combination includes imposed imagery and objective evaluation. In this approach, the phrase/sentence stimulus material is formed by the researcher and objectively evaluated by multiple judges for the degree of commonness/bizarreness represented. The only study which has even approximated this approach is one by Collyer, Jonides, and Bevan (1972) who constructed noun-verb-noun triplets and had ten judges pre-rate the bizarreness of the images evoked by these triplets.

It should be clear from the foregoing discussion of image induction and image evaluation methods that the best procedure for operationally defining bizarreness is to use imposed images and objective evaluation. To recapitulate, using imposed imagery as compared with induced imagery allows for a more precise specification of the conceptualization,
greater control over the nature of the manipulation, and a more consistent and reliable manipulation across subjects. The objective evaluation strategy is superior to the subjective in that the former allows for a clearer specification of what the conceptualization is; and, because of the use of multiple judges, it allows for greater confidence that the stimulus material actually is consistent with the conceptualization. Thus, given the superiority of the imposed-objective method, and the problems associated with the other three methods, the results of studies using those three methods must be interpreted with caution. In addition, this analysis suggests that future research should employ the imposed-objective methods of imagery manipulation and evaluation.

The third category of problems in existing research deals with problems in experimental design. The problems represented under design criticisms will be broken down into three general categories for the sake of clarity. These categories are: 1) factors which, when confounded with bizarreness, increase the probability of finding a bizarreness effect; 2) factors which affect image formation; and 3) factors which interfere with maximum retrieval/recall of correct response words.

Two factors which have been confounded with bizarreness are the use of imagery perse and the degree of interaction present in the image. The presence of these confounds can produce a bizarreness effect which is in fact attributable either to imagery itself or to interaction. Specifically, early studies (Perensky & Senter, 1970; Crovitz, 1969; Briggs, Hawkins and Crovitz, 1970; Johnson, 1972) did not include appropriate (nonbizarre imagery) control groups to ensure that the effects of bizarreness were separated from imagery. These studies
looked at recall as a function of bizarre imagery mediation, verbal mediation, standard P-A instructions, or no instructions at all, and found positive support for the use of bizarre imagery. However, since imagery has been shown to facilitate memory, it is impossible to say whether superior recall performance was due to the use of imagery per se or to the bizarre quality of the image. Therefore, no conclusions regarding the effects of bizarreness can be drawn from these studies.

In other studies (Wood, 1967; Johnson, 1972; Crovitz, 1969; Briggs, Hawkins and Crovitz, 1970; Perensky & Senter, 1970) bizarreness apparently was confounded with the degree of interaction of the images. This potential confounding occurs in studies using both induced and imposed imagery. In those studies where subjects were asked to form their own images (induced imagery) confounding was possible for several reasons.

First, subjects were not explicitly instructed to form either interacting or non-interacting images. In addition, in some cases the example given to illustrate a bizarre image depicted the two nouns in interaction. Finally, in terms of sheer probabilities of occurrence, it is more likely for a bizarre relationship between two nouns to include interaction than it is for a common relationship to include such interaction. For example, consider the noun pair "coffee-book." A common image for coffee-book might be a cup of coffee sitting beside an open book. A bizarre image for this same pair would likely be something like a book floating in a giant cup of coffee. In those studies where imagery was imposed by the use of phrases or sentences, confounding occurred as a result of not holding the degree of interaction represented by the phrase/sentence constant for all noun pairs.
Thus, the above studies failed to control the amount of interaction present in the different images. The problem, as stated previously, is that interactive images result in better recall performance than non-interactive images. Therefore, analogous to the problem of confounding imagery with bizarreness, the effects of using bizarre images cannot be separated from the effects of using images that contain interaction. The implication of these two problems is that although the studies which confounded bizarreness with imagery or interaction have supported bizarreness as a critical variable, such support must be considered potentially artifactual. Thus, they leave unanswered the question of whether bizarreness is important in imagery.

The second category of design criticisms to consider is factors which affect image formation. These are: 1) noun concreteness; 2) inter-item association strength of the noun pairs; and 3) plausibility of meaningfulness of the noun pairs or phrases.

First, whether the stimulus nouns are concrete or abstract affects the ability of subjects to use imagery at all. As mentioned earlier, concrete nouns easily elicit images, whereas abstract nouns do not. To the extent that abstract nouns are used in the stimulus pairs, instructions to use imagery are less effective and the experimenter has less control over the imagery manipulation. Several early studies (e.g. Wood, 1967; Crovitz, 1969; Johnson, 1972) used both concrete and abstract nouns. It is difficult to interpret the results of these studies since an investigation of bizarreness requires effective imagery manipulation.

