Immediate Free Recall of Drug Names: Effects of Availability and Familiarity

Bruce L. Lambert, Ph.D.[a, b]

Associate Professor

Ken-Yu Chang, MPH, Ph.D.[c]

Research Scientist

Swu-Jane Lin, B. Pharm., Ph.D.[a]

Assistant Professor

[a]Department of Pharmacy Administration

[b]Department of Pharmacy Practice

University of Illinois at Chicago

[c]Abbott Laboratories

Abbott Park, IL

This project was done when Dr. Chang was a graduate student in the Department of Pharmacy Administration. She is now a pharmacoeconomist at Abbott Laboratories in Abbott Park, IL.

Short Title: Drug Name Recall

Address: Department of Pharmacy Administration,
833 S. Wood Street (M/C 871), Chicago, IL 60612-7231

Phone: 312-996-2411

Fax: 312-996-0868

Email: lambertb@uic.edu

Word Count: 7230 (main body text only); 8736 (acknowledgments, abstract, body text, references)

Acknowledgments

This research was supported in part by the U. S. Pharmacopeia, Inc. (USP) Fellowship Program, the Drug Information Association, and the Campus Research Board of the University of Illinois at Chicago. The authors acknowledge the assistance of Alan Baddeley, Dan Boring, Bill Brewer, Mike Cohen, Diane Cousins, Gary Dell, Sanjay Gandhi, Robert Gibbons, Prahlad Gupta, Keith Johnson, David Lambert, Robert Lee, James Nairne, Sue Proulx, Don Rucker, Gordon Schiff, and anonymous reviewers of earlier versions of this manuscript. Parts of this research were reported on November 17, 1998 at the Annual Meeting of the American Association of Pharmaceutical Scientists in San Francisco, CA and at the conference on Enhancing Patient Safety and Reducing Errors in Health Care, November 9, 1998, Rancho Mirage, CA. Requests for reprints should be sent to Dr. Lambert.

Abstract

The objective of this study was to assess prescribing frequency, subjective familiarity, and two measures of similarity as predictors of error in immediate free recall of drug names. The design utilized prospective, computer-based, word memory experiments. Thirty pharmacists and sixty-six college students viewed a list of three drug names and were immediately asked to recall and write down the names they saw. The main outcome was number of item errors in free recall. Results showed that pharmacists made fewer errors than college students. Familiarity reliably enhanced item recall among both pharmacists and college students. Prescribing frequency enhanced recall among both pharmacists and college students except when college students recalled generic names (in Experiment 4). Orthographic (i.e., spelling) similarity was reliably associated with item recall in both groups. Fewer errors were made when lists were more orthographically similar. Among pharmacists, there was an inverted U-shaped relationship between phonological (i.e., sound) similarity and item errors, with the fewest errors being made on the most similar lists. Among college students, phonological similarity was not reliably associated with item errors. Frequently prescribed and subjectively familiar drug names are more accurately recalled than rarely prescribed and unfamiliar names. Orthographically similar lists of drug names are easier to recall than dissimilar lists because similarity provides cues that facilitate the retrieval of degraded short-term memories. The effects of similarity, familiarity and frequency on short-term memory for drug names vary as a function of task and stimulus characteristics.

Words in Abstract: 237

Key words — drug names, medication errors, patient safety, pharmacists, free recall, short-term memory, similarity, working memory

INTRODUCTION

Medication errors may involve the wrong drug, the wrong dose, the wrong schedule, the wrong combination, or the wrong route of administration.[1] This report focuses only on drug name confusion errors, those that occur when similarity in spelling and/or pronunciation causes practitioners and/or patients to identify an incorrect drug. Thousands of confusing pairs of names have been identified in the literature (e.g., *Xenical*® and *Xeloda*®; *cisplatin* and *carboplatin*; *amiodarone* and *amrinone*; *Vioxx*® and *Videx*®; *Dynacin*® and *DynaCirc*®; *Retrovir*® and *Ritonavir*®).[2-4] Between 12% and 25% of errors voluntarily reported in the U.S. identify name confusion as the primary cause.[5-7] These errors are made by both health professionals and by patients.[8] In order to make the drug use process safer, it is important to identify the objective characteristics of names that affect confusability. Once these are identified, responsible parties can begin to assess and minimize the risk of confusion by avoiding new names with confusable characteristics and taking steps to avert errors involving existing names.[9]

Focus on Human Factors May Reduce Risk

Although the problem of drug name confusion has been recognized for many years, progress in reducing the incidence of such errors has been slow. Several authors have recommended that taking into account human factors when naming and labeling drugs can minimize these errors.[1, 4, 8-13] The term 'human factors' refers to the characteristics and limitations of human attention, perception, memory, judgment, decision-making, and motor control. Knowing how certain characteristics of names affect human performance (especially memory and perception) could help manufacturers, regulators, and practitioners minimize the risk of name confusion. Researchers have made some progress in this direction. For example, retrospective studies have shown that measures of orthographic (i.e., spelling) and phonological

(i.e., sound) similarity can be used to predict which pairs of names are most likely to appear in national error report databases.[10, 11, 14] *Rynatan*®/*Rynatuss*® and *Urex*®/*Erex*® are examples of orthographically similar pairs—pairs of similarly spelled names. *Nasarel*®/*Nizoral*® or *Cardene*®/codeine or *Xanax*®/*Zantac*® are examples of a phonologically similar pairs—words that sound alike, but may or may not be spelled alike. (Of course, in English words that sound alike are often spelled alike and vice versa, but we return to this issue below.) These same measures of similarity predict pharmacists' and laypersons' performance in a short-term recognition memory task[8] and are significantly associated with laypersons' similarity judgments.[15]

Since the human factors approach has succeeded, at least partially, in identifying name characteristics that predict recognition memory errors, the next logical step was to predict performance in another short-term memory task, namely, immediate free recall. Why study immediate free recall? One reason is that free recall is regarded as a basic cognitive function, one which underlies many skilled performances. But that may not be a good enough reason to study it in a pharmacy context. After all, it is not exceedingly common for pharmacists to have to recall short lists of drug names. However, the point is that there are ample opportunities when reading prescriptions, verifying them, and working with them when drug names are recalled without a visual cue immediately available. What practitioners recall may only be a single name, but presumably the underlying process used to generate that name is affected by the same factors that effect the free recall task we have chosen to study. Thus, we have chosen to study a fundamental memory task, free recall, that we believe underlies a wide range of real-world performances.

The literature in cognitive psychology has identified a wide variety of attributes that affect free recall of word lists.[16] With respect to human error generally, Reason has identified similarity and frequency as fundamental mechanisms.[17] The present study assessed prescribing frequency, subjective familiarity, and two measures of similarity with respect to their ability to predict the number of errors in immediate free recall of drug names. The studies were motivated by the following research questions:

RQ1: Can a measure of orthographic similarity reliably predict pharmacists' and college students' performance on a task involving immediate free recall of visually presented drug names?

RQ2: Can a measure of phonological similarity reliably predict pharmacists' and college students' performance on a task involving immediate free recall of visually presented drug names?

RQ3: Can prescribing frequency reliably predict pharmacists' and college students' performance on a task involving immediate free recall of visually presented drug names?

RQ4: Can subjective familiarity reliably predict pharmacists' and college students' performance on a task involving immediate free recall of visually presented drug names?

THEORETICAL BACKGROUND

Immediate Free and Serial Recall

Depending on the task and setting, drug name confusion errors may involve attention, memory, perception, judgment, decision-making, and/or motor control.[1, 8, 17] Here we focused on errors in immediate recall. In an immediate recall task, participants view a list of items (e.g., words or nonsense syllables) and then attempt to recall them. In a *serial recall* task, participants must recall the items in the order that they initially appeared. In a *free recall* task, participants

must recall all of the items but need not recall them in any particular order. This same distinction is sometimes described in terms of memory for *item information* versus memory for *order information*, and this distinction can be operationalized in the task instructions or in the scoring criteria for task performance.[18, 19] The present experiments focused on free recall (i.e., item information) because we were simulating a situation in which a pharmacist or a layperson must recall a list of drugs that they just read on a chart or prescription. In the real world task, order is rarely important, as long as the correct items (i.e. drug names) are recalled.

