ON THE FUZZINESS OF LEXICALITY

By

ROBYN STELLMAN

Bachelor of Science
University of Illinois
Urbana, Illinois
1985

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1992

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 1996

ON THE FUZZINESS OF LEXICALITY

Thesis Approved:

Larry Hochhaus

Thesis Adviser

Richard Potts

Dominic W Massaro

Rebekah [illegible]

James M. Price

Thomas C. Collins

Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to my major advisor, Dr. Larry Hochhaus, for his seemingly unending patience, constructive criticism, and unfailing positive encouragement, no matter how many revisions he had to read. I hope I can show as much patience and kindness with anyone I may work with in the future. I would also like to thank my other committee members from Oklahoma State University for agreeing to sit on a committee for a topic of which they had no prior knowledge or experience, and to Dr. Dominic Massaro of the University of California at Santa Cruz for providing his expertise in the review of this study.

I also would like to thank my family, however scattered about the country they may be, for their continued encouragement and support throughout my graduate career. I would also like to thank my close friend and constant source of inspiriation, Gabby Richards, who never fails to impress me that no matter what happens, it is possible to live one's dreams.

Finally, I would like to express my thanks to all my co-workers at Pizza Hut number 469, but especially Terry Williams, Cam Botchlet, Jeremy Henderson, Carla Hyatt, Julie Smith, and Travis "Brillo" Jackson for helping me to keep my perspective about where the real world is, and for all their encouragement, understanding, and patience. Because when all is said and done, we all know that Ph.D. stands for "Pizza Hut Delivers."

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF TABLES

LIST OF FIGURES

Chapter One: Introduction

"Everything is vague to a degree you do not realize until you have tried to make it precise." - Bertrand Russell (as cited in Kosko, 1993)

In these days of high technology, precision appears to be of paramount importance. Classifying items as A or not-A is the goal of various software applications throughout the world. Is the stock price high enough to sell? Is the satellite within radio range? Do the symptoms correspond to a given diagnosis? Is the pattern an "A?" Is the letter string a word? Such examples point to the tradition in Western culture of the Aristotlean approach of classification. The present project expects to show that the A or not-A approach to lexicality ("wordness") overlooks a rich source of information. The idea that all letter strings belong to one of two piles (words or not-words) may constrict the amount of information available from data collected in the typical lexical-decision task paradigm. The present results should indicate whether people process letter strings in a continuous ("fuzzy") manner, rather than an either-or (Aristotlean) approach. The remainder of the current section describes some issues in the argument of using bivalence versus multivalence to describe phenomena. Subsequent sections in this chapter explain the fuzzy logic model of perception, word recognition theories, and a fuzzy logic model of lexicality.

The history of modern science is deeply rooted in bivalent, either-or explanations of phenomena. However, there are a few drawbacks to the Aristotlean approach. One problem is that most measured quantities are continuous. When a person claims to weigh 150 pounds, this value has been rounded off. Very few (if any) individuals weigh exactly 150 pounds. Continuous quanitities are rounded off all the time, especially to fit constraints of a computer's 1s and 0s. This rounding error is labelled the <u>quantization problem</u> (Kosko, 1993). Parts of reality are discarded to "fit the grid." Compact disc audio technology was heralded as cleaner than analog recording techniques, yet serious audiophiles complain that the quantization used in digital recording robs the sound of its richness. Even at 44,100 samples per second, parts of the signal are lost to fit the grid (Kosko, 1993). The present disucussion will suggest that current lexicality models lose information due to the quantization problem.

Given a glass with a little water left in it, people round off the amount of water and call the glass empty. As a result, the 5% or 10% left in the glass has been quantized to 0%. Similarly, a glass filled to 80% or 90% of capacity is rounded to 100% when quantized. A glass filled to exactly 50% of its capacity presents a problem to bivalent logic, namely how to round 50%. Does 50% get rounded to 0% or 100%? Bivalent logic demands an answer to the question "Is the glass full or empty?" Similarly, standard lexical decision paradigms sort letter strings into two categories, "word" or "not word." How do people classify a word they commonly misspell? Technically it is not a word, but to the misspeller it is. Again, this example illustrates how information may be lost when the only alternatives are yes or no.

Another classic example of the limitations of bivalent logic is found in Bertrand Russell's barber example. In this example a barber posts a sign in his shop that states "I shave all, and only, those men in town who do not shave themselves." Who shaves the barber? If he shaves himself, he contradicts his sign. If he does not shave himself, his sign claims that he does. He appears to shave and not shave himself at the same time. This problem is deemed a paradox by bivalent logic standards.

The examples of the glass and the barber are called midpoint phenomena (Kosko, 1993). A fuzzy interpretation is one way to avoid these midpoint phenomena. A continuum of response values allows the glass to be half full or half empty. Because bivalence operates on the law of the excluded middle, the barber and the glass present a paradox to the Aristotlean dichotomy. Multivalence allows the midpoint so that a glass can contain 50% of its capacity. An answer of "half" is acceptable in a multivalence approach.

A more modern example considers a new car owner's pride and joy parked in two spaces to avoid those dreaded first door dings. The bivalent view states that the car is parked in whichever space is most covered by the vehicle, necessarily rounded to one car per space. In the bivalent case, the statement "I parked in space A" is either totally true or totally false. Multivalence, however, allows a car parked in several spaces, most of them to degree 0%. Perhaps the new car owner parked 80% in space A and 20% in space B. Then the statement "I parked in space A" is 80% true and 20% false. Fuzzy logic avoids the quantization problem ("I parked in space A" can only be true or false) through the use of intact information ("I parked in space A" is true to the degree 80%).

Many critics of fuzzy logic claim that it is just probability theory in disguise. This claim is false. When a person draws an ellipse, the ellipse can be viewed from a fuzzy perspective (it is a fuzzy circle) or a probabalistic perspective (it is probably a circle). Which perspective is more accurate? Kosko (1993) asks the crucial question of where to find the randomness required by the probability view. In his argument that fuzziness is not probability, Kosko states that probability is a fiction used to round off the excuded middle. Kosko also claims that probability just predicts long-term average behavior based on the past, e.g., flipped coins, but that no one can catch probability in the act. Probability may give structure to competing hypotheses about how the future unfolds, but probability is really all in one's mind. For the ellipse as fuzzy circle example, one could base the fuzziness of the circle on the ratio of the major and minor axes. When these axes are equal, the ellipse is a circle to the degree 100%. When one axis is infinite and the other is zero, the ellipse is a line and thus a circle to the degree 0%. There is no similar analog to the probabilistic view of the ellipse. One may base whether the ellipse is classified as "probably" a circle by polling a number of raters, but there is still no reality involved. There is a big difference between an ellipse with axes of ratio 0.5 (a circle to 50%) and a 50% chance of a shape being labeled a circle.

Similarly, there is also a big difference between a letter string rated with degree of "wordness" 50% and a letter string with a 50% chance of being a word.

From another angle, probability decreases when precision increases. More information diminishes the need for probability. However, when precision increases, fuzziness increases as well (Kosko, 1993). More information helps to pin down the exact shape of the curve between "thing" and "non-thing." Bivalence only describes whether an item is considered part of a category or not. Mulitvalence describes the boundary between thing and non-thing. Therefore, more information only helps clarify which items go on which side of the boundary in the bivalent case, whereas in the multivalent case more information helps clarify the actual boundary.

The present study investigates the multivalent perspective as applied to lexicality ratings. The bivalent perspective considers letter strings as either words or nonwords. More examples of letter strings increase an individual's awareness of what is or is not a word. Neural networks are excellent bivalent classifiers. Seidenberg and McClelland (1989) designed a neural network which classified letter strings as words or nonwords with fairly good (85% accuracy) success. However, the multivalent case describes word strings with degrees of wordness. An increase in the exposure to letter strings and a broad range of the degrees of wordness increases an individual's awareness of the factors that actually contribute to the lexical decision. The main research question of the present study is whether lexical information is lost due to the quantization problem. To answer this question, individuals will be allowed to rate letter strings with degrees of lexicality, as opposed to traditional yes/no approaches. The data should show that individuals process lexicality information according to a multivalent approach.

A Model of Perception

Massaro's (1988a) fuzzy logic model of perception (FLMP) is used as the backbone for the present investigation. This flexible model has been applied to speech perception (Massaro, 1988a, Oden & Massaro, 1978), pattern recognition (Massaro, 1994), and depth perception (Massaro, 1988b). The FLMP for lexicality will be used in the present investigation as a starting point to determine if people will consistently rate letter strings according to a multivalent approach.

The FLMP is a model consisting of three operations - <u>feature evaluation</u>, <u>feature integration</u>, and <u>decision</u> (see Figure 1). Features can be any physical event transduced by the sensory system and are continuously valued from 0 to 1. Features are assumed independent of one another in the detection stage of processing (Oden & Massaro, 1978). Examples of features are the acoustic components of speech syllables such as place and voicing (Oden & Massaro, 1978), auditory and visual cues in speech perception (Massaro, 1988a), and depth perception cues such as size, height, occlusion, and motion parallax (Massaro, 1988b).