The second factor affecting image formation is the inter-item association strength of the noun pairs. To the extent that the two
nouns in a pair are closely related or frequently occur together (such as table-chair), the manipulation of bizarreness will be rendered less effective by 1) increasing the ease and spontaneity with which common images are formed; 2) increasing the number of spontaneous common images elicited for each noun pair; and 3) reducing the need for facilitation of memory through imagery.

First of all, high-associate noun pairs lend themselves more easily to the formation of familiar commonplace images than to the formation of bizarre or unusual images. For example, the pair table-chair immediately calls to mind a familiar image of some table and chair seen in everyday life. This implies that common images will have two advantages over bizarre images. If a subject is instructed to form a bizarre image for table-chair, it would be difficult to follow those instructions. In addition, if a bizarre image is formed, it would probably not be equal in strength or completeness to the common image. As a result, the manipulation of bizarreness will be impaired.

Another effect of using high-associate noun pairs is that multiple images will be elicited. These images are similar to one another, and serve to intensify the "commonness" of the relationship. In addition, having multiple common images but not multiple bizarre images likewise gives common images an advantage.

Finally, the use of high associates results in reduced reliance on imagery mediation to achieve high recall performance. Specifically, although recall is enhanced as a function of imagery, it is also enhanced by the multiple familiar associations between the two nouns (Paivio, 1971). As the associative strength between the noun pair increases, the reliance on imagery for recall decreases.
For all three of these reasons, then, using high-associates makes it more difficult to manipulate bizarreness. That is, it is more difficult to elicit images on the bizarre end of the continuum; thus, the number and strength of bizarre images will not equal the number and strength of common images. In addition, the high-associate pairs will have a recall advantage over the low-associate pairs.

The problems associated with using high-associates can be seen in the results of a study by Delin (1968), who varied the inter-item association strength of three noun pairs. He then looked at recall as a function of both this manipulation and degree of bizarreness of the image. In general the more bizarre images were recalled better. However, the pair that was highest in inter-item association value was recalled most often, but had the lowest bizarreness rating. Other studies simply did not control the inter-item association strength of the noun pairs (e.g. Wood, 1967; Wollen, Weber and Lowry 1972; Senter and Hoffman 1976). Because of this problem, it is difficult to draw conclusions from these studies regarding the effects of bizarre images on memory.

The third factor which affects image formation is the plausibility or implausibility of the noun pair. In order to elicit bizarre images effectively and avoid the problems mentioned above (with high associates), implausible noun combinations (or phrases) have been used both intentionally and by not controlling degree of plausibility. The problem with implausible noun pairs is that they do not easily lend themselves to the formation of meaningful images. According to ancient mnemonic proscriptions (Yates, 1966) as well as more recent work (e.g. Bower, 1972), meaningless images are difficult to generate and do not
facilitate recall. Therefore, to the extent that noun pairs or phrases intended to arouse bizarre images are implausible or meaningless, images will be difficult to form and again the number and strength of bizarre images will not be equal to the number and strength of common images.

Wood (1967) for example, did not control for either inter-item association value or meaningfulness of the noun pairs. Although he used the same noun pairs for bizarre and common imagery groups, many of the pairs were implausible. It is therefore unlikely that subjects were able to follow instruction to form common and bizarre images.

Collyer, Jonides and Bevan (1972), as mentioned earlier, used implausibility as a definition of bizarreness. Given that the common imagery group was presented with plausible, meaningful noun-verb-noun triplets to elicit images, these images would have been vivid and complete. However, since the bizarre group had implausible noun-verb-noun triplets, these images would have been difficult to form. The authors noted that subjects with bizarre triplets may have formed less complete or vivid images. Consequently, manipulation of the bizarreness dimension was impaired, and memory facilitation via bizarre images would not necessarily be expected.

Wollen, Weber and Lowry (1972) and Senter and Hoffman (1976) used pictures (line drawings) to represent interacting-bizarre, interacting-nonbizarre, non-interacting-bizarre, and non-interacting-nonbizarre conditions. Each noun pair was depicted in both conditions. However, the quality of bizarreness was determined subjectively by the experimenters, and the pairs employed did not appear to lend themselves to equally distinct common and bizarre pictures. In addition, bizarreness
was achieved in the non-interacting bizarre condition by distorting the referents of the two nouns. As a result, pictures in both interacting-bizarre and non-interacting-bizarre conditions were not as readily identifiable by the subject as those pictures in the nonbizarre conditions. This was substantiated by subject reports indicating that a higher percentage of pictures in the bizarre conditions were substituted by subjects' own images than in the nonbizarre conditions. To the extent that subjects had problems identifying the content of the bizarre pictures the probability of finding a main effect for bizarreness would be reduced.