Models of Immediate Recall

The articulatory loop. For many years, Baddeley's trace-decay/articulatory loop model has been the most influential model of immediate recall.[20] (A *trace* is the pattern encoded in memory after exposure to some to-be-remembered stimulus.) The articulatory loop consists of a finite-capacity phonological store, and a subvocal rehearsal system. The phonological store can be thought of as a mental tape loop that holds 2 seconds of articulated speech sounds. It is subject to decay. The subvocal rehearsal system refreshes decaying traces in the phonological store (e.g., to remember a phone number, silently repeat the number). Short-term memory is thus conceived as a race between rehearsal and decay. Phonologically similar patterns are hard to recall correctly because decay can obliterate the features that distinguish between two phonologically similar words. Imagine trying to remember a sequence of similar words (e.g., *man*, *cap*, *mad*, *cat*, *can, mat* or *Prostigmin®, Prolastin®, progestin*). As soon as the word *man* is stored, its trace begins to decay. When attempting to recall the sequence, the /*n*/ at the end of *man* may have decayed completely. One must reconstruct the partial trace, and the process of reconstruction is vulnerable to similarity-based errors. Often, the correct word will be recalled,

but occasionally, a similar word will be recalled instead.[20] Thus, the articulatory loop model predicts that similarity will degrade performance in immediate free recall.[21]

Articulatory loop anomalies. A series of anomalous results have recently begun to challenge the articulatory loop hypothesis. For example, it has been observed that frequency enhances item recall without affecting rehearsal[22-28], that suppression of rehearsal does not do away with the word frequency or word class effects[22], that phonological similarity may either enhance[19, 29] or not effect item recall[27], that semantic similarity (i.e., similarity in meaning) enhances item recall[18, 30], that words produce better item recall than nonwords[31], that lexical neighborhood size (i.e., the number of words similar to the target word) enhances serial recall[32], and that the number of associative links to a word in long-term memory enhances recall.[33] None of these results is consistent with the articulatory loop model. Instead, these results suggest that short-term recall of word lists is influenced by lexical representations held in long-term memory.[16, 24, 31-34]

The retrieval-based hypothesis. To explain the effects of long-term memory on immediate recall, researchers have put forward the retrieval-based hypothesis.[31] On this view, each item in a list of to-be-recalled words is represented internally as a list of phonological (and other) features.[35] Performance on the recall task works as follows:

> At the point of recall, phonological representations set up by list presentation are thought to be degraded and cannot be output directly as responses. Instead, they must undergo a reconstruction process that calls upon the long-term representation of the to-be-recalled items. In this process, degraded phonological representations are used as retrieval cues for accessing an acceptable recall candidate.[31]

The retrieval-based hypothesis predicts that item recall will be enhanced by any factor that increases the efficiency of the retrieval process.[31] For example, frequent words are predicted to be recalled better than rare words because they are more available to the retrieval mechanism, perhaps by virtue of their more elaborate representation, greater number of associative links, or higher resting level of activation.[16, 28, 36] Phonological and semantic similarity enhance item recall because the common rhymes or semantic features serve as cues to retrieval.[18, 19, 30, 37] For example, one might not remember a word from a list of similar words, but if one knows it rhymes with *–amine* or one knows it is an antibiotic, then one can use that common feature to search long-term memory for the to-be-recalled word. In contrast, any process that degrades the phonological traces of a word, such as suppression of rehearsal, will tend to hinder item recall because degraded traces serve as poor cues for retrieval.[31] The retrieval-based hypothesis also makes predictions about recall of order information (i.e., serial recall), but since the current study focuses only on item recall, those predictions will not be discussed.

Using the retrieval-based hypothesis as a theoretical framework, we designed experiments to test the following hypotheses:

H1: Availability, as measured by outpatient prescription volume and/or subjective familiarity, will enhance item recall of drug names because enhanced availability should lead to easier retrieval of to-be-recalled items.

H2: A measure of orthographic similarity will be positively associated with item recall of drug names, because orthographically similar names will tend to share letter clusters, rhymes, and semantic features that can be used as cues to increase the efficiency of retrieval.

H3: A measure of phonological similarity will be positively associated with item recall of drug names, because phonologically similar names will tend to share letter clusters, rhymes, and semantic features that can be used as cues to increase the efficiency of retrieval.

EXPERIMENT 1

Method

Design

Each participant was asked to recall 15 lists, each containing three, three-syllable drug names. Intralist orthographic similarity was systematically varied. Prescribing frequency was not allowed to be confounded with similarity but was otherwise allowed to vary between lists (see below). Subjective familiarity was not explicitly manipulated and varied both within and between lists. The number of words forgotten or incorrectly recalled was then examined as a function of similarity, subjective familiarity, and prescribing frequency.

Participants

Fifteen licensed, practicing pharmacists participated in Experiment 1. Participants were recruited from the clinical faculty and pharmacy resident staff of an academic medical center in the Midwest United States. All participants held the Pharm.D. degree. Individuals were not paid for their participation. The protocol was reviewed and approved by the local Institutional Review Board, and all participants consented to participate.

Stimulus Materials

Fifteen sets of three names were constructed: five sets each at varying levels of orthographic similarity (i.e., roughly corresponding to high, medium, and low levels, see Table 1). Names and prescribing frequency data were drawn from the drugs listed in the combined 1992-1994 reports of the National Ambulatory Medical Care Survey (NAMCS).[38] Sets of names

at different levels of similarity were matched for frequency to prevent the effects of frequency from confounding the effects of similarity (see Table 1). Both brand (i.e., proprietary) and generic (i.e., nonproprietary) names were used.

Orthographic similarity was measured by the bigram method with one space added to the beginning and ending of each word.[11, 39] The first step in computing the bigram similarity between two words is to break the words into their two-letter subsequences. For example, the bigrams for the drug *Atarax*® are {*_a, at, ta, ar, ra, ax, x_*}. The bigrams for the drug *Marax*® are {*_m, ma, ar, ra, ax, x_*}. The Dice coefficient is then used to compute a similarity score between 0 and 1:

$$Similarity = 2C/(B+A)$$

where *A* was the number of bigrams in the first word, *B* the number of bigrams in the second word, and *C* the number of bigrams that occur in both words. *Atarax*® and *Marax*® share 4 bigrams {*ar*, *ra*, *ax*, *x_*}. Thus, their bigram similarity score is (2*4)/(7 + 6) = 0.62. The intralist similarity score for each three word set in Table 1 was given as the average of the three pairwise similarity scores between names in the list.

Insert Table 1 about here.

Procedures

Experiments were conducted using the SuperLab® experiment program on a Macintosh computer.[40] Participants arrived at the laboratory, read an informed consent script, and were seated in front of a 17-inch color monitor. Participants were instructed that they were going to take part in a test of their memory for drug names. After giving consent and reading a description of the task, each pharmacist completed a single practice trial and then proceeded to complete

fifteen experimental trials. On each trial, three drug names were displayed sequentially using a 36 point Times Roman font in the center of the computer monitor. Each name appeared on the screen for two seconds. Immediately after the third name disappeared from the screen, participants were asked to recall the three names and to write those names on an answer sheet. Participants took as much time as they needed to recall the presented names. The next trial began when the participant hit the space bar on the computer keyboard. The order of appearance of each set of three names was randomized, as was the order of presentation of the three names within a given list. After the experiment was completed, participants were asked to rate the subjective familiarity of each name in the experiment on a seven-point, semantic differential scale that ranged from –3 (i.e., extremely unfamiliar) to +3 (i.e., extremely familiar).

