On the norming of homophones

DAVID S. GORFEIN

Adelphi University, Garden City, New York

AND

KRISTIN M. WEINGARTNER

Hofstra University, Hempstead, New York

Associative norms for homographs have been widely used in the study of language processing. A number of sets of these are available, providing the investigator with the opportunity to compare materials collected over a span of years and a range of locations. Words that are homophonic but not homographic have been used to address a variety of questions in memory as well as in language processing. However, a paucity of normative data are available for these materials, especially with respect to responses to the spoken form of the homophone. This article provides such data for a sample of 207 homophones across four different tasks, both visual and auditory, and examines how well the present measures correlate with each other and with those of other investigators. The finding that these measures can account for a considerable proportion of the variance in the lexical decision and naming data from the English Lexicon Project provides an additional demonstration of their utility. The norms from this study are available online in the Psychonomic Society Archive of Norms, Stimuli, and Data, at www.psychonomic.org/archive.

Ambiguous words have been employed in the study of a wide range of cognitive phenomena, including, but not limited to, meaning selection with lexical ambiguity (Bubka & Gorfein, 1989; Gorfein, 2001; Gorfein, Brown, & DeBiasi, 2007; Gorfein & Walters, 1989; Simpson, 1984; Simpson & Kang, 1994), parafoveal preview effects during reading (Chace, Rayner, & Well, 2005; Polatsek, Lesch, Morris, & Rayner, 1992), meaning stability (Britton, 1976; Winograd & Geis, 1974), implicit memory (Eich, 1984; Jacoby & Witherspoon, 1982), and spacing effects in free recall (Gartman & Johnson, 1972). Investigations that employ ambiguous words typically generate predictions on the basis of which associated meaning is dominant, a fact that is particularly relevant to the present article. In these cases, investigators often rely on normative data generated by word association tests to determine the relative dominance among alternative meanings of an ambiguous word.

A multiplicity of norms exist for the meanings of homographs—that is, of words with the same spelling and distinct meanings (Cramer, 1970; Geis & Winograd, 1974; Gorfein, Viviani, & Leddo, 1982; Kausler & Kollasch, 1970, 1973; Nelson, McEvoy, Walling, & Wheeler, 1980; Perfetti, Lindsay, & Garson, 1971; Twilley, Dixon, Taylor, & Clark, 1994; Yates, 1978). Typically, these norms are derived from responses to written word association tests in which the participant is asked to write the first word that comes to mind in response to a printed word. Because nothing about the printed word in isolation distinguishes

among its alternative meanings, the meaning provided by the most responders is assumed to be the dominant one.

The focus of the present article is on words that have the same pronunciation but distinct spellings, such as *sun–son*, *night–knight*, and so on. These words have been labeled *homophones* in the word processing literature, although, with few exceptions, most homographs are homophonic as well. We will follow that distinction in this article, using the word *homograph* when we refer to an ambiguous word of one spelling and *homophone* when the ambiguity is present in the spoken form only. (Note that some homophones may be homographic as well. For example, *fair–fare* is a case in which both homophones are homographs: *Fair* means “carnival” as well as “just,” and *fare* means “transportation cost” as well as “kind of food served.”)

Despite the substantial use of homophones in the word processing literature, as compared with homographs, fewer studies have been devoted to the norming of homophones. Notably, most nonhomographic homophones are unambiguous in print. Given this situation, a number of investigations have employed measures of frequency of occurrence in print (e.g., Francis & Kučera, 1982; Kučera & Francis, 1967). However, as demonstrated by Kreuz (1987), this approach can be misleading. Kreuz compared homophone frequencies in print, derived from the Kučera and Francis counts, with rated familiarity (from *very common* to *very uncommon*) as predictors of performance in the lexical decision task. As in earlier work with homographs (Gerns-

D. S. Gorfein, dgorfein@yahoo.com

bacher, 1984), rated familiarity was a significantly better predictor of lexical decision times for homophones.

Beyond rated familiarity, normative data exist for a small number of additional subjective measures of homophones. Recently, White and Abrams (2004) used two measures for 96 homophone sets. The first measure was a written word association task, and the second was a dominance rating task, in which participants were presented with homophone pairs (or in some cases, triplets) and asked to choose the individual word they felt to be more dominant (or to choose "equal" if they considered the words to be equally dominant). The results indicated no significant correlations between responses for the two measures. However, for both measures, there was some consistency between the responses from young adults (17–25 years old) and old adults (62–91 years old). The White and Abrams norms should be particularly useful for those who choose to undertake cognitive aging studies, in that the authors identified a subset of homophones that are well matched across young and old adults for both measures.