It is clear that there is a conflict between 1) constructing noun pairs/phrases which contain no obvious associates, and 2) constructing noun pairs/phrases which are plausible and meaningful. An attempt must be made to accomplish both of these, but this adds to the difficulty inherent in obtaining the extremes of common and bizarre images for conceptualization and effective manipulation purposes. The fact that it is so important to control for association value and meaningfulness further supports the need for presenting subjects with pre-rated experimenter-generated phrases in order to develop effectively the manipulation of bizarreness and to obtain reliability of the treatment across subjects.

The third category of design criticisms to consider is factors which interfere with retrieval/recall of the correct response words. These are 1) multiple use of stimulus words; and 2) high intra-list association value of nouns. These two factors result in the problem of interference being created either 1) between the response nouns of similar stimulus words, or 2) between the response nouns of multiple uses
of the same stimulus word. This interference inhibits retrieval of the appropriate response noun, and thus potentially decreases the level of recall. In the case of multiple use of stimulus nouns, interference is created due to the fact that the subject is asked to pair different response words to the same stimulus word, and then to remember which response is paired with the different presentations of the stimulus word. As results from Wood (1967), Bower (1972) and Bower and Reitman (1972) have indicated, this interference could minimize the possibility of finding a bizarreness effect and/or lead to alternative explanations.

Hauck, Walsh and Kroll (1976) used a methodology which exemplifies this problem. They instructed subjects to form either common or bizarre images for a list of noun pairs. They repeated this process on five consecutive days, using the same stimulus words each day paired with different response words. It was only at the end of the fifth day that subjects were tested for recall of the response words for all lists. Results showed no difference in recall of nouns from common and those from bizarre images. It may be that no difference was found due to interference washing out any differential effects.

The second issue is the intra-list association value of the nouns. If stimulus nouns have high association value, interference will be created when a given stimulus noun is presented for recall, and several response nouns come to mind. An analogous problem occurs when the response nouns have high association value. Nappe and Wollen (1973) used a list of stimulus nouns which contained obvious associates. As might be expected from potential confounding bizarreness with association value, Nappe and Wollen found no difference in recall between the common and bizarre conditions.
Implications of Existing Literature for the Present Study

A number of issues in the areas of conceptualization, operational definition, and design that have plagued past research on bizarreness have been presented. This section details the ways in which the present study attempts to avoid these problems.

The first problem examined was the lack of a clear conceptualization of the construct of bizarreness. This study approaches the construct as one comprised of multiple dimensions which converge to form the quality commonly called bizarreness. In order to develop a conceptual base, three dimensions related to bizarreness were identified. These were 1) common-bizarre, 2) familiar-unfamiliar, and 3) conceivable-inconceivable. The common-bizarre dimension is the critical one in that it is the basic quality of images under investigation. The familiar-unfamiliar dimension may also be related in that the familiarity dimension may enhance the bizarre or nonbizarre quality of images. That is, generating images that are both unfamiliar and bizarre may intensify the bizarreness quality. More importantly, unfamiliar images may result in less interference than familiar images since familiar images are less distinctive. This implies that controlling familiarity of the images should lead to more distinct images in the bizarre group. Thus, one aspect of bizarreness which would lead to better recall is unfamiliarity. Finally, the conceivable-inconceivable dimension may be related to bizarreness in ways similar to unfamiliarity. That is, inconceivable images imply something wild and very much out of the ordinary. Inconceivability, then should further increase the distinctiveness and the bizarre quality of the images.
The use of these three dimensions, then, defines the quality of bizarreness for this study. Furthermore, through the operationalization described below, these three dimensions allowed for the generation of clearly distinct groups of common and bizarre stimulus materials.

The second problem identified in past research involved operationalization; in particular, problems related to the induced-imposed issue and the subjective-objective rating issue. It was concluded earlier that the most appropriate strategies are 1) to use the imposed method of imagery manipulation, and 2) to rate objectively the bizarreness of phrases and select phrases that are consistent with the concept of bizarreness to be used as stimulus materials.

Specifically, in this research, ratings were obtained on phrases as to their degree of bizarreness, unfamiliarity and inconceivability. Based on these ratings the phrases in the "bizarre" group are those rated as being high on bizarreness, unfamiliarity, and inconceivability. Those in the "common" group are those rated as being high on commonness, familiarity, and conceivability.