During the <u>feature evaluation</u> stage, features are evaluated according to prototype descriptions in memory. Features are assumed to be independent - "the value of one feature does not influence the value of another" during the feature evaluation stage (Oden & Massaro, 1978, p. 173) Prototypes, which are representations in long-term memory coincident with perceptual units, are generated for the task at hand. Prototypes are not just a collection of features, "but rather are propositions that may be, in principal, arbitrarily rich in logical structure" (Oden & Massaro, 1978, p. 174). For example, the phoneme /b/ would be represented by the propostition (labial) AND (voiced). Similarly, /p/ would be represented by the proposition (labial) AND [not (voiced)] (Oden & Massaro, 1978).

Features of prototypes correspond to ideal values that an exemplar should have if it is a member of a given category. Feature evaluation provides information concerning the degree of match between the feature of the stimulus and the corresponding feature in the prototype. The FLMP assumes that a feature is not simply detected, but is "perceptually more or less present" (Oden & Massaro, 1978. p. 175). For example, Massaro (1988a, p. 129) explains

> Although the prototype /ba/ is supported to degree .999 by a visual /ba/ articulation, the prototype for /bda/ also is supported .331. Thus, a visual /ba/ is fairly consistent with the alternative /bda/. Similarly, a visual /da/ supports not only /da/ .937, but also /tha/ to degree .535.

During the <u>integration stage,</u> features within each prototype are conjoined. The result of this combination is that the FLMP reports back the degree to which each protoype matches the stimulus. In other words, prototypes are generated for the task, the stimulus is broken down into features, the degree of match between stimulus feature and prototype feature is produced, and the degrees of all the matches are combined. The result is an overall degree of match between the stimulus and each of the generated prototypes.

Finally, a <u>decision</u> is made as to the match of the relevant prototype relative to the sum of the matches of all the relevant prototypes. The sum of the matches gives a rating judgment of the degree to which the stimulus matches a prototype. Some prototypes may match the stimulus to the degree of 0%, others may match to the degree of 50%, while still others may match to the degree 90% or 100%. The relative goodness of match provides a rating judgment corresponding to the degree to which the stimulus matches the generated category (Massaro, 1988a, 1993). To put it another way, because perception is a noisy process, the decision stage is a probabalistic process due to the fact that a given physical stimulus can be perceived different ways at different times. As a result, the probability of identifying a given stimulus should correspond to the goodness of match for that prototype relative to the sum of the goodness-of-match values for all prototypes under consideration (Oden & Massaro, 1978).

The processes of the FLMP just described may appear very similar to Selfridge's (1959) pandemonium model. In that model, Selfridge's computational demons would do the feature evaluation, the cognitive demons would perform the prototype match, and the decision demons would choose among the prototype alternatives. Oden and Massaro explain the differences in the models:

> [W]hereas the fuzzy logical model is intended to describe the cognitive processes that are actually used by humans to identify speech sounds, Selfridge (1959) was primarily interested in the problem of learning to make correct identifications. Thus, Selfridge concentrated on how Pandemonium might be made to come to discover the appropriate features and feature integration rules over the course of training trials. In contrast, the [fuzzy logical model]...relied on intuition and on the analytic linguistic description of the phonemes to formulate the corresponding prototype specifications (Oden & Massaro, 1978, p. 187).

Massaro (1988a, 1988b, 1993, 1994) and Oden and Massaro (1987) have tested the FLMP in several different areas (speech perception, pattern recognition, and depth perception). Most of this work has dealt with speech perception, and the findings have been consistently in favor of the FLMP over more traditional models such as a categorical model (Massaro, 1988a) or additive model (Oden & Massaro, 1978). In the speech perception domain, Massaro (1987) has successfully demonstrated the FLMP fits observed data significantly better than a categorical representation using a visual continuum only, an auditory continuum only, combined visual and auditory contiua, forced choice trials (rate the given syllable as /ba/ or /da/), open-ended trials (rate the given syllable as /ba/, /da/, /dba/ (a /ba/-/da/ hybrid), /va/, /ga/, or other), and across age groups. Oden and Massaro (1978) varied voice onset times and place of articulation of speech syllables and found the fuzzy model fit the data better than the more traditional additive model.

Similarly for depth perception, the FLMP was found to fit the data significantly better than an additive model of perception (Massaro, 1988b). Oden (1977) also found a fuzzy model provided an excellent fit for the degree of match of exemplars to categories (A robin is a bird, A butterfly is a bird). Because of the consistent findings of fuzzy model studies, the FLMP will be used in the present experiments as a springboard by which to investigate whether people will rate letter strings according to a fuzzy model.

One of the main ideas of the FLMP used in the present investigation is that as one informational component becomes ambiguous or unusable, individuals will rely on other available informational components to make response decisions. For example, in the speech perception domain, if the auditory component becomes ambiguous, individuals have been found to rely on the visual component (Massaro, 1988a). Similarly, the phenomenon of the "McGurk effect" indicates that visual information can influence what is heard when the combination of nonsensical visual and auditory stimuli are interpreted as a meaningful speech event (Massaro, 1988a). One of the main working assumptions of the present study is that, when rating a letter string's degree of wordness, people will rely on the phonological component of the string if the orthographic component gives no helpful information and vice versa.

The present project is undertaken on the basis of two working assumptions. The first, the substitutability assumption is consistent with the FLMP in that individuals will rely more on alternative information sources when a particular source is ambiguous. In the speech perception case, if the visual information is ambiguous, individuals will rely on the auditory component of the stimulus for response selection. Similarly for depth perception, if occlusion is ambiguous, individuals will rely more on size, height, and motion parallax for their final depth perception rating. The substitutability assumption in the present study is that, as one component of written language becomes ambiguous, the other(s) will be relied upon more to determine the degree of a letter string's wordness. For example, given two illegal nonwords (e.g., BNARE, FTARI), the decision of which is more wordlike may rest on the stimuli's respective neighborhoods or bigram frequencies because other components of written language (e.g., word frequency, phonology) are of no help to the judgment of wordness for illegal nonwords.

The second, generalizability assumption of the present study claims that a fuzzy model of lexicality will agree with current theories of word processing. In other words, the fuzzy model of lexicality should support previous lexical processing results. An example would be that, all other components being equal, wordness may be rated as a function of bigram frequency (cf. Massaro, Jastrzembski, & Lucas, 1981), phonology (cf. Van Orden, Johnston, & Hale, 1988), or frequency (cf. Stone & Van Orden, 1993). The generalizability assumption claims a multivalent model will provide richer information about lexical processing than a bivalent model. The elimination of the quantization problem by using a multivalent approach should expand previous research, not fly in the face of it.

Word Recognition

One common methodology in the study of word recognition is the lexical decision task. Rubenstein, Garfield, and Millikan (1970) first used the lexical decision task to test the hypothesis that word recognition involves use of a mental lexicon. As in most subsequent lexical decision experiments, participants were required to respond "yes" if they thought a stimulus was a word and "no" if they believed the stimulus was not a word. Dependent variables were reaction times and percent correct. Data indicated that the longest reaction times for word stimuli were in the low frequency-nonhomographic case. Low frequency-nonhomographic words (e.g., CAFE) were recognized more quickly than nonwords (VERK), but more slowly than homographs or polysemous words (BANK). Other significant variables influencing Rubenstein et al.'s reaction time scores were frequency (low vs. high), homography (many meanings vs. one meaning), and an interaction between concreteness (perceptible through the senses versus abstract) and homography. These findings have been supported throughout the lexical-decision literature (see Gernsbacher, 1984, for a review).

Other variables investigated with the lexical-decision task are neighborhood (the number of words made by changing one letter in a given stimulus) and bigram frequency (the frequency values of letter pairs, e.g., ea or th, within a word). Coltheart, Davelaar, Jonasson, and Besner (1977) defined the neighborhood (N) measure and used it to study neighborhood effects on lexical decision latencies. Coltheart et al. found no neighborhood latency effects for words, but found a difference in response latency between high N and low N nonwords. Andrews (1992) took Coltheart et al.'s study further by investigating N and bigram frequencies. Andrews found N effects were due to lexical similarity and not orthographic redundancy. However, Gernsbacher (1984) has shown that only experiential familiarity (not printed frequency) significantly affected recognition reaction times.

Other studies using the lexical decision task have investigated word superiority effects (Paap, Newsome, McDonald, & Schvaneveldt, 1982). The lexical-decision task has also been used as a means of supporting the continuous debate as to the existence of a mental lexicon (Besner, Twilly, McCann, & Seergobin, 1990; Fera & Besner, 1992; Seidenberg & McClelland, 1989, 1990; Morton, 1969). The latter issue arose out of developments in neural net modeling and pertains to whether local representations of words are required to perform the lexical-decision task, or whether distributed representations of words are sufficient to perform the task.

The lexical decision literature (see Massaro, et al., 1981) suggests word frequency and word regularity play some role in lexical-decision processing. Reguarity is the extent to which English words can be pronounced by analogy to other English words. In other words, regularity is how much a word violates grapheme-to-phoneme correspondence rules. The present generalizability assumption claims that a fuzzy model of lexicality will not contradict previous lexical-decision findings. For that reason, both printed frequency and regularity will be manipulated to determine their roles (if any) in the fuzziness of lexicality (i.e., in wordness ratings). Previous research (Massaro, et al., 1981) suggests that higher frequency or greater regularity will lead a letter string to be rated as more wordlike than stimuli of low frequency or less regularity.

One way to investigate the frequency and regularity variables is to use pairwise comparisons. Given two words matched for phonology and frequency (e.g., homonyms), the string with the greater regularity should be chosen as more wordlike if regularity plays a role in wordness ratings. In the present study, homonyms matched for frequency will be compared to determine if regularity is a component of lexicality. Alternatively, homonyms matched for regularity will be compared to determine if frequency is a component of lexicality ratings.