Whereas White and Abrams (2004) collected association responses to visually presented homophones, other researchers have used auditory presentation. One advantage of using auditory presentation is that one can obtain association responses simultaneously with a second indicator of homophone dominance: spelling. Galbraith and Taschman (1969) used both measures and found them to be highly correlated ($r = .87$). Gorfein and O'Brien (1985) retested the Galbraith and Taschman words, but they replaced the association task with a "cued" spelling condition. Gorfein and O'Brien reasoned that the written word response in the spelling task might be ambiguous with respect to meaning if the participant either spelled both meanings of a homophone in the same manner (e.g., if the participant thought that *belle*, meaning "Southern beauty," and *bell*, meaning "chime," were both spelled *bell*) or had confused the spellings and meanings. In the cued spelling condition, participants were instructed that they would hear a word (e.g., *son–sun*) and that they would find on an answer sheet a printed word related to the word they heard (e.g., *tan*). Participants were asked to write the correct spelling of the word they heard beside the printed cue on the answer sheet. The responses were scored for correctness with respect to the definition of the cue. Consistent with the purpose of the methodology, it was discovered that participants were not always able to supply the correct spelling for the auditory homophone–written cue combination.

One purpose of the present study was to compare subjective normative responses for homophones presented either aurally or in written form. To do this, we used five measures of homophones: spelling under two conditions of auditory presentation (cued, as in Gorfein & O'Brien [1985], and uncued); word association with the written form of the homophone; a sentence comprehension task, in which participants had to decide whether the written form of a homophone made sense, given either an appropriate or an inappropriate sentence context; and frequency in print (Kučera & Francis, 1967). We will present the results separately, with a minimum of discussion, and then

we will examine the intercorrelations among the measures and comment on their potential utility for research.

EXPERIMENT 1 Spelling Tasks

Experiment 1 was a follow-up to the Gorfein and O'Brien (1985) norms and was designed to examine the spelling of aurally presented homophones under two conditions: cued and uncued.

Method

Participants. A total of 240 students enrolled in introductory psychology courses at the University of Texas, Arlington, participated as partial fulfillment of a course requirement. All were selected to meet a stringent criterion that they be native English speakers—they had to indicate on a questionnaire that English was spoken at home at least 95% of the time during their preschool years.

Materials. The materials were 207 homophone sets selected from a variety of sources. Of these sets, 81 had appeared in both the Galbraith and Taschman (1969) and the Gorfein and O'Brien (1985) norms. The other 126 were selected on the basis of appearing in one or more studies in the homophone literature. Several of the sets contained more than two alternative spellings. The homophones were separated into two lists in order to obtain the number of observations necessary to complete the design. Data collection for the second list occurred approximately 1 year following data collection for the first. For each list, the homophones were randomly ordered and assigned a number. Both the number and the corresponding word were read aloud in a male voice at a rate of one number–word pair every 7 sec. These readings were tape-recorded for use during norm collection. Three separate answer sheets were created: Cued Meaning A, Cued Meaning B, and uncued. For both of the cued answer sheets, the cue appeared in parentheses following the number corresponding to the appropriate number–word pair from the tape recording. For example, if [sʌn] was spoken following the number "three" on the tape recording, then *father* was printed following the number 3 on the Cued Meaning A answer sheet, and *tan* was printed following that number on the Cued Meaning B answer sheet. For both cued answer sheets, the cue was followed by a blank line on which participants were asked to print a response. For homophones that had more than two spellings, the aural form was included in both lists in order to evaluate the effects of a cue on third and fourth meanings: *pair* (*pear*, *pare*), *you* (*ewe*, *yew*), *vain* (*vein*, *vane*), *sight* (*cite*, *site*), and *air* (*heir*, *ere*, *err*). For example, for List 1, a cue related to *air* appeared on one answer sheet, and a cue related to *heir* appeared on the other; for List 2, a cue related to *ere* appeared on one answer sheet, and a cue related to *err* appeared on the other. The uncued answer sheet contained only numbers (corresponding to the number–word pairs from the tape recording) followed by a blank line.

Procedure. The participants were tested in groups ranging in size from 10 to 23. Within each group, the three different answer sheets were randomly assigned to participants, with the restriction that no answer sheet was used for the $N+1$ time until all had been used N times. Forty participants served in each condition (Cued Meaning A, Cued Meaning B, and uncued) for each of the two homophone lists.

All participants were informed that they were part of a norming study and that they would hear words that had more than one spelling. We chose to state the nature of the materials to avoid having different participants "catch on" at different points in the experiment. Specific instructions for the experiment appeared on the cover page of each answer sheet. Participants in the uncued task were instructed to listen to the spoken words and write the first spelling of the word that came to mind. These individuals received the uncued answer sheet, which contained each number followed by a blank line for the answer. Participants in the cued task were instructed to write

the word they heard that corresponded to or was associated with the written cue. These individuals received either the Cued Meaning A or Cued Meaning B answer sheet. After all of the participants within a group had read the instructions, the appropriate tape recording was played, and the participants printed on the answer sheet their spelling responses to each word.