The last set of problems found in existing research were problems of methodology. Basically, this research attempts to eliminate past design problems by controlling for potential confounds, and maximizing the chance for effective image formation, bizarreness manipulation, and recall. First, the present study uses a control group instructed to image nonbizarre (common) phrases so that the effects of imagery can be separated from those of bizarreness. Second, interaction between the noun pair is present in all phrases so that degree of interaction will be held constant. Third, only nouns which are high in concreteness, meaningfulness and imageability are used in order to ensure complete, vivid
images. Fourth, there are no noun pairs which are obvious associates. Thus, a stronger bizarreness manipulation is possible. Fifth, there are no obviously implausible noun pairs/phrases so that the formation of meaningful images is enhanced. Sixth, an attempt is made to use nouns with low intra-list association value in order to minimize interference in recall of the correct response word. Finally, each stimulus and response noun is used only once, again to avoid interference in recall.

In addition to investigating the contribution of bizarreness in a way that attempts to avoid the problems encountered in previous research, the present study attempts to extend such research by considering two additional variables. The first variable is retention interval. Most studies have looked only at the effects of bizarre imagery on immediate recall. However, mnemonists suggest that bizarre imagery is most useful for long-term retention. The argument is that bizarre imagery helps minimize the effects of interference that occur over time and weaken recall. Only two studies have looked at long term retention (Delin, 1968; Andreoff and Yarmey, 1976). Both studies found evidence supporting the hypothesis that bizarre imagery facilitates recall on a long term retention test. In addition, Andreoff and Yarmey found that superior recall of bizarre over common images increased after a one-week interval. This issue is examined in the present study by testing subjects both immediately after presentation of the stimulus material, and again after one week.

The second variable to be considered is rehearsal. Ancient mnemonic techniques proscribe rehearsal of the image to increase its resistance to interference (Yates, 1966). Thus, it may be that any
beneficial effects of using bizarre images can be increased with such rehearsal. To explore this hypothesis, the present study presents phrases at different frequency levels. In this way, the amount of rehearsal is controlled, and the effects of different levels of rehearsal on recall can be assessed for both common and bizarre images.

In order to assess the effect of the manipulated variables, two dependent variables are used. The first is a cued recall task and the second is a frequency estimation task. Cued recall is the primary measure of recall in this study and also has been used in the majority of studies reviewed. The frequency estimation task is primarily an exploratory measure. It has been found that subjects can estimate fairly accurately the number of times they have seen a stimulus item (Peterson and Beach, 1967). Therefore, frequency estimation should prove useful in interpreting the effects of the different frequency levels of presentation and is one other aspect of memory that could help clarify the contribution of bizarreness.

Hypotheses

With these general issues in mind, we now turn to the specific hypotheses of the study. First, recall should be higher when subjects are instructed to image than when they are instructed to simply rehearse. This can be thought of in some ways as a check on the imagery manipulation in that this finding is clearly established and should be replicated here. Second, recall for bizarre phrases is expected to be higher than for common phrases. This is the basic prediction of the study. Third, based upon previous research, general recall should be higher immediately after stimulus presentation than
after the one week retention interval. Fourth, consistent with the assumption that the superiority of bizarre over common images should increase over time, it is predicted that the difference between recall for nouns from bizarre and common images should be greater in the delayed recall test than in the immediate test.

No specific predictions are made regarding the exploratory measure of frequency estimation other than that frequency estimates should increase as frequency of presentation increases.
Method

Subjects and Design

The subjects were 68 Rice University students who participated in the study for course credit. Four subjects were discarded—two for failure to follow instructions and two for failure to complete the experiment. Data from 64 subjects (32 males and 32 females) were analyzed.

The basic task for the subjects was to learn phrases of various types. There were four independent variables in the study. The first variable was instructional condition: half the subjects were asked to form actual images of the phrases presented, while the other half were asked to rehearse the phrases in rote fashion.

The other three independent variables in the design were within-subject factors. The first of these was the bizarreness manipulation where half the phrases used for a given subject were bizarre and half were common. The second within-subjects factor was retention interval, with subjects tested immediately after the list presentation and again after one week. The last within-subjects factor was frequency presentation level with each phrase presented 1, 2, 4, or 6 times.