In the nonword case, bigram frequency is used instead of regularity. Therefore, an illegal nonword (e.g., BNARE) should be selected as more wordlike than an illegal nonword with lower bigram frequency (e.g., TNEZT). In this case, frequency and phonology are not variables and thus would not enter into the decision. Such pairwise comparisons will allow an isolation of variables pertinent to wordness ratings.

Oden (1977) used pairwise comparisons to determine whether individuals would consistently rate categorical exemplars according to a fuzzy model. Participants chose which of two sentences were more representative of a category (A ROBIN IS A BIRD, AN OSTRICH IS A BIRD) and rated the subjective difference between the stimuli. Oden found exemplars of categories fit a fuzzy model. The present study changes Oden's stimuli from sentences to letter strings. This change results in a modification of the typical lexical-decision task. Types of letter strings can then be manipulated to determine the lexicality of those strings. The present study manipulates the letter strings for frequency, regularity or bigram frequency, and phonology. The specific manipulations will be described in the next section. These manipulations will help determine which (if any) variables are components in fuzzy lexicality processing.

Prior research suggests people will consistently rate letter strings according to a fuzzy model. Gernsbacher (1983, as cited in Gernsbacher, 1984) found printed frequency to be a fuzzy, not a crisp, concept. She had participants rate same-frequency words according to familiarity. Her results showed that low-printed-frequency words differed substantially across experiential familiarity. Balota and Chumbley (1984, p.352) suggested

> the basic notion is that words and nonwords differ on a familiarity/meaningfulness (FM) dimension. A particular letter string's value on this FM dimension is based primarily on its orthographic and phonological similarity to actual words. The word and nonword distributions on the FM are separated but overlap.

Gernsbacher's (1983) findings and Balota and Chumbley's suggestion indicate that frequency is a fuzzy variable. If that is the case, frequency should contribute to the fuzziness in lexicality ratings because a decision based on a fuzzy quantity is in itself a fuzzy quantity (Pedrycz, 1991).

For the phonological case, Van Orden, Johnston, and Hale (1988) found that false positive errors to pseudohomophones were more likely than false positive errors to nonhomophonic nonword control foils. For example, false positive errors were found more often in the case "SUTE" as AN ARTICLE OF CLOTHING than in the case "SULE" as AN ARTICLE OF CLOTHING. Van Orden et al. stated that phonological characteristics of nonword foils are critical to word identification. Gernsbacher (1984) also states that pronounceable nonwords are harder to reject as nonwords than nonpronounceable nonwords. Because phonology appears to contribute to nonword lexicality, it will be included in the present experiments. Learning that phonology contributes to wordness ratings, if the hypothesis is confirmed, will provide converging evidence that pronounceable nonwords are more similar to words than are phonologically illegal nonwords. This evidence will also support the idea that all nonword lexicality is not equal.

Finally, the generalizability assumption encourages the inclusion of reaction time data in the present experiments. Pisoni and Tash (1974, as cited in Oden & Massaro, 1978) found decision latencies for a choice that two sounds were the same phoneme were dependent on the phonemes' degree of similarity with respect to voice onset time. Stone and Van Orden (1993) found that correct reaction times to words and the advantages of high-frequency over low-frequency words were greater when the nonword foils were more wordlike. These results were true in the case of illegal versus legal nonwords and legal nonwords versus pseudohomophones. The present experiments include reaction time (surreptitiously measured) as a variable in order to show that these previous findings can be included in a multivalent model of lexicality.

To recapitulate the concepts introduced so far, (a) fuzzy logic claims that most events are a matter of degree, (b) the FLMP predicts that individuals make decisions based on the degrees to which various stimulus components contribute to the overall psychological experience of the stimulus, c) decisions based on fuzzy quantities are fuzzy decisions, and d) the word recognition literature suggests that phonology, orthography, frequency, and neighborhood each contribute to lexical decisions.

<u>A model of computed lexicality</u>

The main idea of the present discussion is that prior word recognition studies have overlooked information due to the quantization problem, or rounding off the excuded middle. The proliferation of bivalent models of lexicality may be due to the idea of a lexicon, or mental dictionary. Lexicon-based models assume people have a separate node for each word they have learned. These models constrict the concept of lexicality to the bivalent case; either a letter string is in the lexicon (is a word) or it is not in the lexicon (is a nonword).

Seidenberg and McClelland (1989) developed a parallel distributed processing (PDP) model to show that a lexicon is not necessary for word recognition or pronounciation. The PDP model involved the training of a neural network on a corpus of almost 3000 words and their pronounciations. Seidenberg and McClelland found that their trained model could perform the lexical decision task fairly well without the use of a lexicon. Thus, the door has been opened to the use of non-lexicon-based models.

However, neural networks are not the only answer. Many areas in science and engineering are now combining neural networks with fuzzy systems to enhance the capabilities and accuracies of neural networks alone (Kosko, 1993). The time may have come for the Seidenberg and McClelland (1989) model to be combined with a fuzzy system. If the expected results obtain, this combination of neural network and fuzzy system may be the next step in computational models of lexicality.

<u>Variables used in the present study</u>

Previous word recognition research as mentioned above has indicated that some of the important variables in lexical-decision tasks are word frequency, regularity, orthograpy, and phonology. Each of these variables will be manipulated to determine what role (if any) it plays in the fuzziness of wordness decisions. The manipulations fall within two broad categories, various types of word and and various types of nonword letter strings. Words will not be compared with nonwords. However, several comparisons will be made within each category.

In the nonword category, pseudohomophones (BRANE, PHLAG), legal nonwords (SOOK, BLORE), and illegal nonwords (BSETE, WRATL) will be used to investigate the orthographic and phonological variables. Each letter string will be compared to the other letter strings in the nonword category. The results are expected to show that relative truthfulness ratings of the comparisons (e.g., BRANE IS A WORD, SOOK IS A WORD) will follow a pattern consistent with the fuzzy model. Results are expected to show that 1) pseudohomophones will be selected as more wordlike than legal and illegal nonwords, 2) legal nonwords will be selected as more wordlike than illegal nonwords, and 3) when two letter strings of the same type are presented (e.g., two different illegal nonwords) the string with the larger bigram frequency will be selected as more wordlike.

In the word category, homonyms will be matched for frequency or regularity. The use of homonyms (e.g., BLEW, BLUE) eliminates phonology as a component in the lexicality decision. Thus, the higher frequency or more regular homonym should be selected as more wordlike than the lower frequency or less regular alternative. Irregular words (HAVE, AISLE) will be matched for frequency, with the assumption that higher frequency irregular words will be chosen as more wordlike than lower frequency irregular words. In the irregular word case, orthography is held at a low level to compare the stimuli for frequency effects. Similarly, irregular words will be compared with regular words matched for frequency. The data should show that regular words are considered more wordlike than irregular words of the same frequency. In this case, orthography is manipulated between the regular and irregular words.

A visualization of one possible continuum of lexicality ratings is shown in Figure 2. Illegal nonwords should most likely be considered least wordlike of the stimuli. Similarly, the word stimuli should all be rated as fairly wordlike, but the actual order of frequency, regularity, and phonology will be determined by the data. The right side of Figure 2 indicates how traditional bivalent models group lexicality. The figure suggests how some of the richness of the data can be lost in the bivalent case.

Hypotheses

The current hypotheses, based on previous research findings and the present discussion are listed below. Figure 3 illustrates in tabular form the hypotheses and the stimuli expected to be favored for each one.

For the nonword category:

1) Pseudohomphones will be selected as more wordlike more often than illegal nonwords.

2) Pseudohomophones will be selected as more wordlike more often than legal nonwords.

3) Legal nonwords will be selected as more wordlike more often than illegal nonwords.

4) When legal nonwords are compared with other unique legal nonwords, the stimulus with the larger neighborhood will be selected as more wordlike more often.

5) When illegal nonwords are compared with other unique illegal nonwords, the stimulus with the larger neighborhood will be selected as more wordlike more often.

For the word category:

6) Given a homonym pair matched for frequency, the word with greater regularity will be chosen as more wordlike more often than the less regular word.

7) Given a homonym pair matched for spelling, the higher frequency word will be chosen as more wordlike more often than the lower frequency word.

8) High frequency irregular words will be chosen as more wordlike than low frequency irregular words.

9) When matched for frequency, regular words will be chosen as more wordlike than irregular words.

10) Response latencies should be longer for stimuli which are closer together on the lexicality continuum than stimuli which are farther apart (e.g., the response latency for the lexicality decision between a pseudohomophone and an illegal nonword should

be much shorter than the response latency to decide the greater lexicality of two unique illegal nonwords).

Chapter Two: Method

Design. The current investigation used a within-subjects design. All participants evaluated all of the letter strings. The reason for this approach is that frequency is an internal dimension and the comparison of subjects across groups may produce invalid results. Archer and Wang (1991) have shown that grade of membership is subjective and context-dependent, therefore, stimulus ambiguity will vary differently for different individuals. Similarly, regional dialects may play a role in the extent to which some stimuli are considered pseudohomophones. As a result, participants were exposed to all the manipulations of orthography, phonology, and frequency.

Subjects. Thirty Oklahoma State University undergraduate volunteers (nine male, twenty-one female, average age 22.8 years) participated in the experiment. All were native speakers of English with normal to corrected-to-normal vision and received a small amount of course credit for their participation.