Analysis Plan

The dependent variable, number of errors, was an ordinal variable that ranged from 0-3. Initially any misspelling was coded as an error (see below for a discussion of more lenient scoring rules). The independent variables were: (a) orthographic similarity, a continuous variable reflecting the average pairwise bigram similarity for each three-name list; (b) log frequency, a continuous variable reflecting the log (base 10) of the average NAMCS prescribing frequency of the names in a given list; (c) familiarity, a continuous variable representing the average subjective familiarity of the names in a given list; and (d) trial, an ordinal variable representing the sequential position of a given list within the set of fifteen trials.

Data were analyzed using MIXOR, a system for doing mixed effects, ordinal logistic regression modeling.[41-43] The mixed-effects logistic regression model accommodates nesting of experimental conditions within subjects for an ordinal outcome and a mixture of discrete and

continuous covariates that can vary either at the level of the subject or the experimental condition.[41-43] All statistical tests used alpha = 0.05 as the criterion for significance.

Our modeling strategy included multiple steps. The first step was to identify the correct scale for each independent and control variable. We did this by separately plotting the log odds of error as a function of each independent or control variable. If these plots were linear, terms were entered as simple linear terms. If the plot revealed an obvious nonlinearity, we selected a scale to fit the nonlinear form of the function.[44, 45] In this case, we primarily considered quadratic terms. Having identified the appropriate scale for each independent and control variable, we used Kleinbaum's method of backward elimination to decide which variables to include in the final model.[46] Using this method, the analyst begins with a full model and then proceeds to eliminate as many terms as possible, using likelihood ratio tests (analogous to partial F-tests in ordinary least squares regression) to decide which terms contribute significantly to the model's fit. Higher order terms (e.g., interaction terms, squared terms) are eliminated first, then first order terms. Interested readers should consult the text for a complete description of the approach.[46]

The final step in our modeling strategy was to assess goodness-of-fit. Unlike the case of ordinary least squares regression, where R^2 provides an agreed-upon measure of fit, in logistic regression, there is no consensus measure of goodness-of-fit.[44-47] For each model, we graphically showed the fit between observed and predicted probabilities of error at selected levels of similarity. To compute predicted probabilities, we evaluated the fitted model with all control variables set to their subgroup means and similarity set to the average value within a given level.[8]

Results

The mean number of errors per trial was 0.85 (SD = 0.91). Mean familiarity was 0.56 (SD = 1.47). The correlation between familiarity and frequency was 0.45 ($p < 0.01$). Overall, 44% of the lists were recalled with zero errors, 32% with one error, 18.7% with two errors, and 5% with three errors. Estimated coefficients to the ordinal logistic regression model are given in Table 2, along with the associated standard errors and probabilities. Figure 1 provides a graphical display of the observed and predicted error rates. The fit between model and data was good.

Insert Table 2 and Figure 1 about here.

As Table 2 shows, orthographic similarity had a significant effect on pharmacists' immediate recall of drug names ($p < 0.0001$). As similarity increased, the number of errors in recall *decreased*. Both log frequency and familiarity had significant effects on the number of recall errors ($p < 0.001$ and $p < 0.0001$ respectively), with fewer errors being made for more frequent and more familiar names.

Upon further examination, it was clear that errors fell into three categories: omissions, substitutions involving misspellings (e.g., *Anuphin* instead of *Anuphen*®) and syllable deletions (e.g., *Distine* instead of *Dihistine*®), and substitutions involving other drug names (e.g., *Decadron*® instead of *Decaspray*®). Substitutions involving misspellings and transpositions were by far the most common, accounting for 158 of 192 total errors (82%). Most of these substitutions were quite close to the target names. To see how close, we used the edit distance measure to compute the distance between target name and substituted name. Edit distance measures the number of insertions, substitutions, and/or deletions needed to transform one word into another.[10, 39] Of the 158 substitution errors, 68 (43%) were within one edit, 29 (18%) were

within 2 edits, and 23 (15%) were within 3 edits. The remaining 38 (24%) were more than 3 edits away from the target. The large number of errors that appeared to be misspellings suggested that a different method of tallying true recall errors might have been appropriate. It is possible that some of the apparent recall errors were actually recalled correctly but subsequently misspelled or mistyped. These "output" errors, although clearly unacceptable in pharmacy context, should be distinguished from true errors in recall. To examine this possibility, the error data were reanalyzed with a more liberal definition of a correct response (assuming close spelling errors were not true recall errors). In separate analyses, we defined a correct response as any response within 1, 2, or 3 edits respectively. With one exception, in each of these analyses, increased familiarity, frequency, and similarity were still associated with improvement in recall ($p < 0.01$, details not shown). When names misspelled by 3 edits were coded as correct, frequency no longer had a reliable effect. Among the 34 omissions, some words were more common than others. Nineteen names accounted for all 34 omissions, with *Aldactone*® and *Cetapred*® both being omitted 5 times each. Only three errors involved potentially the most dangerous kind of substitution, when one drug name is substituted for another. *Imdur*® was recalled in place of *Imuran*®. *Decadron*® was recalled instead of *Decaspray*®, and *Aldactone*® was recalled as *Aspercreme*®.

Discussion of Experiment 1

More frequent and familiar names were recalled more accurately than their less frequent counterparts, regardless of similarity. Thus the data supported Hypothesis 1. This is a straightforward example of the well-known word frequency effect, an effect thought to be caused by the greater cognitive availability of more frequently occurring words.[16, 28, 48, 49] Results also supported Hypothesis 2. The bigram measure of orthographic similarity was a reliable predictor

of immediate free recall, with errors decreasing as similarity increased. This result is consistent with the retrieval-based hypothesis and contradicts the articulatory loop hypothesis. Similar lists are easier to recall because similarity provides additional cues that aid retrieval from long-term memory. It is important to note that, even though orthographic similarity reliably predicted recall error rates, the effects may not be due to orthography but to semantic and phonological similarity which, in the drug name lexicon, are confounded.[19, 30, 50] For example, the medium and high similarity lists from Table 1 are more likely to contain rhymes than the low similarity lists. Rhymes are known to enhance item recall.[19] In addition, names with high and medium levels of orthographic similarity are more likely to share semantic features than low similarity lists. For example, high similarity lists are more likely to contain drugs with the same indication or drugs from the same pharmacologic category. Semantic similarity (e.g., shared category membership) is also known to enhance item recall.[18, 30]

EXPERIMENT 2

Method

Design and Participants

Experiment 2 was designed to parallel Experiment 1, the main difference being that this experiment operationalized similarity in terms of phonology (i.e., the sound pattern of a word) rather than in terms of orthography (i.e., spelling). Fifteen licensed, practicing pharmacists participated in Experiment 2. This group of pharmacists was distinct from those who participated in Experiment 1, although both sets of participants were recruited from the same large group of practitioners at an academic medical center in the Midwest United States. The protocol was reviewed and approved by the local Institutional Review Board, and all participants consented to participate.

Stimulus Materials

Materials for this experiment were 15 3-word lists of generic drug names, 5 each at 3 levels of similarity (roughly corresponding to high, medium, and low; see Table 3). The majority of names had four syllables. Each 3-word list had a total of 12 syllables. To be included in this experiment, names had to be listed both in the NAMCS database and in the USP Dictionary of USAN and International Drug Names[51] because prescribing frequency data were taken from the NAMCS database, and pronunciation guides were taken from the USP Dictionary.

A measure of phonological similarity was developed for this experiment.[10] Several phonological characteristics had been identified as important in previous research on similarity and memory.[52] These included number of syllables, location of stressed syllable, initial syllable, terminal syllable, and stressed vowel. Based on these features, similarity was defined as follows:

$$PhonoSim(word_i, word_j) = 0.5\left(\frac{2C}{B+A}\right) + 0.5\frac{(2D+E+F+G+H)}{6}$$

where *A* was the number of syllables in one word, *B* was the number of syllables in the other word, *C* was the number of common syllables, *D* was a binary feature representing a match between initial syllables, *E* was a binary feature representing a match between terminal syllables, *F* was a binary feature representing a match between accented syllables, *G* was a binary feature representing a match between accent positions, and *H* was a binary feature representing a match between number of syllables. This measure gave half of the weight to the commonality in syllables and half of the weight to specific phonological features, with initial phoneme getting twice the weight as the other phonological features.