Finally, two major dependent variables were used to assess recall: a cued recall task and a frequency estimation task. However, a given subject was tested on only one of these dependent variables.
Stimulus Materials

The stimulus materials consisted of 32 phrases selected from 142 phrases rated in preliminary research. This preliminary research was conducted in order to accomplish three objectives: 1) to set up a conceptual base to define the quality of bizarreness; 2) to manipulate bizarreness by using imposed imagery techniques; and 3) to determine the degree of bizarreness associated with the stimuli through an objective rating strategy. To accomplish these objectives, three groups of 40 subjects rated 142 phrases (using a 7-point scale) on one of three dimensions thought to be part of the quality called bizarreness. These three dimensions were 1) common-bizarre, 2) familiar-unfamiliar, and 3) conceivable-inconceivable. The phrases to be rated were derived by first selecting 88 nouns having imagery ratings of at least 5.0 and concreteness ratings of 5.8 or higher from the Paivio, Yuille & Madigan (1968) norms, as well as frequency ratings ranging from 1 to AA/million in the Thorndike and Lorge (1944) frequency count. Next, the nouns were divided into 44 pairs, avoiding any obvious interitem associations, such as table-chair. Several phrases were constructed for each noun pair, on the basis of three primary considerations: 1) the stimulus and response terms (the noun pair) of each phrase must be interacting in some way; 2) the range of phrases included should span the continuum of the three dimensions; and 3) all phrases, no matter how strange or unusual, should arouse an image. After the phrases were rated, means and standard deviations were computed for each dimension, as well as for each dimension collapsed across phrases. These statistics are included in Appendix A.
The 32 stimulus phrases were then selected to form the two distinct groups corresponding to the concept of common and bizarre phrases. The group of common phrases was formed by selecting 16 phrases which were rated as at least one-half a standard deviation below the mean on all three dimensions. The group of bizarre phrases was formed by selecting 16 phrases which were rated as at least one-half standard deviation above the mean on all three dimensions. Thus, the common phrases selected for the study were those rated as common, familiar, and conceivable; while the bizarre phrases selected were those rated as bizarre, unfamiliar, and inconceivable. The resulting list of 32 phrases appears in Appendix A.

These 32 phrases, with repetitions, totalled 104 presentations. One basic order for phrase presentation was constructed. First, the 104 list positions were partitioned into six segments with one presentation each for the eight phrases (frequency level = 6) assigned to each segment. Next, the positions were divided into four equal segments for assignment of frequency level = 4 phrases and two equal segments for assignment of frequency level = 2 phrases. Frequency level = 1 phrases were assigned one each to eight segments. The minimum lag between identical repetitions was 9 list positions, and no more than three common or bizarre phrases were presented in sequence. In order to counterbalance for specific-phrase effects, each phrase appeared at every frequency level across subjects. To achieve this control, it was necessary to have four different phrase lists, while maintaining the basic order. Once the order had been constructed, the four different lists were derived by first randomly assigning the phrases to the four frequency levels for List A, then repeating the process for List B with the restriction that
the phrase had not already been assigned to that frequency level, and repeating this process each time for Lists C and D. Thus, the positions occupied by different frequency level phrases were fixed by the basic order, but the phrases presented in these positions were different in each of the four lists. In addition, each phrase occurred with a different set of phrases at different frequency levels.

Presentation of the stimulus phrases was carried out in an experimental cubicle using a PDP-8L computer equipped with a remote TV monitor. Each phrase appeared typed in the middle section of the screen with a subject and a predicate noun (keywords) enclosed in parentheses. The subject noun would be used to cue recall of the predicate noun in the test phase of the experiment. In addition, the word COMMON or the word BIZARRE (identifying the respective phrases) appeared in the lower left corner of the screen. The purpose of this identification was to ensure further that the images formed by subjects in the imagery condition would be appropriately common or bizarre, and was similar to procedures carried out by Nappe and Wollen (1973) and Andreoff and Yarmey (1976).

**Procedure**

Upon arrival, each subject was assigned randomly to one of the four experimental conditions and one of the four control lists, with the restriction that an equal number of males and females be assigned to each. The subject was given a typed sheet containing the instructions for the experimental session. Subjects in the imagery condition were instructed to form an image of what the phrase represented, and to concentrate on the image until the next phrase was presented. Subjects in the rehearsal condition were instructed to repeat the entire phrase over and over until
the next phrase appeared on the screen. Both groups were also told to pay particular attention to the keywords and to whether the phrases were common or bizarre, but to learn the whole phrase. An example was given, along with a picture of a phrase as it would appear on the screen. The word READY then appeared on the screen, followed by presentation of the phrases. Each phrase remained on the screen for 6 seconds, with a 1 second pause between phrases. The 6 second presentation rate was determined on the basis of pilot testing. Results from this pilot work indicated that a 10 second presentation time resulted in rehearsal of previous phrases, and three seconds was not sufficient to form a complete image. Total presentation time for the entire set of phrases was about 15 minutes.