Apparatus. The stimuli were presented on an Apple II computer with a monochromatic screen. Stimulus presentation and scoring programs were written in BASIC.

Stimuli. Nonword stimuli were taken from Stone and Van Orden (1993) and Andrews (1989). Homonym pairs can be found in McRae, Jared, and Seidenberg (1990). A list of all the stimuli is included in Appendix A. Average bigram frequency for the illegal nonword case was one for the low condition and 33 for the high condition. For the legal nonword case, the average bigram frequency was 90.5 for the low condition and 191 for the high condition. Pseudohomophone bigram freqencies were 76.5 for the low condiiton and 178 for the high condition. Bigram frequencies were computed from tables in Mayzner and Tressalt (1965). Values of average counterpart word frequencies for the pseudohomophones (Kucera & Francis, 1967) were 39.67 for the high condition and 2 for the low condition.

The mean word frequencies (taken from Kucera & Francis, 1967, with frequencies listed as per million) of the stimuli used were as follows. For the matched frequency homonym case, the mean frequencies were 40.6 for the regular words and 60.5 for the irregular words. For the matched spelling homonym case, the high frequency condition mean was 181.0, while the low frequency condition mean was 4.5. For the irregular low frequency case, the mean frequency was 3.1, while the regular low frequency case mean frequency was 5.4. For the irregular high frequency case, the mean frequency was 167.6, while the irregular low frequency case mean was 7.8.

Procedure. Participants were presented with two sentences (e.g., BRANE IS A WORD, PHLAG IS A WORD). One sentence was presented above the other sentence, separated by a blank line of type. Individuals were instructed to enter their choice of which sentence had a greater truth value and then enter a rating of how much more truthful the selected stimulus was as compared to the remaining stimulus. Ratings ranged from "9" (very much more truthful) to "1" (about the same amount of truthfulness). Each trial was surreptitiously timed by a software clock (Price, 1979) from the moment the stimulus appeared on the screen until the participant responded by pressing the appropriate response key ("1" for the top sentence, "2" for the bottom sentence).

Participants were asked to compare all nonword stimuli with all other nonword stimuli. Each nonword letter string was compared to all the other nonword letter strings. As a result, 190 nonword comparisons were made, 45 for the illegal nonword/illegal nonword comparison, 10 each for the legal nonword/legal nonword and pseudohomophone/pseudohomophone comparisons, 50 for the legal nonword/illegal nonword comparison, 50 for the illegal nonword/pseudohomophone comparison, and 25 for the legal nonword/pseudohomophone comparison. Ten unique illegal nonwords and ten unique legal nonwords (five legal, five pseudohomophones) comprised the nonword stimulus set. Ten frequency-matched homonym pairs, ten spelling-matched pairs, ten high-frequency regular words, ten high-frequency irregular words, and five high-frequency and five low-frequency irregular words comprised the word stimulus set.

Participants were only asked to compare words of similar categories (e.g., irregular/regular words, matched homonym pairs). In all, 145 word stimuli comparisons were made. Thus, participants compared 335 stimulus pairs. Each stimulus pair presentation was initiated by the participant (e.g., a self-paced schedule was used).

The order of stimulus presentation was randomized in several ways. The stimuli were randomized within category (e.g., all the illegal nonwords were randomized amongst themselves), within presentation pair (whether a particular stimulus was presented first or second in a given comparison), and across the experiment (e.g., the order of presentation of the various types of comparisons). The randomization reduced the amount of confounding that might have occurred due to order effects or practice effects.

Three dependent variables were measured. The first variable was reaction time between stimulus presentation and selection of the sentence rated as more truthful. The second variable was the choice of letter string in each trial selected as more wordlike. The third variable was the rating of how much more truthful the selected sentence was as compared to the first sentence.

Chapter Three: Results

Nonword comparisons

Significant differences were found for most of the hypotheses listed in Chapter 1. The mean number of pseudohomophones chosen over illegal nonwords (hypothesis 1) was 46.67 (out of 50 comparisons), a value which differs significantly from a chance value of 25, $t(29) = 33.84$, $p < .00001$. Similarly, the mean number of times a legal nonword was chosen over an illegal nonword (hypothesis 3) was 47.4, again a reliable difference from the chance value of 25, $t(29) = 45.25$, $p < .00001$. Of a total of 25 comparisons, the mean number of pseudohomophones chosen over legal nonwords (hypothesis 2) was 16.27, a value also significantly different from the chance value of 12.5, $t(29) = 6.06$, $p < .00001$.

For the cases in which like stimuli were compared (e.g. an illegal word with an illegal word, MRAUB/LREOT), bigram frequency counts were used to separate the letter strings into two groups. In the case of illegal nonwords, the letter strings with higher bigram frequency were chosen significantly more often than the strings with lower bigram frequency. The mean number of illegal strings with higher bigram frequency chosen over those with lower bigram frequency (hypothesis 5) was 15.33 (out of 25 comparisons). This value differs significantly from the chance value of 12.5, $t(29) = 5.50$, $p < .00001$.

When legal nonword strings were compared with each other (hypothesis 4) no significant differences were found based on bigram frequency. The mean number of higher bigram frequency strings chosen (out of 6 comparisons) was 3.1, compared to the chance value of 3, a non-significant difference, $t(29) = 0.45$, $p > .05$.

The instance of comparing pseudohomophones with each other suggested two ways to analyze the data. One was to look at the bigram frequencies, in order to be consistent with the measures of the other two (illegal and legal) nonword cases. Because these strings are pseudohomophones, however, they were also compared according to the frequency of the words they sound like. For example, the pseudohomophone SHURT would be assigned the frequency of the real word SHIRT. In the case of bigram frequency, no significant difference was found between high and low bigram frequency pseudohomophones. The mean number of times a higher bigram frequency pseudohomophone was selected over the lower frequency pseudohomphone (out of 6 comparisons) was 2.63, a value not reliably different than the chance value of 3, $\underline{t}$ (29) = 1.59, $\underline{p} > .05$.

A different finding is apparent when the pseudohomphones are compared using the frequencies of their real-word counterparts. The mean number of times a string with a higher frequency counterpart was chosen over the lower frequency counterpart (out of 6 comparisons) was 3.53, a significant difference from the chance value of 3, $\underline{t}$ (29) = 2.39, $\underline{p} < .025$.

Word comparisons

For the case of homonyms (BETTER/BETTOR) matched for spelling with different frequencies (hypothesis 7), the word of higher frequency was chosen as more wordlike significantly more often than the word of lower frequency. The mean number of times the higher frequency word was chosen over the lower frequency word (out of 10 comparisons) was 7.43, a significant difference from the chance value of 5, $\underline{t}$ (29) = 8.38, $\underline{p} < .00001$.

Homonyms with different regularity and similar frequencies (e.g., SITES/ SIGHTS, hypothesis 6) also produced frequency effects. Words with higher frequency were chosen more often than those of lower frequency. The mean number of higher frequency words chosen (out of 10 comparisons) was 6.37, again a reliable difference from the chance value of 5, $\underline{t}$ (29) = 5.44, $\underline{p} < .00001$. However, no reliable regularity effect was found because the average number of higher regularity words chosen was 5.5 (out of 10 comparisons), which is not statistically different from the chance value of 5, $\underline{t}$ (29) = 1.82, $\underline{p} > .05$.

For words of similar frequency but different regularity (hypothesis 9), regular words (e.g., CLOWN) were chosen as more wordlike more often than irregular words (e.g., RESIGN). The mean number of times a regular word was chosen as more wordlike (out of 100 comparisons) was 68.77, a reliable difference from the chance value of 50, t (29) = 9.03, $p < .00001$.

In the case of matched irregularity and different frequency (e.g., CAFE/BUILD, hypothesis 8), the higher frequency words were chosen as more wordlike more often than the lower frequency words. The mean number of times the higher frequency irregular word was chosen (from 25 comparisons) was 16.60, a significant difference from the chance value of 12.5, t (29) = 6.59, $p < .00001$.

Reaction times

Mean reaction times for each comparison type are listed in Table 1. Table 2 describes which of the reaction time differences are significant. A graph of all the reaction times is shown in Figure 4.

Chapter 4: Discussion

The following discussion consists of two parts. The fist part will describe how the present data show that participants used graded judgments of wordness. The second part will describe a tentative process model of those wordness ratings. The second part will be divided into explanations of reaction time data and explanations of preference data.

Evidence against a categorical model

To review, a categorical model implies that judgments are made in a simple yes or no fashion. The categorical approach to lexicality suggests that people rate letter strings as "words" or "nonwords." Given a large number of letter strings, people (according to the categorical view) would divide the strings into two piles only, a word pile and a nonword pile. The lexical decision task forces individuals to use this dichotomous approach. In the lexical decision task, participants are requested to state whether a letter string is a word or not. While this approach has provided a wealth of information about the processes used in the perception of written language, it leaves untouched the question of whether all words (or nonwords) are created equal with respect to lexical status.

On the other hand, a continuous model allows graded judgments. Responses of "sort of," "mostly," and "not really" are allowed in addition to "yes" or "no." The continuous approach to lexicalty implies people rate letter strings according to a continuum. Given the same large number of letter strings suggested in the preceding paragraph, any number of wordlike piles may be used, according to how finely a person wants to discriminate the letter strings.