To compute the phonological similarity between lincomycin and tobramycin, we began by retrieving the respective pronunciation guides from the USP Dictionary of USAN and International Drug Names: (*lin koe mye' sin*) and (*toe bra mye' sin*). An apostrophe indicated the accented syllable. These names shared two syllables {*mye*', *sin*}. They had different initial syllables ($D = 0$). They had the same accented syllable ($E = 1$) in the same accent position $F = 1$). They had the same number of syllables ($G = 1$) and the same terminal syllable ($H = 1$). Thus, the similarity between these two names was calculated as follows:

$$PhonoSim(lincomycin, tobramycin) = 0.5\left(\frac{2\times 2}{4+4}\right) + 0.5\frac{(2\times 0+1+1+1+1)}{6} = 0.58$$

The similarity for a list of names was computed as the average of all pairwise similarities between names in the list.

Insert Table 3 about here.

Procedures and Analysis Plan

The procedure and analysis plan were identical to those used in Experiment 1.

Results

The mean number of errors per trial was 0.57 (SD = 0.77). Mean familiarity was 2.21 (SD = 0.98). The correlation between frequency and familiarity was 0.26 ($p < 0.01$). Overall, 59.52% of the lists were recalled with zero errors, 27.14% with one error, 11.4% with two errors, and 1.90% with three errors. Parameter estimates for the logistic regression model given in Table 4. Figure 2 displays the results graphically, and it indicates a good fit between model and data. As in Experiment 1, similarity had a significant effect on the number of errors made in recall ($p < 0.01$) (see Table 4). But in this case, so did a quadratic term for similarity ($p < 0.001$). The

relationship between similarity and the probability of error was nonlinear. Figure 2 reveals an inverted U-shape, with probability increasing from low to mid levels of similarity and decreasing from mid to high levels. Familiarity had a significant effect on recall errors ($p < 0.0001$) as did prescribing frequency ($p < 0.05$, see note to Table 4). Frequently prescribed and familiar words were recalled more accurately than rarely prescribed or unfamiliar words.

Insert Table 4 and Figure 2 about here.

As in Experiment 1, inspection of the 129 actual errors revealed 9 omissions (6.9%) and 120 substitutions/spelling errors (93.3%). Fifteen of the 120 substitutions (12.5%) involved other drug names. Close misspellings were again very common. Fifty-seven of the 120 misspelling errors (47.5%) were within one edit of the target. Twenty-four (20%) were within 2 edits, and 7 were within 3 edits. The 9 omissions involved only 5 different names. When a more lenient scoring criterion was used, allowing names within an edit distance of 3 edits to be coded as correct, the results were unchanged (details not shown).

Discussion of Experiment 2

Experiment 2 supported Hypothesis 1. Familiarity and frequency were reliably associated with item recall, as predicted by the word frequency effect and the retrieval-based hypothesis. The most plausible explanation is that, compared to rare and unfamiliar words, common and familiar words are more cognitively available and hence easier to retrieve from long-term memory when cued by degraded traces from short-term memory. Results of Experiment 2 also partially supported Hypothesis 3. Phonological similarity did reliably predict the number of errors in free recall, but the shape of the relationship was not as predicted. We had predicted a monotonic increase in errors with increasing similarity. Instead, the relationship was quadratic,

in the form of an inverted U. At present we have no satisfactory explanation for the nonlinear effects of similarity observed here.

EXPERIMENT 3

Method

Design and Participants

Experiment 3 was designed to be identical to Experiment 1 except with respect to the participants. The sample for this experiment was larger and included no health professionals. Participants in Experiment 3 were 33 college students. The majority of students were undergraduate psychology majors who participated in exchange for course credit. A small number of participants were recruited from the general student population. These students were each paid 10 dollars for their participation. The protocol was reviewed and approved by the local Institutional Review Board, and all participants consented to participate.

Materials, Procedures and Analysis Plan

The stimulus materials, experimental procedures, and analysis plan were identical to those used in Experiment 1 (see Table 1).

Results

The mean number of errors per trial was 1.94 (SD = 0.98). Mean familiarity was –1.84 (SD = 1.43). The correlation between frequency and familiarity was –0.03 (n.s.). Overall, 9.89% of the lists were recalled with zero errors, 21.61% with one error, 33.53% with two errors, and 34.95% with three errors. Parameter estimates for the ordinal logistic regression model are in Table 5. Figure 3 charts error rate as a function of orthographic similarity and illustrates the fit between observed and predicted values. Similarity had a significant effect on the number of errors made in free recall ($p < 0.0001$). As similarity increased, the probability of error

decreased. Log frequency and familiarity both significantly affected the probability of error ($p < 0.01$ and $p < 0.001$ respectively), with frequent and familiar names being recalled more accurately than less frequent and less familiar names.

Insert Table 5 and Figure 3 about here

As in Experiments 1 and 2, we examined the pattern of students' recall errors. Out of a total of 958 errors, 87 were omissions (9.1%), 866 were substitutions/misspellings (90.4%), and 5 were dangerous substitutions (0.52%), where the recalled name was another drug name. Of these errors, 246 (25.7%) were within a single edit of the target word, 159 (16.59%) were two edits away, and 114 (11.9%) were three edits away. When the data were analyzed with more lenient scoring criteria, i.e., when errors within three edits were coded as correctly recalled, the effects of similarity, frequency, and familiarity were unchanged (details not shown).

Discussion of Experiment 3

The pattern of effects was generally consistent with what was observed for pharmacists in Experiment 1. Hypotheses 1 and 2 were supported. Accuracy in recall improved as similarity increased, supporting the retrieval-based hypothesis and contradicting the articulatory loop-based hypothesis. As in Experiments 1 and 2, frequency and familiarity enhanced recall. Overall error rates were higher for college students than for pharmacists, presumably reflecting pharmacists' greater familiarity with drug names. Specifically, the retrieval hypothesis states that short-term recall depends on cued retrieval of words from long-term memory. To the extent that pharmacists have more elaborate and highly interconnected long-term representations of drug names than do college students, pharmacists would be expected to have better item recall. The

explanation for the observed pattern of effects, described in detail in the discussion of Experiment 1, is essentially unchanged.

EXPERIMENT 4

Method

Design, Participants, Materials, Procedures and Analysis Plan

The design, stimulus materials, experimental procedures, and analysis plan were identical to those used in Experiment 2 (see Table 4). The participants were 33 college students drawn from the same population as in Experiment 3. The protocol was reviewed and approved by the local Institutional Review Board, and all participants consented to participate.

Results

The mean number of errors per list was 2.58 (SD = 0.65). Mean familiarity was –2.18 (SD = 1.15). The correlation between frequency and familiarity was –0.09 ($p < 0.01$). Participants recalled 1.29% of the lists with zero errors, 4.76% with one error, 28.78% with two errors, and 65.15% with three errors. Figure 4 displays these results graphically. Parameter estimates for the logistic regression model are in Table 6. Only familiarity was significantly associated with the number of errors made in free recall ($p < 0.001$). More familiar words were recalled more accurately than less familiar words. None of the other coefficients was significantly different from zero.

Insert Table 6 and Figure 4 about here

From a total of 1286 errors, 178 were omissions (13.8%), 1108 were substitutions (86.2%). Five of the substitutions involved other drug names. One hundred forty eight of the substitutions (11.5%) were within a single edit of the target name; 156 (12.1%) were 2 edits

away, and 158 (12.3%) were 3 edits away. The significant effect of familiarity was unchanged by the use of a more lenient scoring criterion, wherein any name within three edits of the correct name was scored as correct (details not shown).