When the last phrase had disappeared from the screen, each subject was given a questionnaire which asked if he/she had 1) used the instructed form of mediation (i.e. imagery or rehearsal); 2) used an alternative or additional form of mediation; and 3) actually made use of the imposed phrases to learn the keywords. Each subject was then given either the cued recall or frequency estimation task. In the cued recall conditions, subjects were asked to write the response keyword beside the stimulus keyword. In the frequency estimation task subjects were given the list of phrases and asked to estimate the number of times each phrase was presented. Unlimited time was allowed for both of these tasks. The subjects were told that the experiment had one additional part and were instructed to return one week later. During the session one week later, subjects were retested on the same task they had received for the immediate test.
Results

As discussed previously, two dependent variables were used: cued recall and frequency estimation. The results of the cued recall variable will be presented first, followed by those of the frequency estimation variable.

Figure 1 presents the recall data in terms of mean number of phrases recalled correctly as a function of instructional condition, type of phrase, and retention interval. Inspection of the figure indicates that the prediction that recall would be higher under imagery than under rehearsal instructions was confirmed, $F(1,30)=4.52; p < .04$.

The finding of greatest interest was that as predicted, overall recall of bizarre phrases was significantly better than of common phrases, $F(1,30)=7.65; p < .01$. Also as expected, retention declined considerably over the week intervening between the immediate and delayed tests, $F(1,30)=140.15; p < .001$. It was also predicted that for the imagery condition the superiority of bizarre images over common images would be greater after the one week delay than in the immediate test. Inspection of the figure suggests that this is indeed the case, and an a priori comparison of the means indicates that the difference between the differences is significant ($t(30)=1.73; p < .05$).

Examination of Figure 2 reveals that the effects of frequency level (presentation frequency) on cued recall were quite pronounced, $F(3,90)=31.69; p < .001$. That is, recall increased substantially with
presentation frequency. Figure 2 further shows that frequency level interacted with retention interval, $F(3,90)=5.95; p < .001$. In the immediate test, recall increased rapidly between 1 and 2 presentations and reached a ceiling at frequency level 4. In the retest condition, total recall was much lower, with a more consistent increase in recall with presentation frequency.

The second dependent variable was frequency estimation. The mean frequency judgments for phrases presented at the different frequency levels as a function of type of phrase and retention interval are presented in Figure 3. As expected, there was a highly significant main effect of frequency level on the mean frequency estimates, $F(3,90)=304.78; p < .001$. That is, frequency judgments increased quite reliably with presentation frequency. The mean frequency estimate was lower in the retest than in the immediate test, $F(1,30)=3.86; p < .001$. However the retention interval by frequency level interaction was significant, $F(3,90)=26.75; p < .001$. Inspection of the means reveals that the immediate and one-week retention tests reflected different patterns of frequency judgments. The data indicate that the frequency judgments displayed a more restricted range in the retest than in the immediate test. Specifically, subjects slightly overestimated the true low frequencies and underestimated the true high frequencies. This restriction in range is a typical finding, and generally is more pronounced in a delayed test than in the immediate test (e.g. Peterson and Beach, 1967; Begg, 1974).

A main effect for type of image was not found in the frequency estimation task. In fact, the overall mean frequency judgments were identical for common and bizarre phrases, collapsed across frequency levels.
Figure 1. Recall by Instructions, Session and Type.
Figure 2. Recall by Session and Frequency.
There was one other significant interaction in the frequency estimation data, i.e. the interaction of retention interval with type of phrase, $F(1,30)=10.44; p < .003$. The appropriate means are plotted in Figure 4 where it is apparent that in the immediate test the common phrases were estimated at a higher frequency than were the bizarre phrases, but this pattern was reversed after one week.

Taken as a whole, the major results show that for cued recall, imagery is superior to rehearsal, and bizarre phrases are superior to common phrases. As expected there was a decrease in recall after one week, and the higher the presentation frequency, the better the recall. In addition, the superiority of bizarre over common imagery was greater after one week than immediately after presentation of the stimulus material. Finally, increased frequency level showed a somewhat different pattern of facilitory effects for the immediate as opposed to the delayed recall test.