In a lexicality rating task such as the one used in the present experiment, participants chose which letter string was more wordlike and indicated the degree of difference, thus allowing a continuous rating scale. If participants were actually using a categorical model of lexicality, the results of a rating task given comparisons of the same type (e.g., two words or two nonwords) should produce chance results. Thus, the rating task should show whether or not all words (or nonwords) are given equal responses. The purpose of the present investigation was to answer the question of discrete vs. graded responses and perhaps to go on from there as to the factors involved in the rating process.

The present results support a continuous model of lexicality. For both the word and nonword cases, significant differences were found in the choices made by participants. In every case in which the participants compared dissimilar nonwords or dissimilar words, the individuals significantly chose letter strings with higher orthographic or phonological consistency or higher frequency. For example, legal nonwords were chosen as more wordlike significantly more often than illegal nonwords. This legal over illegal preference was almost totally exclusive (95%). Similarly, for the case of pseudohomophones compared with illegal nonwords, the pseudohomophones were also almost exclusively chosen over the illegals (93%). Also, the comparison of pseudohomophones and legal nonwords resulted in pseudohomophones chosen significantly more often than the legal nonwords (65%). These results by themselves provide strong evidence against a categorical model.

However, the word data present strong evidence against a categorical model as well. For homonym pairs (either same-spelling, e.g., BETTER/BETTOR, or different spelling, e.g., FLEX/FLECKS), the word with the higher frequency was chosen significantly more often. For irregular spelling/regular spelling nonhomonym pairs of similar frequency (e.g., YACHT/SPOON), the regular word (SPOON) was chosen significantly more often than the irregular word (YACHT). Finally, in the case of two irregular words (e.g., ANSWER/ACHE) of differing frequencies, a reliable frequency effect was obtained. Therefore, the present word data show that all words are not created equal.

To summarize, a categorical model of lexicality predicts that, given two letter strings of the same class (word or nonword), people will show no significant preferences for one type of letter string (e.g., pseudohomophone or regular) over another. A continuous model of lexicality predicts signficantly different preferences of letter string type. The data support the continuous model in both word string and nonword string cases.

Development of a lexicality model

As mentioned in Chapter 1, the model developed by Seidenberg and McClelland (1989) could be combined with a fuzzy logic model to enhance the neural network only approach. To review, the Seidenberg and McClelland model used parallal distibuted processing (PDP) to show that a mental dictionary, or lexicon, is not necessary to process letter strings. The network was trained on 2897 words and pronounciations and then tested on new words. The PDP model could successfully pronounce the words and perform the lexical decision task with about 85% accuracy.

It should be noted that the model presented here has been developed post hoc based on the data obtained. Therefore, all instances of the model's predictions are the model's explanation of the present data. However, the model can be used to predict outcomes of future work, and can be used as a starting point for a more definitive model of fuzziness in lexicality ratings.

The PDP model provides two types of error scores, a phonological error score (PES) and an orthographic error score (OES). The PES is the sum of the squared differences between the target activation value for each of the 460 phonological units and the actual activation computed by the network. The OES is the sum of squares of the differences between a) the values computed by the network's hidden units and fed back to the orthographic units and b) the actual inputs given to the orthographic units.

As just outlined, the PDP computes a phonological score and an orthographic score. The frequency of an input can be found in the connection weights within the network. As such, the PDP is an excellent candidate to begin an outline of a fuzzy model of lexicality. The remainder of the present discussion will describe a way to modify the Seidenberg and McClelland (1989) PDP model to incorporate the present findings, with additions from components of Massaro's (1988a) FLMP .

A fuzzy computational model of lexicality (FCML)

Strictly speaking, the Seidenberg and McClelland (1989) model is based on a categorization task in that it has been used to simulate the lexical decision task, itself a categorical method by definition. In Seidenberg and McClelland's version of the lexical decision task, the orthographic error score (OES) is set to a certain criterion. If the OES is greater than the criterion, the string is labelled a nonword. If the OES is less than the criterion, the string is labelled a word. If the present data indicated only the orthographic component contributed to lexicality ratings, converting Seidenberg and McClelland's model to a fuzzy model would be relatively easy. The model would compute the OES for the first letter string, compute the OES for the second letter string, compare the two, and select the string with the lower OES as more wordlike.

The advantage of the FCML over a lexicon-based model is apparent in that the type of input string is irrelevant. The model can accept two words, a word and a nonword, or two nonwords and can generate a sensible response to any of these combinations. This input stage flexibility is one of the reasons the Seidenberg and McClelland model was selected as a springboard for a fuzzy lexicality model. However, the present data indicate three factors are contributing to the lexicality rating decision. Thus, the Seidenberg and McClelland model would require more ability than that involving the comparison of the OES for two letter strings. The remainder of the discussion describes what changes could be made to the Seidenberg and McClelland model to incorporate the present findings. This new model will be referred to as a Fuzzy Computational Model of Lexicality (FCML).

The FCML computes three decision components at the same time. Thus, an orthographic error score (OES), a phonological error score (PES), and a frequency/familiarity rating are computed in parallel. Note that the phonological error score would be a rating of the pronouncability of a letter string, e.g., how unambiguously an individual can find a way to pronounce that string. Seidenberg and McClelland (1989) produced OES and PES in their model. They stated that the model did show frequency effects, and that these effects were based on the connection weights produced during the training phase. Frequencies are introduced into the model at the training stage as a pobability. This probability is based on the log of a word's frequency as found in Kucera and Francis (1967). These familiarity/frequency values will be used in the FCML.

The FCML and decision times

Given the present task, the FCML would compute the decision components and compare them. The string with the smallest error scores and/or highest familiarity rating would be chosen as more wordlike than the other letter string. The three comparisons may get equal weight or, if future research demands it, the model could be modified to allocate different weights to the three components, orthography, phonology, and familiarity. The factor (OES, PES, or frequency) with the largest difference given the two letter strings would determine the choice of letter string. Large intrafactor differences (in one, two, or all three factors) between the letter strings would result in rapid decision times, while small intrafactor differences produce longer decision times. For example, given an illegal nonword (MRAUB) and a pseudohomophone (SKORE), a large OES is produced for MRAUB, but a small OES is produced for SKORE. Similar results would be found for the PES, while the familiarity ratings for both letter strings should be low, but possibly a little higher for SKORE. The FCML predicts a fast reaction time for the illegal nonword/pseudohomophone comparison because two factors have large intrafactor differences, the OES and the PES. Also, the error scores' differences would be big enough to produce highly consistent choices of the pseudohomophone over the illegal nonword. The present data support both the fast reaction time prediction and the consistency prediction. Up to this point, however, the FCML has not provided any information not available from a categorical model.

The comparison of legal nonwords provides the evidence needed to demonstrate the superiority of the FCML over a categorical approach. The FCML predicts slower decision times for stimuli which are more similar on all three dimensions of comparison. For example, a legal nonword (SLURT) and a pseudohomophone (SKORE) would produce more similar OES and PES values than the illegal nonword/pseudohomophone case, but the two still have similar familiarity ratings because both stimuli are nonwords.

Because the error scores are more similar in the legal nonword/ pseudohomophone case than in the illegal nonword/pseudohomophone case, the model predicts that decision times for the legal nonword/pseudohomophone case should increase and the consistency of choices should decrease relative to the illegal nonword/pseudohomophone case. The FCML predicts that subjects will consistently choose the pseudohomophone over the legal nonword. The categorical model predicts chance consistency because both strings are legal nonwords and neither excite any lexical entries enough to make consistent discriminative decisions. The data support the predictions of the model, in that decision times did increase over the illegal nonword/pseudohomophone case, and that consistency also declined but was still statistically different than chance in the legal nonword/pseudohomophone comparison.

Finally, the model predicts that cases with no discernable intrafactor differences will require the longest decision time and the lowest consistency in choice of preferred letter string. The data again support both predictions. For nonwords, a comparison of two legal nonwords (e.g., TREST and SLURT) produced the longest decision times and least consistent selection of individual items. According to the model, the OES and the

PES are not different enough to produce a consistent choice of which letter string is more wordlike, and the ambiguity results in longer decision times. Again, because these are nonword strings, the familiarity ratings are very low. Therefore, the nonword data support the FCML, while a categorical model is not supported by the same nonword data.

For the comparisons involving actual words, the model still holds. For the case of same-spelling variable-frequency homonyms (e.g., BETTER/BETTOR), the model predicts rapid decision times if the difference in frequency is very large. The decision times in this case can only depend on frequency because phonology and orthography are held constant. Therefore, the decision times should depend on the difference in frequency. The data support the model. The difference in mean frequencies was 176.5, a fairly large difference. The mean reaction time was in the midrange of all of the reaction times measured. A future study could vary the difference in frequency of the same-spelling homonym strings to determine if there is an increase in decision time as the frequency difference decreases.

The pairs containing irregular words of differing frequency provide still more support for the FCML. In this case, the orthographic regularity is considered low for both groups of words, but the frequencies differ (difference of the mean frequencies is 159.8). Thus, the OES values should be about equal, but the familiarity ratings should be very different. It is difficult to determine at this point how the PES values are computed in this case, but the difference in phonology should decrease the decision times compared to the same-spelling homonym case above. An added difference in one of the factors (in this case a phonology difference is added to the frequency difference) should decrease the decision times. The data support the FCML in that a frequency effect was obtained and the decision times decreased as a result of the differing phonologies.