Discussion of Experiment 4

Results of Experiment 4 partially supported Hypothesis 1. Availability, as measured by familiarity, but not frequency, was reliably associated with errors in free recall of drug names. It may be the case that familiarity overwhelms the effects of frequency, especially at high levels of unfamiliarity such as those observed in Experiment 4. Hypothesis 3 was not supported by the results of Experiment 4. Phonological similarity was not a reliable predictor of item recall performance for college students. However, these results are still consistent with the retrieval-based hypothesis. According to the retrieval-based hypothesis, similarity aids in recall by providing cues (e.g., rhymes, semantic category information) that increase the efficiency of retrieval from long-term memory. But, if college students have no long-term memory representations of the generic drug names used in Experiment 4, as evidenced by their lack of familiarity with the names, then the retrieval process is inoperative, and the additional similarity-based retrieval cues would have no impact on item recall.[27] Overall, the very poor item recall for the generic names in this experiment is reminiscent of Saint-Aubin's study of recall for non-words.[31]

LIMITATIONS

Only clinical pharmacists and college undergraduates from the Midwest region of the U.S. participated in these experiments. Generalization of the observed effects to non-Midwesterners, non-pharmacist health professionals or to the lay population at large may or may not be warranted. Caution should be used when generalizing the results to community

pharmacists, since our participants were all clinical pharmacists from an academic medical center. Because of sampling error in the original NAMCS survey, some of the prescribing frequency data which we used lacked precision, especially estimates below 2 million prescriptions per year (i.e., log frequency less than 6).[53] This lack of precision may have clouded the effects of frequency on recall. Experiments 1 and 3 included very few generic drug names. Thus, the effects we observed in those experiments can be generalized with greatest confidence only to other brand names. Similarly, Experiments 2 and 4 were based exclusively on generic names and may or may not be generalizable to brand names. In each of the experiments, the absolute error rate, as opposed to the trend in error rates, should not be overemphasized, since it may have been affected by stimulus duration, font size, list length, and other extraneous characteristics of the experimental task.

Finally, our experimental design limited what one can validly conclude from the results reported above. For example, when we manipulated orthographic similarity in Experiments 1 and 2, we did not simultaneously control for phonological and semantic similarity, even though, in the real drug name lexicon, these factors are confounded. That is, names that are spelled similarly tend also to be pronounced similarly and, to a slightly lesser extent with brand than generic names, tend to share semantic features (e.g., pharmacologic category, indication, mechanisms of action). One *can* conclude from our results that our measure of orthographic similarity will reliably be associated with enhanced recall of drug names, but one *cannot* necessarily conclude that this effect is *caused by* orthographic similarity. In order to draw that conclusion, one would have to vary orthographic similarity while holding phonological and semantic similarity constant. The same limitations apply to Experiments 3 and 4, where we manipulated phonological similarity without controlling semantic or orthographic similarity. One

can conclude that our phonological similarity measure will be reliably associated with drug name recall (among clinical pharmacists), but one must be silent about the precise cause of the observed effects (i.e., they may be due to phonological, orthographic or semantic similarity, or some combination thereof). Given these limitations, one might question our design choices. Why not control for confounding and examine the unique effects of each type of similarity?

The long-term goal of our research program is to design, build and test tools that help to predict and prevent drug name confusions in the real world. We were focusing on an applied question—will this particular measure orthographic or phonological similarity predict recall performance among pharmacists and laypersons?—rather than a theoretical question—how does phonological (orthographic, semantic) similarity affect immediate free recall? Had we been able to construct sets of stimulus materials that varied each type of similarity while holding the others constant (and it is not clear we could have done so even if we wanted to), we would have, in effect, created an artificial drug name lexicon, one which bore little resemblance to the real lexicon. Since these types of similarity (orthographic, phonological, and semantic) are, in fact, confounded in the real drug name lexicon, anything we learned from the artificial lexicon would have limited applicability to the real world. Since we were primarily interested in evaluating the performance of specific measures on real world data, we chose to conduct less rigorously controlled, but more realistic, experiments. Our design decisions allowed conclusions to be drawn about the usefulness of specific measures in the context of the real drug name lexicon. Conclusions about the precise causal role of each type of similarity, although important, must await the results of future experimentation. In this regard, at least three additional experiments are needed: one which examines the effect of phonological similarity while holding semantic and orthographic similarity constant, one which varies semantic similarity while holding

phonological and orthographic similarity constant, and one which varies orthographic similarity while holding phonological and semantic similarity constant.

One final limitation is worth noting in regard to the phonological similarity measure used here. Although based explicitly on features identified in previous studies, the measure is totally reliant on pronunciation guides from the USP Dictionary. This raises several problems. First, one can only use the measure on USAN names or names whose pronunciation has manually been coded into the USAN format. Second, the USAN pronunciation guide is itself somewhat ad hoc and does not rely on accepted phonological formalisms such as the International Phonetic Alphabet or the ARPAbet system.[54, 55] These limitations can be overcome in the future by using text-to-phoneme translation algorithms that transform the orthographic representation of a word into a more standard phonological representation.[54-56] Future publishers of drug information should consider providing pronunciation guides, in a standard formalism, for all brand and generic drug names.

GENERAL DISCUSSION AND POLICY IMPLICATIONS

The authors of the Institute of Medicine's report on medical errors have argued that patient safety will be improved on a large scale only when experts from human factors, cognitive psychology, and industrial quality control (among others) begin to focus on understanding and preventing medical errors.[57] The current project was motivated by the desire to do just that—to bring the theories and methods of cognitive psychology to bear on the problem of drug name confusion errors.[8] The best available theory predicted that similarity, frequency, and familiarity would enhance immediate recall of drug names, and this, for the most part, is what we found. What are the policy implications of these findings? Should companies strive to make names as similar as possible in order to facilitate recall? Such a suggestion runs counter to all of our other

work, where we have argued that similarity should be minimized in order to reduce the name confusion error rate.[8, 10, 11, 14, 58]

It now appears that the effects of similarity on performance depend on the nature of the task. In recognition memory and in visual perception, similarity degrades performance.[8, 58] In immediate free recall, it enhances performance. To decide how to handle similarity from a policy point of view, one must have an understanding of the distribution of tasks in the real-world practice environments. If most tasks involve immediate recall, we would expect a large scale study of name confusion errors to reveal that more similar pairs of names are less likely to be confused. On the other hand, if most tasks in the real world involve recognition memory and visual or auditory perception, one would expect such a study to reveal that more similar names are more likely to be confused. In both of our large-scale analyses of error reports, we have shown that similarity is strongly associated with increased risk of name confusion.[10, 11] We therefore conclude that most tasks in the real world must involve cognitive processes that are undermined by similarity (e.g., recognition memory, visual perception, auditory perception). This conclusion also accords with common sense, which suggests that people rarely need to recall lists of drug names, but they often need to recall and recognize individual names, and they must often decipher ambiguous written or spoken names. As a result, we maintain the belief that similarity is an important risk factor for drug name confusions. The present experiments provided a more complex and differentiated picture, emphasizing the task-dependent nature of similarity effects. On balance, though, when it comes to drug names, similarity still appears to do more harm than good.

The effects of frequency and familiarity also have potential policy implications. The main implication being that frequent and familiar names, those that are more cognitively available for perception and memory, are less likely to be forgotten or misperceived (although this effect may

also be somewhat task dependent). It is not clear, however, how to put this type of knowledge into practice because one cannot explicitly manipulate familiarity or frequency the way one can manipulate similarity (e.g., by changing the spelling of a name). At the time a name is evaluated for approval by FDA, its eventual prescribing frequency or familiarity may not be known with any precision. What's more, even if a drug is likely to become very widely used (because it is an especially novel, safe or effective remedy for a common ailment), it is not clear how long it may take for the familiarity effects to take hold. Immediately post launch, any new name, regardless of its eventual familiarity or prescribing frequency, will be *relatively* unfamiliar when compared to common older names. If reliable projections can be made about frequency and familiarity, it may make sense to tolerate slightly more similarity for very frequent or familiar names, given their inherent advantages in memorability and perceptibility, but this question requires additional examination before a final policy can be formulated.