The frequency estimation data were less sensitive to the manipulated variables. As expected, the higher the frequency of presentation the higher the estimated frequency, and estimated frequencies were lower after one week than for the immediate test. Finally, the estimated frequency for common and bizarre phrases was somewhat different for the immediate and retest conditions and the effects of varying frequency of presentation were somewhat different for the immediate and retest conditions.
Figure 3. Frequency Estimates by Session, Type and Frequency Level.
Figure 4. Frequency Estimates by Session and Type.
Discussion

The most interesting outcome of this study was the finding that bizarre phrases were in fact recalled better than common phrases. Thus, when attempts were made to eliminate the conceptual, operational definition, and design problems discussed in the introduction, the results were supportive of the facilitory effects of bizarreness.

However, there is one problem with this interpretation which is related to the nonsignificant instruction by type of phrase interaction. Specifically, it was predicted that bizarreness should enhance the effectiveness of imagery. In fact, bizarre phrases enhanced recall in both the imagery and rehearsal conditions. While it was not formally predicted, it might have been expected that there would be no difference in recall between common and bizarre phrases when subjects were asked to merely rehearse. That the difference did, in fact, occur could be seen as casting doubt on the advantages of bizarreness for imagery.

However, several factors in the present study serve to reduce the importance of this finding. Specifically, there are reasons to expect that considerable imagery was occurring in the rehearsal condition, even though subjects were not instructed to image. In the first place, it is well established that imagery can be induced by stimulus characteristics such as concreteness and imagery value of the nouns, as well as by the experimental manipulation of instructions (see Paivo, 1971; Reese, 1977). The use of high-imagery concrete stimuli in the present
study would therefore be expected to elicit spontaneous imagery regardless of the instructions given. Second, post-experimental questionnaire data indicated that subjects in the rehearsal condition were, in fact, using a considerable amount of imagery. Thus, to the extent that subjects were using imagery in the rehearsal condition, one would expect bizarre phrases to be recalled better than common phrases.

A second major finding of this study was that the facilitative effects of bizarreness in the imagery condition were greater in the retest than in the immediate test. That is, the major difference in recall for words from common and those from bizarre phrases occurred in the retest condition. Although as mentioned earlier, most studies investigating bizarreness have not addressed long-term retention, ancient treatises definitely stressed that bizarre images are especially useful because they are more resistant than commonplace images to interference, fading, etc. over time (see Yates, 1966). Results of the present study are consistent with this belief, and replicate findings in the retest of Andreoff and Yarmey (1976) and in the long term test of Delin (1968). This indicates that part of the key to demonstrating a strong and replicable bizarreness effect experimentally may lie in testing long-term retention.

There were three other findings in the recall data that are worthy of mention in that while they are not surprising, they were predicted from previous research and, in fact, appeared. These were that imagery produced better recall than rehearsal; recall was better in the immediate test than in the delayed test; and the higher the frequency of presentation of the phrase, the better the recall. Taken together, these
findings demonstrate that the study produced differences where differences were expected.

The results were much less striking for the frequency estimation data. As predicted, frequency estimates increased as actual presentation frequency increased. This could be considered more of a check on the manipulation than a substantive finding. It was also found that upon retest subjects overestimated low frequencies and underestimated high frequencies. This finding is typical of previous research and does not really shed any light on the primary concern of the present study. Finally, collapsed across frequency levels, common phrases received higher frequency estimates in the immediate test while bizarre phrases received higher frequency estimates in the retest. However, this interaction is not readily interpretable.

Probably the more important results with the frequency estimation data deal with what was not found. There was no significant effect for instructional condition or type of phrase. While using frequency estimation as an index of memory was exploratory in this study, it seems that it was not particularly useful as a dependent measure in demonstrating differences between recall for common and for bizarre phrases.

Several points are worth mentioning in terms of directions for future research. To reemphasize, when attempts were made in this study to eliminate previous research problems, a bizarreness effect was found in both immediate and long-term retention tests. In addition the superiority of bizarre images increased between the immediate and delayed tests. Taken together, these findings lend sound positive support to bizarreness as a critical attribute of images.
However, extensions of the approach used in this study would help to clarify further the concept of bizarreness and determine the facilitory role bizarreness plays in imagery mediation. First, although the present study made progress toward developing a conceptual base for bizarreness, additional research in this area would be useful. Specifically, three dimensions were identified to define the bizarreness construct. It may be that additional dimensions converge to form this facilitative quality of images. Several such dimensions might be distinctiveness (discussed by Saltz, 1963), vividness (Neisser and Kerr, 1973), novelness (Bower, 1969), uniqueness, etc. To accomplish this, techniques (similar to the ones used in this study) could be employed where subjects are instructed to rate phrases on the multiple dimensions. The two groups of phrases rated at the extremes of these dimensions would then be selected for the experiment. Alternatively, it would be interesting to try to establish more empirically what people are referring to by "common" and "bizarre" images; that is, defining bizarreness according to what it means for different people. This approach would involve just having phrases rated on the bizarreness dimension without specifying a definition for subjects to use. Likewise, these same phrases would be rated on multiple other dimensions. Finally, through correlational analyses the experimenter could identify those dimensions which relate most highly with bizarreness as generally defined. Then the value of bizarreness could be tested by conceptualizing bizarreness in terms of these dimensions and selecting for stimulus materials only those phrases which are rated at the extremes of these dimensions.