For the case of same frequency words with differing regularity, the model predicts a regularity effect. Recall that irregular words (e.g., CAFE) cannot be pronounced by analogy to other English words, that is, they violate grapheme-to-phoneme correspondence rules. From the nonword data, large orthographic differences resulted in the shortest decision times. Thus, the model predicts fairly rapid decision times for a large difference in nonword orthography. The model therefore also predicts shorter decision times for more regular words (more consistent orthography). Again, it is difficult at this point to determine the exact contribution of the differences in phonology. But, that a difference exists surely contributes to the decision times. The data support the FCML regularity prediction here as well, with decision times for more regular words less than those times based on items with similar frequency but low regularity.

Finally, the similar frequency homonym comparisons support the FCML as well. In this case, both strings have identical phonology, low OES, and similar familiarity. Thus, their decision times should be very long. The data show the longest decision times in the case of similar frequency homonyms. The model predicts that one small difference and two indiscernable differences produce longer decision times. The data in the similar frequency homonym condition support that prediction because two of the differences are very small (phonology, orthographic regularity) and one is only slightly larger (frequency).

To summarize, the model predicts longer decision times for stimuli with similar factor scores, suggesting an inversely proportional relationship for decision time and difference in error scores or familiarity ratings. Thus, one or more very large error score differences will produce rapid decision times. A few smaller differences would produce longer decision times. No discernable differences or highly ambiguous stimuli on all factors would result in the longest decision times. The present data have shown consistent support for the decision time explanations of the FCML.

The FCML and lexicality decisions

The FCML also explains lexicality decisions which are reflected in the present data. The lexicality decisions also depend on the differences in the error scores or familiarity ratings. Larger differences should produce more consistency (choosing the same type of letter string consistently) in the choice of greater lexicality. Smaller differences should produce less consistency. The data reflect the choice model as follows. In all circumstances, the conditions that lead to slower decision times can be expected to create less consistent classifications of wordness.

For the cases involving nonwords, an illegal nonword paired with any legal nonword string (including pseudohomophones) will result in a large difference in OES and thus a high consistency in the choice of the legal nonword string. The data support this prediction, showing the highest consistency for the illegal nonword/legal nonword case (95%). The illegal nonword/pseudohomophone case also very high (93%) and is not significantly different from the illegal nonword/legal nonword case in consistency. In the comparison of legal nonwords and pseudohomophones, the pseudohomophones should be chosen consistently over the illegal nonwords, but with a lower consistency because the orthography is the same. The consistency with which participants chose pseudohomophones over nonwords is based on the phonology differences of the two strings, which appears to have a lesser contribution than orthography. The data support this view, as the consistency of pseudohomophones chosen over legal nonwords dropped to 65%, a number more in line with the word findings.

The FCML also predicts lower consistencies for the nonword cases in which the stimuli are highly similar. The pseudohomophone/pseudohomophone comparison showed a lower consistency of choice (59%), which appeared to be based primarily on counterpart word frequency. The FCML predicts that the higher counterpart frequency pseudohomophone will be chosen more often than the pseudohomophone with a lower counterpart frequency. This prediction is supported by the data in that the mean number of times the higher counterpart frequency string was chosen was 3.53, significantly different than the chance value of 3.0.

A higher consistency may be found with a larger difference in counterpart frequencies. The mean frequency difference for the present stimuli was only 37. A future study could vary the counterpart frequencies of the pseudohomophones to determine if higher consistencies are found with larger counterpart frequency differences. The FCML predicts a direct relationship between lexicality choice consistency and pseudohomophone counterpart frequency differences. The FCML predicts non-significant differences for the cases of legal nonwords compared with other legal nonwords, and illegal nonwords with other illegal nonwords. The data support this prediction, which contradicts the original hypothesis that the choice would be based on bigram frequency.

For the comparisons involving word stimuli, the data also support the FCML. In the comparison of same-spelling homonyms, the present model predicts the consistency of the choice is dependent on the difference in the frequencies. The consistency found for the present data was the highest for the word comparisons (74%). Future work could, again, systematically vary the difference in frequency to determine the relationship between the frequency difference and the consistency of choices. The model predicts that as the frequency difference increases, so will the consistency of the lexicality choice.

For differences in regularity, the model predicts regular words to be chosen over irregular words with a consistency less than that of the same-spelling homonyms. The data show a reliable consistency (69%), but that may be due to the fact that regularity is not as concrete a factor as frequency. Further work could vary the regularity according to rules similar to those of Massaro, Venezky, and Taylor (1979). The model predicts consistency will increase as the difference in regularity increases. Another problem with regularity is that it is, at this juncture, hard to determine how much of regularity effects are due to orthography and how much are due to phonology. The present study assumes that the majority of the difference is due to orthography, but further work needs to be done to confirm or refute this assumption.

When frequency differs for low regularity words, the model predicts that the higher frequency words will be chosen as more wordlike, but again with lower consistency than the case of the same-spelling homonyms. This prediction is based on the contribution of orthographic regularity, familiarity, and the relationship between the two. Lower regularity words' familiarity may vary across individuals according to familiarity more widely than is characterized by the words' frequencies. As Gernsbacher (1984) pointed out, words of similar frequency can vary widely according to familiarity. Therefore, the familiarity of low regularity words may vary widely within the same frequency range and thus lower the consistency of which word is chosen as more wordlike. Additional work comparing low regularity words with differing familiarity (as opposed to printed frequency) may provide clarification to this problem.

Finally, in the case of word comparisons, the model predicts that same-frequency homonyms of differing regularity should show a regularity effect. The present data suggest a regularity trend in similar frequency homonyms, but the effect is not statistically significant. The data also indicated a frequency effect, even though the difference of the frequency means was only 20. The FCML explains these findings in that the only intrafactor difference big enough to show a reliable consistency difference is that of frequency. A choice of variable-regularity homonyms varying more in regularity values with closer frequency (or familiarity) values may determine if the model is describing the data correctly.

Massaro's FLMP (1988a) suggests that the auditory and visual factors are independent in the evaluation stage (see Figure 1). This independence implies that when a feature is assigned to one source of information it remains independent of values assigned to other sources of information. This type of independence does not imply statistical independence, as operations that occur in other parts of the model (e.g., integration) can introduce statistical dependence. The fuzzy computational model of lexicality (FCML) may be operating under the same type of independence, but it is beyond the scope of the present investigation to determine if that is the case. Figure 5 presents the FCML in a similar fashion to the FLMP (Figure 1). The orthographic, phonological, and familiarity components are all computed in the evaluation stage according to algorithms similar to those described by Seidenberg and McClelland (1989). The psychological values of these factors (given by the lower case letters in the figure and represented by the OES, PES, and familiarity rating, respectively, for the orthographic, phonological and familiarity components) are presented to the integration stage. The integration may be a minimization function, a multiplicative function, or some other combinatorial function of the three factors. The psychological value of the integration stage is output to the classification stage, where it is converted to a response. The classification stage can be bivalent (as in the lexical decision task), a rating of some kind, or selection of an item as more wordlike (as in the present task).

Figure 6 presents a timeline for the FCML. The time required to compute the OES, PES, and familiarity values for the two letter strings should be approximately the same, regardless of the string type (word or nonword). Also, the response time in the present task to press the appropriate computer key for the top or bottom letter string should be consistent across trials. Therefore, the actual comparison stage should determine the variance in decision times for the different types of comparisons. Large intrafactor differences lead to shorter comparison times because the differences in the OES, PES, and familiarity differences are more obvious. Similarly, small intrafactor differences will lead to longer comparison times because the OES, PES, and familiarity differences are more ambiguous.

As stated at the beginning of this chapter, the present discussion is not intended to prove the existence of fuzzy logic and lexicality decisions, but to act as a springboard for future research into a multifactor approach to how people interpret fuzzy quantities such as letter strings. The FCML is merely a first step in trying to integrate the present results.

The next step in the present line of investigation is to incorporate the present suggestions into the Seidenberg and McClelland (1989) model to compare two letter strings and see how the FCML data compare with human data, much like Seidenberg and McClelland did with lexical decision data and their parallel distributed processing model. Also, work could be done to further develop Massaro's FLMP to incorporate specific degrees of wordness as he has done with degrees of speech perception. Finally, more extensive replications of the present findings could be done for each of the comparisons herein. This is especially true with the homonyn comparisons.

Methodological contributions

As stated earlier, the lexical decision task has led to a wealth of information concerning how individuals process written language. However, the lexical decision task, by design, forces a bivalent categorization of the input stimuli into two categories: words and nonwords. The lexical decision task has also generated wide controversy due to its susceptibility to strategy factors (Seidenberg, Walters, Sanders & Langer, 1984). The present investigation avoids both shortcomings of the lexical decision task by eliminating the categorical bias of the lexical decision task.

The purpose of using the present method was to determine if lexicality did, in fact, follow a continuum. The present results have shown that lexicality does have varying degrees. As a result, a whole new avenue of written language research has opened up. The present discussion suggests several ways to improve upon the current findings, but only by choosing stimuli according to stricter guidelines. The actual method has been shown to produce reliable results, results that indicate individuals process written language in a fuzzy, not a categorical, manner.

REFERENCES

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? Journal of Experimental Psychology: Learning, Memory & Cognition, 18, 234-254.

Archer, N. P., & Wang, S. (1991). Fuzzy set representation of neural network classification boundaries. IEEE transactions on systems, man, and cybernetics, 21, 735-742.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance, 10, 340-357.