The underlying causes of the observed similarity effects were not unequivocally identified in the present study due to lack of controls for confounding between different types of similarity. This is the first of our studies that has shown a processing advantage for similar drug names. Overall, it appears that the effects of similarity are task dependent. In some cases, similarity will improve performance; in others, it will undermine performance. When deciding how to use similarity information in drug naming decisions, it will be important to consider the distribution of tasks and how similarity is likely to influence those tasks. Our previous work suggests that, across all relevant tasks, the effects of similarity do more harm than good. Therefore, we stand by our previously published recommendation that objective measures of similarity should be used by decision-makers when evaluating the acceptability of new drug

names, and that similarity between new and existing drug names should be minimized whenever possible.

CONCLUSION

Subjective familiarity, prescribing frequency, and the bigram measure of orthographic similarity are each reliable predictors of item recall in pharmacists' and laypersons' short-term memory for drug names. Familiarity, frequency, and orthographic similarity enhance item recall. Phonological similarity, measured using USAN pronunciation guides, is less reliable, showing inconsistent or null patterns of association with item recall performance. Results lend support to the retrieval-based model which suggests that short-term memory performance is influenced by long-term memory representations.

References

1. Cohen M, ed. *Medication errors*. Washington, DC: American Pharmaceutical Association; 1999.

2. Davis NM. Drug names that look and sound alike. *Hosp Pharm.* 1997;32(11):1558-1570.

3. U. S. Pharmacopeia. USP Quality Review. *U. S. Pharmacopeia*. March. Available at: http://www.usp.org/reporting/review/qr76.pdf. Accessed July 19, 2001.

4. Keneagy JW, Stein GC. Naming, labeling, and packaging of pharmaceuticals. *Am J Health-Syst Pharm.* 2001;58(21):2033-2041.

5. U. S. Pharmacopeia. *USP Quality Review.* Rockville, MD: U. S. Pharmacopeia; July 1995. 49.

6. U. S. Pharmacopeia. *Drug product quality review.* Rockville, MD: U. S. Pharmacopeia; March 1993. 34.

7. U. S. Pharmacopeia. Similar names. *USP DI Update Volume I and II.* 1997:4-7.

8. Lambert BL, Chang KY, Lin SJ. Effect of orthographic and phonological similarity on false recognition of drug names. *Soc Sci Med.* 2001;52:1843-1857.

9. Grasha AF. Into the abyss: Seven principles for identifying the causes of and preventing error in complex systems. *Am J Health-Syst Pharm.* 2000;57(March 15):554-564.

10. Lambert BL, Lin S-J, Gandhi SK, Chang K-Y. Similarity as a risk factor in drug name confusion errors: The look-alike (orthographic) and sound-alike (phonological) model. *Med Care.* 1999;37(12):1214-1225.

11. Lambert BL. Predicting look- and sound-alike medication errors. *Am J Health-Syst Pharm.* 1997;54:1161-1171.

12. Grasha AF, O'Neill M. Cognitive processes in medication errors. *U.S. Pharmacist.* 1996;June:1-9.

13. Grasha AF, Reilly S, Tranum D. *A cognitive systems perspective on human performance in the pharmacy: Implications for accuracy, effectiveness, and job satisfaction.* Cincinnati, OH: University of Cincinnati; October, 2000.

14. Lambert BL, Lin S-J, Gandhi SK, Chang K-Y. Predicting and preventing drug name confusion errors: A summary of findings. In: Scheffler AL, Zipperer LA, eds. *Proceedings of Enhancing Patient Safety and Reducing Errors in Health Care*. Rancho Mirage, CA: National Patient Safety Foundation; 1999:221-225.

15. Lambert BL, Donderi D, Senders J. Similarity of drug names: Objective and subjective measures. *Psychology and Marketing.* 2002;19(7-8):641-661.

16. Rubin DC. Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronounciability for 925 nouns. *Mem Cognit.* 1986;14(1):79-94.

17. Reason J. *Human error*. New York: Cambridge University Press; 1990.

18. Saint-Aubin J, Poirier M. Semantic similarity and immediate serial recall: Is there a detrimental effect on order information. *Q J Exp Psychol* 1999;52A(2):367-394.

19. Fallon AB, Groves K, Tehan G. Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT but not MAN. *Int J Psychol.* 1999;34(5/6):301-307.

20. Baddeley AD. *Working memory*. Oxford: Oxford University Press; 1986.

21. Coltheart V. Effects of phonological similarity and concurrent irrelevant articulation on short-term-memory recall of repeated and novel word lists. *Mem Cognit.* 1993;21:539-545.

22. Tehan G, Humphreys MS. Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bull Psychon Soc.* 1988;26(4):293-296.

23. Roodenrys S, Quinlan PT. The effects of stimulus set size and word frequency on verbal serial recall. *Memory.* 2000;8(2):71-78.

24. Hulme C, Roodenrys S, Schweickert R, Brown GDA, Martin S, Stuart G. Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *J Exp Psychol Learn Mem Cogn.* 1997;23(5):1217-1232.

25. Watkins MJ. The intricacy of memory span. *Mem Cognit.* 1977;5:529-534.

26. Roodenrys S, Hulme C, Alban J, Ellis AW, Brown GDA. Effects of word frequency and age of acquisition on short-term memory span. *Mem Cognit.* 1994;22(6):695-701.

27. Poirier M, Saint-Aubin J. Immediate serial recall, word frequency, item identity, and item position. *Can J Exp Psychol.* 1996;50(4):408-412.

28. Gregg V. Word frequency, recognition, and recall. In: Brown J, ed. *Recall and recognition*. New York: John Wiley & Sons; 1976:183-216.

29. Gathercole SE, Gardiner JM. Modality and phonological similarity effects in serial recall: Does one's own voice play a role. *Mem Cognit.* 1982;10(2):176-180.

30. Poirier M, Saint-Aubin J. Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Q J Exp Psychol* 1995;48A(2):384-404.

31. Saint-Aubin J, Poirier M. Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis. *Psychon Bull Rev.* 2000;7(2):332-340.

32. Roodenrys S, Hinton M. Sublexical or lexical effects on serial recall of nonwords. *J Exp Psychol Learn Mem Cogn.* 2002;28(1):29-33.

33. Stuart G, Hulme C. The effect of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *J Exp Psychol Learn Mem Cogn.* 2000;26(3):796-802.

34. Saint-Aubin J, Poirier M. The influence of long-term memory factors on immediate serial recall: An item and order analysis. *Int J Psychol.* 1999;34(5/6):347-352.

35. Nairne JS. A feature model of immediate memory. *Mem Cognit.* 1990;18(3):251-269.

36. Anderson JR, Bothell D, Lebiere C, Mantessa M. An integrated theory of list memory. *J Mem Lang.* 1998;38:341-380.

37. Nairne JS, Neumann C. Enhancing effects of similarity on long-term memory for order. *J Exp Psychol Learn Mem Cogn.* 1993;19(2):329-337.

38. U. S. Department of Health and Human Services. 1993 National Ambulatory Medical Care Survey. *National Center for Health Statistics* [CD-ROM]. Available at: http://www.cdc.gov/nchswww/products/catalogs/subject/cdprice.htm#namcscdprice. Accessed September 23, 1999, 1999.

39. Stephen GA. *String searching algorithms*. River Edge, NJ: World Scientific; 1994.

40. Cedrus. SuperLab. *Cedrus*. Available at: http://www.cedrus.com/. Accessed October 8, 1999, 1999.

41. Hedeker D, Gibbons RD. MIXOR: a computer program for mixed-effects ordinal regression analysis. *Comput Methods Programs Biomed.* 1996;49:157-176.

42. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics.* 1994:933-944.

43. Hedeker D. MIXOR/MIXREG Website. *University of Illinois at Chicago* [HTML page]. Available at: http://www.uic.edu/~hedeker/mix.html. Accessed June 4, 1999.

44. Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: John Wiley & Sons; 1989.

45. Selvin S. *Statistical analysis of epidemiological data*. New York: Oxford University Press; 1996.

46. Kleinbaum DG. *Logistic regression: a self-learning text*. New York: Springer-Verlag; 1994.

47. Pedhazur EJ. *Multiple regression in behavioral research: Explanation and prediction*. New York: Harcourt Brace; 1997.

48. Deese J. Frequency of usage and number of words in free recall: The role of association. *Psychol Rep.* 1960;7:337-344.

49. Anderson JR, Lebiere C, eds. *The atomic components of thought*. Mahwah, NJ: Erlbaum; 1998.

50. Baddeley AD. Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Q J Exp Psychol* 1966;18:362-365.

51. U. S. Pharmacopeia. *USP dictionary of USAN and international drug names*. Rockville, MD: U. S. Pharmacopeia; 1998.

52. Drewnowski A, Murdock BB, Jr. The role of auditory features in memory span for words. *J Exp Psychol [Hum Learn].* 1980;6:319-332.

53. Woodwell DA. National Ambulatory Medical Care Survey: 1997 summary. *Adv Data.* May 20, 1999 1999;305:1-26.

54. Jurafsky D, Martin JH. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall; 2000.

55. Manning CD, Schutze H. *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press; 1999.

56. Fisher WM. A statistical text-to-phone function using ngrams and rules. Paper presented at: IEEE International Conference on Acoustics, Speech and Signal Processing; March 15-19, 1999; Phoenix, AZ.

57. Kohn LT, Corrigan J, Donaldson MS, eds. *To err is human: building a safer health system*. Washington, DC: Institute of Medicine; 2000.

58. Lambert BL, Chang K-Y, Lin S-J, Gupta P. Effect of prescribing frequency, neighborhood frequency, and neighborhood density on pharmacists' visual perception of drug names. *PharmSci.* 2000;2:S-1276.

Table 1. Stimulus words for recall experiments using orthographic similarity measure (Experiments 1 and 3)

Frequency	Similarity	Names		
5.21	0.10	Correctol	formalin	Bellergal
5.18	0.39	Midamor	Cetamide	Blephamide
5.18	0.73	Nolamine	calamine	Alamine
5.90	0.11	Florinef	Fedahist	Beclovent
5.89	0.49	Lubriderm	Estraderm	Eryderm
5.81	0.72	Thorazine	norazine	Clorazine
4.47	0.16	Dihistine	filgrastim	Decaspray
4.47	0.36	Cataflam	Cotazym	Azactam
4.42	0.68	Prostigmin	Prolastin	progestin
5.39	0.11	Flaxedil	Cyclogyl	Cetapred
5.47	0.44	Iberet	Fibermed	Fibercon
5.40	0.69	Calcidrine	Dalcaine	Alcaine
5.79	0.16	Eldercaps	Aspercreme	Aldactone
5.72	0.36	Imuran	Iophen	Anuphen
5.74	0.62	Pertussin	Histussin	Detussin

Note. Brand names are capitalized in the tables to distinguish them from generic names. In the actual experiments, all names were capitalized.

Table 2

Experiment 1: Parameter estimates for ordinal logistic regression model predicting number of recall errors made by pharmacists (orthographic similarity)

Variable	Estimate	Stand. Error	Z
Intercept	7.47	1.42	5.24**
Similarity	-4.37	0.85	-5.12**
Log Frequency	-0.94	0.26	-3.68*
Familiarity	-0.64	0.15	-4.29**

* $p < 0.001$, ** $p < 0.0001$

-2 Log Likelihood = 464.59

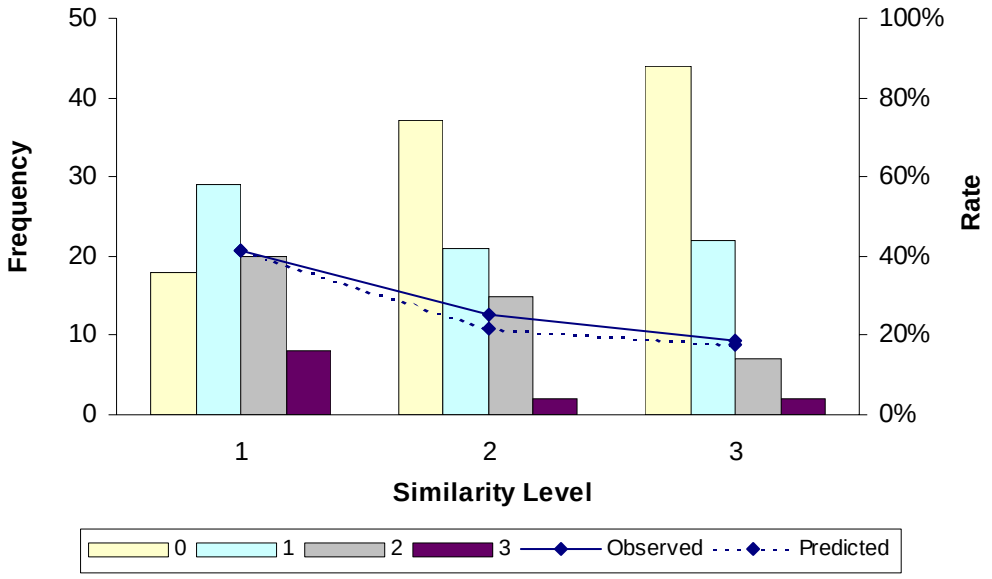

Figure 1. Errors in pharmacists' recall as a function of orthographic similarity. Any misspelling was coded as an error. Vertical bars represent frequency of responses with 0, 1, 2, or 3 errors. Trend lines illustrate observed and predicted trend in overall error rate (i.e., number of errors divided by total opportunities for error). Although similarity was a continuous variable in our analyses, we divided it into 3 ranges for the purposes of these illustrations. Levels 1 through 3 correspond to mean bigram similarity values of 0.13, 0.40, and 0.68 respectively.

Table 3

Stimulus words for recall experiments using phonological similarity measure (Experiments 2 and 4)

Log Frequency	Similarity	Names		
5.45	0.63	lincomycin	tobramycin	vancomycin
5.41	0.39	cimetidine	minoxidil	simethicone
5.44	0.21	clotrimazole	cytarabine	temazepam
4.75	0.58	carbidopa	levodopa	methyldopa
4.79	0.40	astemizole	indapamide	miconazole
4.78	0.21	adenosine	chlorzoxazone	nevirapine
4.89	0.58	thioguanine	thiotepa	thiothixene
4.91	0.39	methenamine	methimazole	metolazone
4.89	0.21	carbamazepine	clozapine	isradipine
5.01	0.54	famotidine	nizatidine	ranitidine
5.13	0.38	alprazolam	triazolam	trimethoprim
5.07	0.21	acetone	amiodarone	norfloxacin
5.39	0.51	fenoprofen	ketoprofen	metoprolol
5.39	0.36	chlorthalidone	piroxicam	risperidone
5.32	0.21	amoxapine	cefazolin	prednisolone

Note. All names are generic.