A second variable which might be a contributor to the facilitative effects of bizarreness is rehearsal. In this study, frequency of
presentation was used in an exploratory fashion to assess the effects of controlled rehearsal. The manipulation of this variable did not appear to reveal any meaningful differences between common and bizarre phrases in cued-recall or frequency estimation. Another way to incorporate rehearsal into research is through elaborative rehearsal as opposed to maintenance (Craik and Lockhart, 1972). Typically, it takes longer to form a bizarre image than a common one (e.g. Nappe and Wollen, 1973). Perhaps though, the distinctiveness of bizarre images would increase more with elaboration than would the distinctiveness of common images. Using frequency of presentation to vary the amount of rehearsal, combined with instructions to elaborate the image with each presentation could clarify whether rehearsal to develop a more fully elaborated image enhances the bizarreness effect.

A final line of investigation which appears to hold promise for being related to the bizarreness effect is encoding uniqueness. Lesgold and Goldman (1973) looked at the effects of unique encoding (of the relationship between the cue words and response words) on cued-recall. Their evidence indicated that a focus on unique imagery encoding increased recall of the cue word over a condition of non-unique encoding. They suggested that perhaps this "encoding uniqueness" variable could explain the mnemonic success attributed to bizarreness. Although the present study supports bizarreness as a facilitory variable, the relationship between bizarreness and unique encoding might further clarify which aspects of bizarreness do account for its facilitory effects.
References


Bower, G. H. How to ... uh ... remember. Psychology Today, October 1973, 63-70.


Delin, P. S. Learning and retention of English words with successive approximations to a complex mnemonic instruction. Psychonomic Science, 1969, 17, 87-89.


### Appendix A

**I. Statistics for 144 pre-rated stimulus phrases:**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common-Bizarre</td>
<td>3.74</td>
<td>2.43</td>
</tr>
<tr>
<td>Conceivable-Inconceivable</td>
<td>3.16</td>
<td>2.46</td>
</tr>
<tr>
<td>Familiar-Unfamiliar</td>
<td>4.47</td>
<td>2.35</td>
</tr>
</tbody>
</table>

**II. Each of the 32 selected stimulus phrases had means that were at least .5 standard deviation from the mean of all three dimensions.**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Phrase Mean for &quot;Common&quot; Group</th>
<th>Phrase Mean for &quot;Bizarre&quot; Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common-Bizarre</td>
<td>&lt; 2.52</td>
<td>&gt; 4.96</td>
</tr>
<tr>
<td>Conceivable-Inconceivable</td>
<td>&lt; 3.29</td>
<td>&gt; 4.39</td>
</tr>
<tr>
<td>Familiar-Unfamiliar</td>
<td>&lt; 1.93</td>
<td>&gt; 5.65</td>
</tr>
</tbody>
</table>

**III. List of 32 stimulus phrases:**

**Common**

1. Holding a foam pillow in your arms  
2. A football leader watching the clock  
3. A child hitting its fork on the table  
4. A teacher talking to the village  
5. A king holding butter on the Imperial commercial  
6. Wiping an elbow clean with cotton  
7. Dirt blowing into an open cellar  
8. A coffee stain on a book  
9. An arrow covered with blood  
10. Drinking wine on a ship  
11. Reading a letter in front of a warm fire  
12. A doll with a flower in her hair  
13. A judge riding his horse  
14. A cat licking jelly off of a plate  
15. A hammer lying in the grass  
16. A maiden wearing a party costume
Appendix A (continued)

Bizarre

1. Biting bullets which melt like candy
2. An infant erased by a twirling pencil
3. A leopard buying a journal
4. A chin swimming in a lake
5. A door smoking a cigar
6. A mule defrosting an icebox
7. A frog with lemons for eyes
8. A prison on the floor of the ocean
9. Stuffing a brain with pudding
10. Shoes driving a car
11. Money growing in a window
12. A lobster being marched to jail
13. A cabin sneezing pepper
14. A tall wheat growing in a bowl
15. A priest walking across the water to a yacht
16. Baking a painter in the oven