Besner, D. (1987). Phonology, lexical access in reading, and articulatory suppression: A critical review. The Quarterly Review of Experimental Psychology, 39A, 467-478.

Besner, D., Twilly, L., McCann, R. S., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? Psychological Review, 97, 432-446.

Chandler, J. P. (1969). Subroutine STEPIT - finds local minima of a smooth function of several parameters. Behavioral Science, 14, 81-82.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), Attention and performance VI (pp. 535-555). Hillsdale, NJ: Erlbaum.

Fera, P., & Besner, D. (1992). The process of lexical decision: More words about a parallel distributed processing model. Journal of Experimental Psychology: Learning, Memory, & Cognition, 18, 749-764.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. Journal of Experimental Psychology: General, 113, 256-281.

Kosko, B. (1993). Fuzzy thinking: The new science of fuzzy logic. New York, NY. Hyperion.

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

Massaro, D. W. (1988a). Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1988b). Ambiguity in perception and experimentation. Journal of Experimental Psychology: General, 117, 417-421.

Massaro, D. W. (1994). A pattern recognition account of decision making. Memory & Cognition, 22, 616-627.

Massaro, D. W., & Cohen, M. M. (1993). The paradigm and the fuzzy logical model of perception are alive and well. Journal of Experimental Psychology: General, 122, 115-124.

Massaro, D. W., Jastrzembski, J. E., & Lucas, P. A. (1981). Frequency, orthographic regularity and lexical status in letter and word perception. In G. H. Bower (Ed.) The Psychology of Learning and Motivation, 15, 163-200.

Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and digram frequency counts for various word-length and letter position combinations. Psychonomic Society Monograph Supplements, 1, 13-32.

McRae, K., Jared, D., & Seidenberg, M.S. (1990). On the roles of frequency and lexical access in word naming. Journal of Memory & Language, 29, 43-65.

Morton, J. (1969). Interaction of information in word recognition. Psychological Review, 76, 165-178.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. Psychological Review, 89, 573-594.

Pedrycz, W. (1991). A referential scheme of fuzzy decision making and its neural network structure. IEEE transactions on systems, man, and cybernetics, 21, 1593-1604.

Price, J. M. (1979) Software timing for 6500 series microcomputers. Behavior Research Methods & Instrumentation, 11, 568-571.

Oden, G. C. (1977). Fuzziness in semantic memory: Choosing exemplars of subjective categories. Memory & Cognition, 5, 198-204.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. Psychological Review, 85, 172-191.

Seidenberg, M. S. (1990). Lexical access: Another theoretical soupstone? In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds). Comprehension Processes in Reading (pp. 33-71). Hillsdale, NJ: Erlbaum.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96, 523-568.

Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon: A reply to Besner, et al. (1990). Psychological Review, 97, 447-452.

Seidenberg, M. S., Waters, G., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. Memory & Cognition, 12, 315-328.

Stellman, R. (1994). Fuzzy logic in the lexical decision task: Lexicality as a matter of degree. Paper presented to the Southwestern Psychological Association, April, 1994, Tulsa, Oklahoma.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 19, 744-774.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. Journal of Experimental

Psychology: Learning, Memory, & Cognition, 14, 371-386.

Appendix A

Stimulus lists

Illegal Nonwords

MRUAB	TLSEI
RNEGA	LREOT
LKOGI	PQAUK
TNIOF	RBEJU
RSEUV	RHUTD

Legal Nonwords

SNEEK	SNERK
TREET	TREST
SHURT	SLURT
SKORE	SKIRE
SWERL	SWARL

Matched Frequency

CANNON (7)	CANON (5)
AISLE (6)	ISLE (5)
WOE (5)	WHOA (1)
BERRY (9)	BURY (6)
SITES (16)	SIGHTS (15)
PAST (281)	PASSED (157)
FLEX (2)	FLECKS (1)
RED (197)	READ (173)
WRAP (5)	RAP (2)
RIGHTS (77)	RITES (41)

Matched Spelling

BETTER (414)	BETTOR (1)
CURRENT (104)	CURRANT (1)
COURSE (465)	COARSE (10)
FEET (283)	FEAT (6)
DEAR (54)	DEER (13)
LIE (59)	LYE (1)
REAL (260)	REEL (2)
BIRTH (66)	BERTH (4)
SLOW (60)	SLOE (2)
STEEL (45)	STEAL (5)

Table 1 - Mean Reaction Times

Nonword Comparisons		Word Comparisons	
Stimuli	RT (ms)	Stimuli	RT (ms)
legal/legal	3446	Homo/matched F	3906
legal/ph	3126	Homo/matched spell	2989
ph/ph	3013	Irreg - vary F	2513
illegal/illegal	2952	Low F - vary reg.	2316
illegal/legal	2149		
illegal/ph	2095		

ph = pseudohomophone

homo = homonym
irreg = irregular
F = printed word frequency

Table 2 - Significant differences between reaction times

Comparison 1	RT	Comparison 2	RT	*p* value
legal/legal	3906	legal/ph	3126	> .05
legal/legal	3906	ph/ph	3013	< .01
legal/legal	3906	illegal/illegal	2952	< .025
illegal/illegal	2952	illegal/legal	3906	< .0001
illegal/legal	2149	illegal/ph	2095	> .05

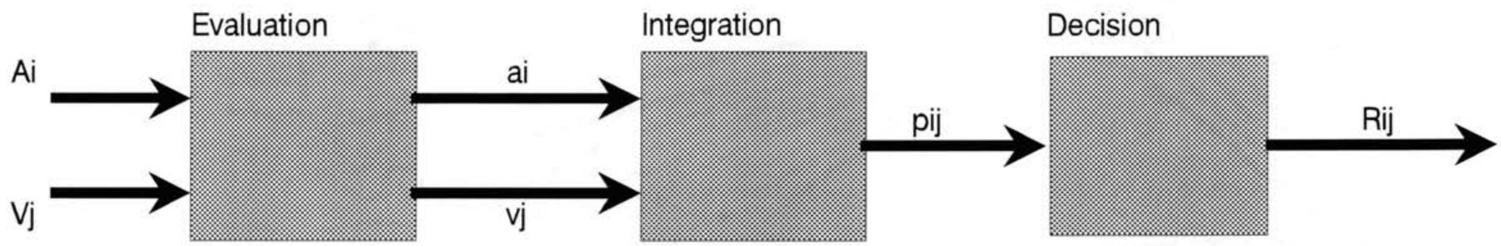
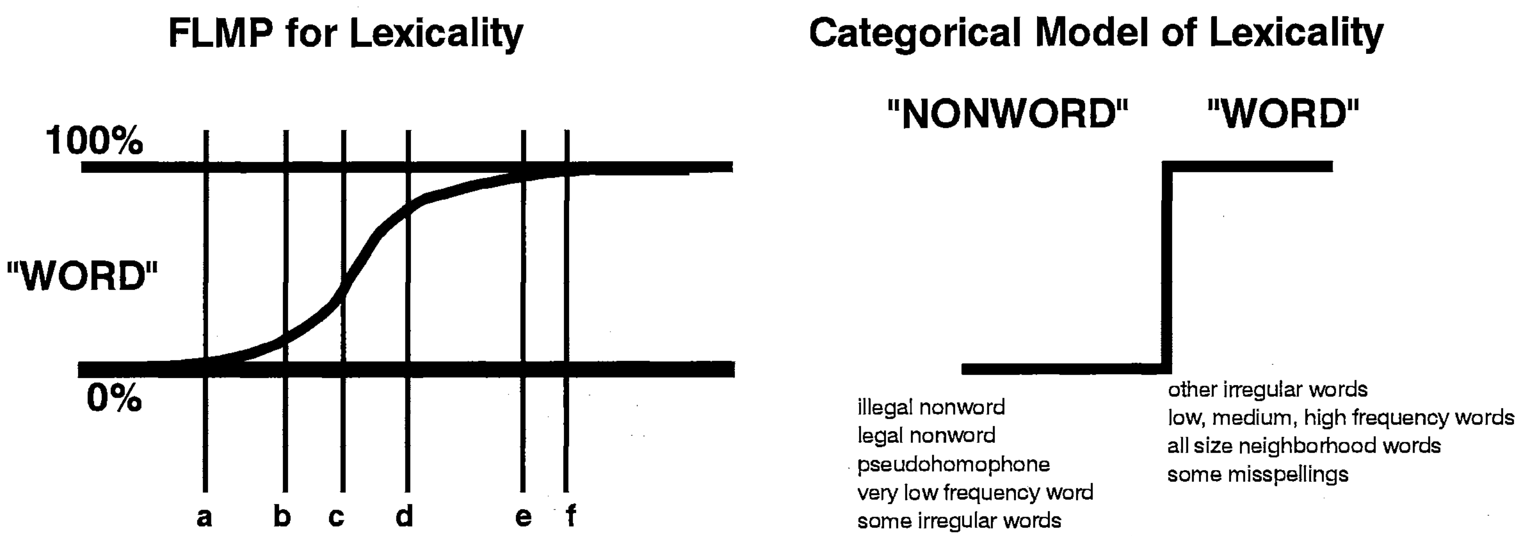


Figure 1. Illustration of the FLMP operations. Upper case letters represent sources of information (audio or visual), while lower case letters indicate psychological values. The overall degree of support following the integration stage is represented by pij, and the response is shown by Rij.



line a: illegal nonword

line b: very low frequency word/legal nonword

line c: irregular word

line d: pseudohomophone

line e: high neighborhood word

line f: high frequency word

Figure 2. Comparison of hypothetical representations of the FLMP and Categorical models for lexical decisions.