Table 4

Experiment 2: Parameter estimates for ordinal logistic regression model predicting number of recall errors made by pharmacists (phonological similarity)

Variable	Estimate	Stand. Error	Z
Intercept	8.04	6.23	1.28
Similarity	18.84	6.48	3.05*
Similarity^2	-29.07	8.23	-3.68**
Log Frequency	-1.77	1.18	-1.51
Familiarity	-0.80	0.20	-4.12***

* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

-2 Log Likelihood = 355.42

Note: Terms were kept in the model if likelihood ratio tests of their removal were significant. In some cases, likelihood ratio tests were significant at $\alpha = 0.05$, but z-scores (or Wald tests on the parameter estimates) were not. In such cases, likelihood ratio tests are believed to be more reliable, and hence they were used.[44, 46]

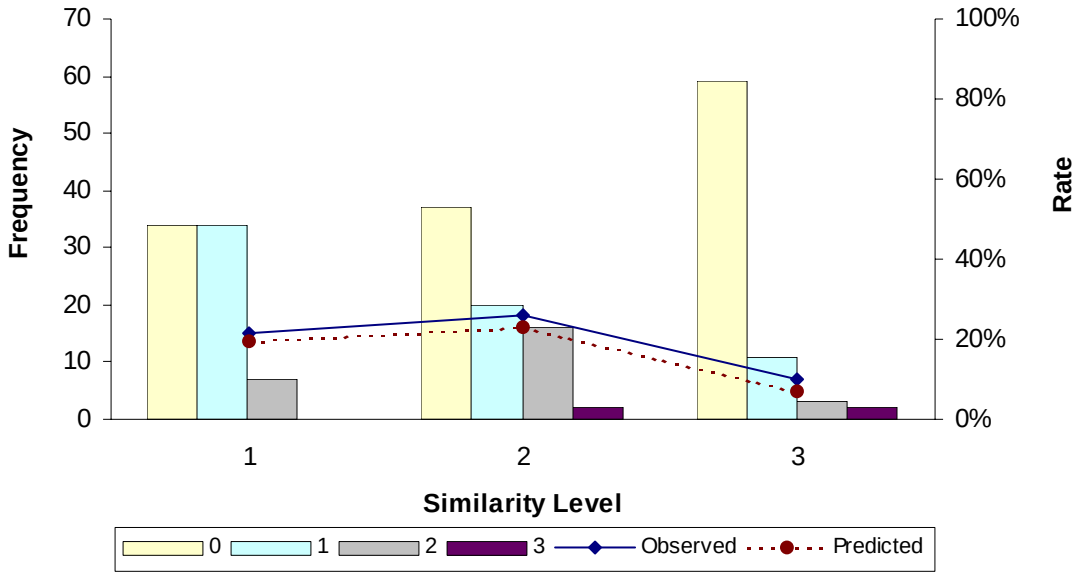

Figure 2. Errors in pharmacists' recall as a function of phonological similarity. Any misspelling was coded as an error. Vertical bars represent frequency of responses with 0, 1, 2, or 3 errors. Trend lines illustrate observed and predicted trend in overall error rate (i.e., number of errors divided by total opportunities for error). Although similarity was a continuous variable in our analyses, we divided it into 3 ranges for the purposes of these illustrations. Levels 1 through 3 correspond to mean bigram similarity values of 0.21, 0.38, 0.57 respectively.

Table 5

Experiment 3: Parameter estimates for ordinal logistic regression model predicting number of recall errors made by college students (orthographic similarity)

Variable	Estimate	Stand. Error	Z
Intercept	6.92	1.42	4.85***
Similarity	-3.16	0.57	-5.57***
Log Frequency	-0.64	0.21	-2.99*
Familiarity	-0.29	-0.08	-3.40**

* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

-2 log likelihood = 1157.65

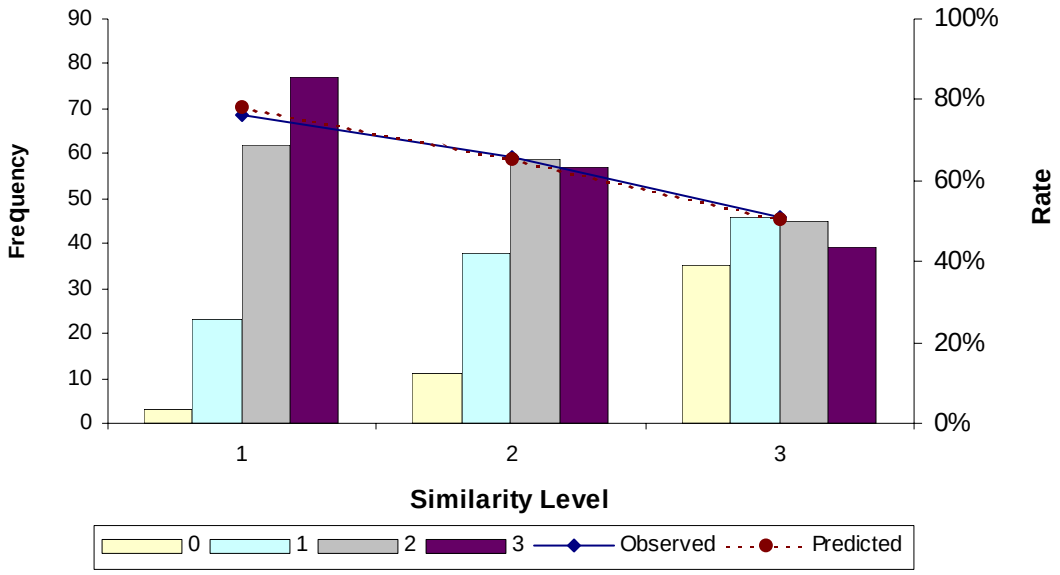

Figure 3. Errors in college students' recall as a function of orthographic similarity. Any misspelling was coded as an error. Vertical bars represent frequency of responses with 0, 1, 2, or 3 errors. Trend lines illustrate observed and predicted trend in overall error rate (i.e., number of errors divided by total opportunities for error). Although similarity was a continuous variable in our analyses, we divided it into 3 ranges for the purposes of these illustrations. Levels 1 through 3 correspond to mean bigram similarity values of 0.13, 0.40, and 0.68 respectively.

Table 6

Experiment 4: Parameter estimates for ordinal logistic regression model predicting number of recall errors made by college students (phonological similarity)

Variable	Estimate	Stand. Error	Z
Intercept	3.78	0.48	7.83**
Familiarity	-0.41	0.11	-3.72*

* $p < 0.001$, $p < 0.00001$

-2 log likelihood = 743.434

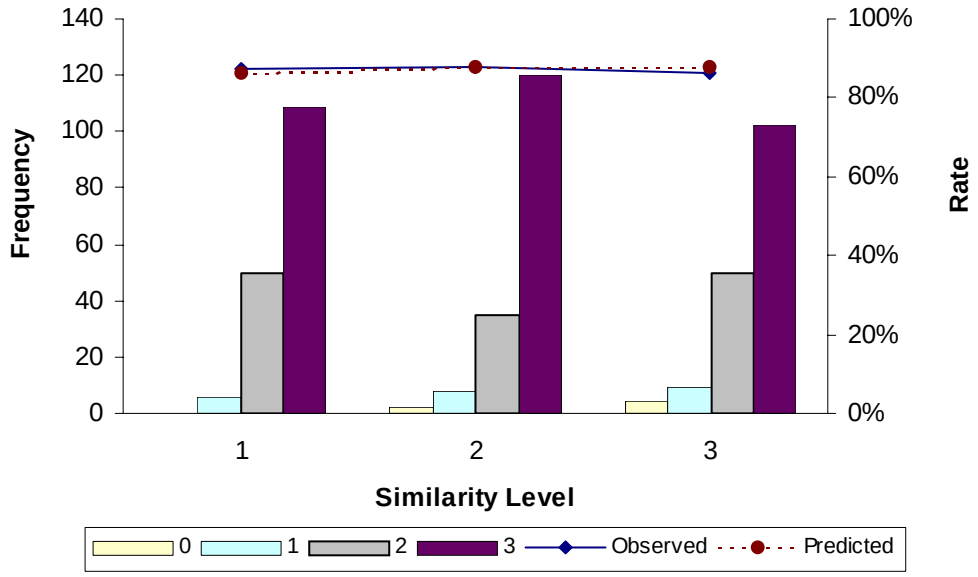

Figure 4. Errors in college students' recall as a function of phonological similarity. The trend lines illustrate the observed and predicted decline in the error rate. Although similarity was a continuous variable in our analyses, we have divided it into 3 ranges for the purposes of creating these illustrations. Levels 1 through 3 corresponded to mean bigram similarity values of 0.21, 0.38, 0.57 respectively.