	Nonword Catgory: Illegal Nonword	Nonword Catgory: Legal Nonword
Pseudohomophone	Pseudohomophones chosen more often than illegal nonwords	Pseudohomophones chosen more often than legal nonwords
Legal Nonword	Legal nonwords chosen more often than illegal nonwords	Choice determined as as function of bigram frequency - Larger f chosen
Illegal Nonword	Choice determined as as function of bigram frequency - Larger f chosen	Legal nonwords chosen more often than illegal nonwords

Word Catgory
High frequency homonyms chosen over low frequency homonyms
Large neighborhood homonyms chosen over small neighborhood homonyms
High frequency irregular words chosen over low frequency irregular words
Matched-frequency regular words chosen over irregular words

Figure 3. Tabular representation of the various lexicality hypotheses.

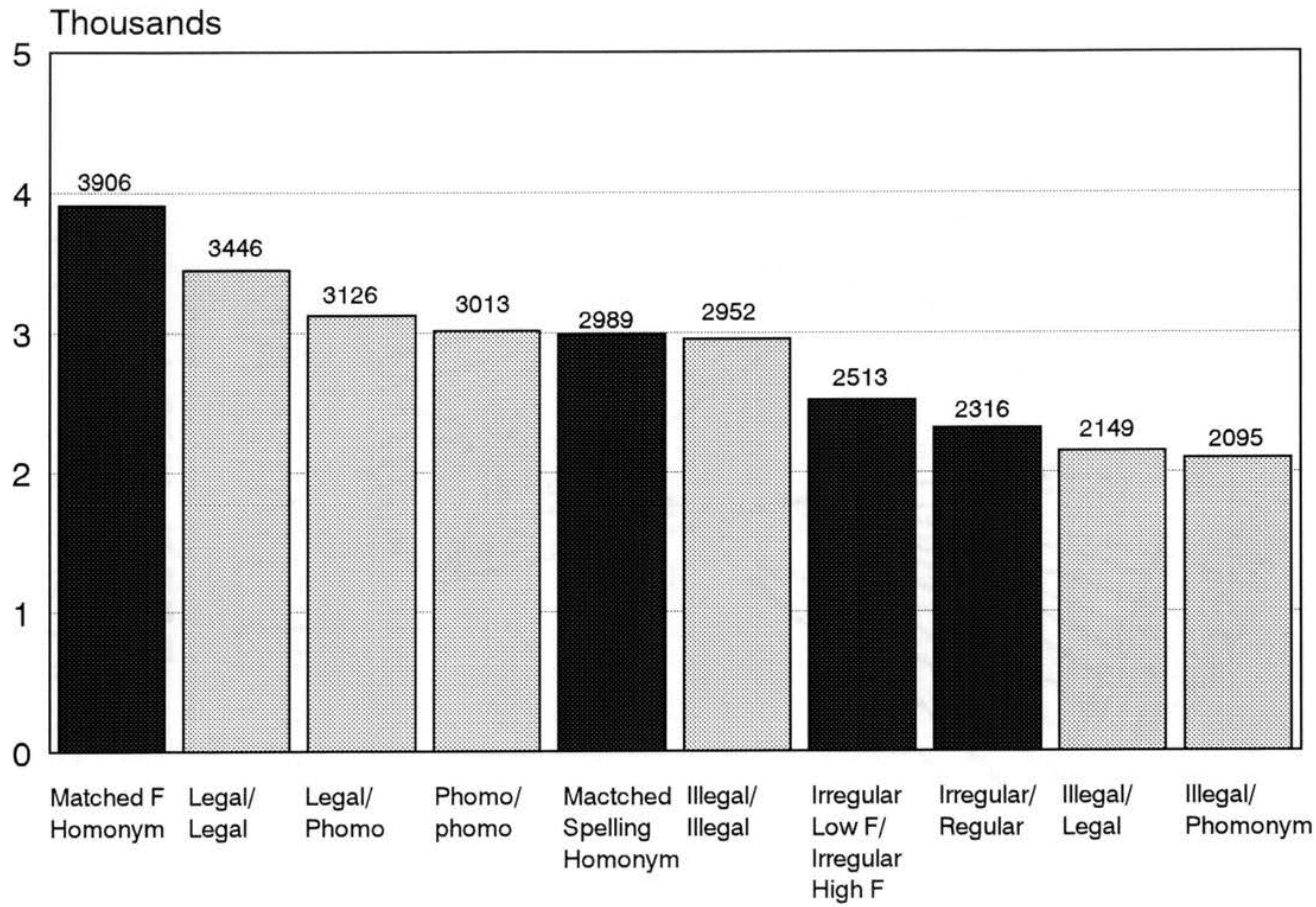


Figure 4. Mean decision times (in ms) for all comparisons

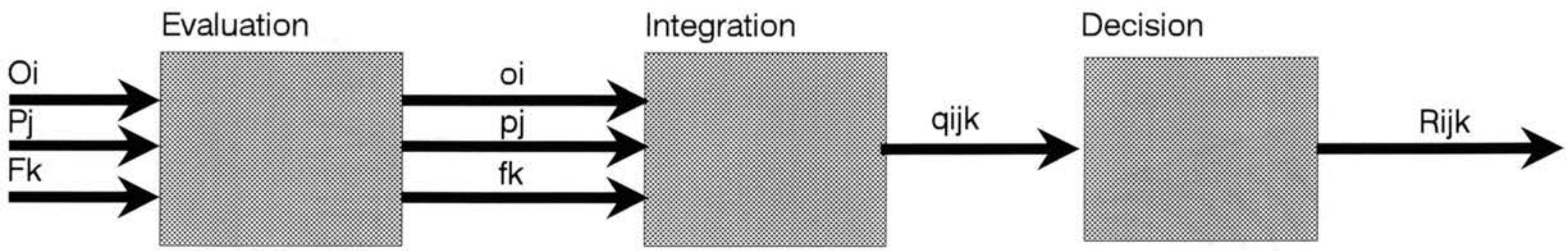


Figure 5. Illustration of the Fuzzy Computational Model of Lexicality (FCML) using orthography (Oi), phonology (Pj), and frequency (Fk) inputs.

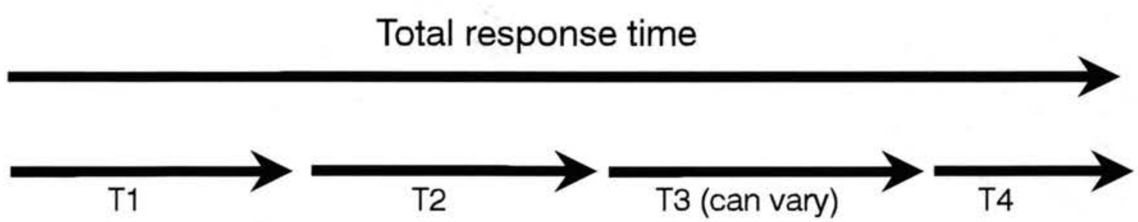


Figure 6. Illustration of the FCML timeline. T1 and T2 are the times needed to evaluate and integrate the input strings. T3 is the time needed to compare the input strings, and can vary. T4 is the time needed to make a response (keypress).

VITA

Robyn Stellman

Candidate for the Degree of

Doctor of Philosophy

Thesis: ON THE FUZZINESS OF LEXICALITY

Major Field: Psychology

Biographical:

Education: Graduated from Rich Central High School, Olympia Fields, Illinois in June 1980; received Bachelor of Science degree in Electrical Engineering from the University of Illinois, Urbana, Illinois in January, 1985; received Master of Science degree in Psychology from Oklahoma State University, Stillwater, Oklahoma, December, 1992. Completed the requirements for the Doctor of Philosophy degree with a major in Experimental Psychology at Oklahoma State University in May 1996.

Experience: Worked as an electrical engineer for General Dynamics Fort Worth Division from January, 1985 until August, 1990. Employed by Oklahoma State University Psychology Department as a teaching assistant from August, 1991 until December, 1995.

OKLAHOMA STATE UNIVERSITY
INSTITUTIONAL REVIEW BOARD
HUMAN SUBJECTS REVIEW

Date: 05-15-95 **IRB#:** AS-95-053

Proposal Title: ON THE FUZZINESS OF LEXICAL DECISIONS

Principal Investigator(s): Larry Hochhaus, Robyn Stellman

Reviewed and Processed as: Expedited

Approval Status Recommended by Reviewer(s): Approved

ALL APPROVALS MAY BE SUBJECT TO REVIEW BY FULL INSTITUTIONAL REVIEW BOARD AT NEXT MEETING.
APPROVAL STATUS PERIOD VALID FOR ONE CALENDAR YEAR AFTER WHICH A CONTINUATION OR RENEWAL REQUEST IS REQUIRED TO BE SUBMITTED FOR BOARD APPROVAL.
ANY MODIFICATIONS TO APPROVED PROJECT MUST ALSO BE SUBMITTED FOR APPROVAL.

Comments, Modifications/Conditions for Approval or Reasons for Deferral or Disapproval are as follows:

Provisions received and approved.

Signature: John H. Wheeler

Chair of Institutional Review Board

Date: June 21, 1995