Results and Discussion

We begin with a discussion of the results from the uncued task. For each alternative form of a homophone, a score was computed that represented the proportion of the participants who provided that particular spelling. Depending on the number of alternative forms provided across all participants, each homophone had anywhere from two to four scores associated with it. These scores appear in online Appendix A. The assignment of scores to individual columns was based in part on whether the word forms provided had later been cued for in the cued task. That is, the scores for word forms that were later cued appear under the headings "First" and "Second" in order of dominance (e.g., the spelling adopted by the most responders appears under the heading "First"). Note that the same convention applies to the four homophone sets in which the effects of a cue were evaluated for a third (five sets) and fourth (one set) meaning. Thus, for those five sets, the alternative word forms provided in the uncued task appear under the headings "First," "Second," "Third," and (for one set) "Fourth." Because the auditory forms of these five homophones appeared in both of the auditory lists, the scores for these homophones in Appendix A reflect an average of the proportions obtained from the two lists. For several homophone sets, a third word form was provided that had not been later cued for in the cued task. The scores for these forms appear under the heading "Third" in the Appendix. Finally, homophone sets are listed in alphabetic order of the dominant spelling. This order of presentation is preserved throughout the results for the subsequent studies. The reader should be alerted to the fact that, for both of the tasks in the present experiment and for the subsequent experiments, proportions for a few homophone sets are missing because of a faulty recording of the stimuli or because the responses were difficult to unambiguously code. To maintain consistency across the tables for all of the experiments, all homophones are listed in all of the tables, and a dash ("—") appears in place of any nonexistent proportions.

As we suggested in the introduction, an issue of interest is how well frequency in print (Kučera & Francis, 1967) corresponds to meaning dominance. This is particularly of interest when the homophone is presented in its auditory mode. At the most conservative level, we would predict that the word form with the highest frequency in print will

in general be dominant. This was the case for 140 of the 206 coded homophone sets, whereas a lower-frequency word form was dominant for 58 homophone sets (for 8 of the sets, the alternative spellings had equal print frequencies). Similar results were obtained with frequencies drawn from the HAL database (Balota et al., 2007).¹ The higher-frequency homophone was dominant for 142 of the 206 coded homophone sets, whereas the less-frequent homophone was dominant in the spelling norms 64 times.

For the 81 homophone sets common to this study, Galbraith and Taschman (1969), and Gorfein and O'Brien (1985), we computed product-moment correlations between the proportion scores for the dominant forms obtained in the present experiment and in (1) the Gorfein–O'Brien spelling task, (2) the Galbraith–Taschman spelling task, and (3) the Galbraith–Taschman word association task. These correlations are reported in Table 1. For ease of discussion, three additional correlations are reported: the correlations between the proportion scores for the dominant forms obtained in the Gorfein–O'Brien spelling task and each of the Galbraith–Taschman tasks, as well as the correlation between the scores obtained in each of the Galbraith–Taschman tasks. The present sample correlates more poorly with the Galbraith–Taschman word association scores than do the Gorfein–O'Brien and Galbraith–Taschman spelling norms. The correlations among the spelling scores are reasonably high, although in some cases there are large differences among the samples. This is not too surprising, given that the three studies are separated both by different geographical locations and, perhaps more crucially, by 30+ years.

We next turn to the results for the cued task. The primary goal of this task was to examine the degree to which participants could correctly spell an aurally presented homophone when provided with a semantically related printed cue. The responses were scored for correctness with respect to the meaning of the cue, and a score was also computed that represents the proportion of times the item was spelled in the direction of an alternative meaning. Online Appendix B contains the proportion of correctly spelled words under each cue condition, the proportion of times an alternative spelling of the word was given, and the proportion of times participants provided a spelling that did not match any form in the homophone set (i.e., "Other").

We predicted that there would be a good deal less competition among the alternative spellings in the cued task than in the uncued task. However, given the conclusions of Gorfein and O'Brien (1985), we expected participants to produce inappropriate-to-cue spellings on some occasions, either because they thought the cued meaning was spelled that way or because they were unfamiliar with one

Table 1
Intercorrelations Among Four Measures of Homophone Dominance

	Gorfein–O'Brien Spelling (Uncued)	Galbraith–Taschman Spelling (Uncued)	Galbraith–Taschman Word Association
Present norms spelling (uncued)	.682	.616	.513
Gorfein–O'Brien spelling (uncued)		.679	.679
Galbraith–Taschman spelling (uncued)			.722

meaning, and therefore used the only spelling that they knew. Inspection of Appendix B indicates that the mean proportions of correct spellings in the cued conditions were .833 for the dominant cue and .615 for the secondary cue. The proportions of alternative spellings were .061 for the dominant cue and .204 for the secondary cue. The .143 difference between the incorrect spellings (.204 – .061) suggests that the mean dominance score in the uncued condition is an overestimate, since it contains a proportion of responses that resulted from participants' thinking of a secondary word and then misspelling it.

EXPERIMENT 2

Word Association to the Printed Form of the Homophone

In the second experiment, word associations to the printed forms of the homophones from Experiment 1 were collected. To the extent that participants knew the differences among the spellings of the alternative meanings for the homophones, the associates provided for one spelling of a homophone would overlap little with those provided for an alternative spelling. In other words, there should be few meaning intrusions across alternative spellings.

Method

Participants. A total of 200 students enrolled in introductory psychology courses at the University of Texas, Arlington, participated as partial fulfillment of a course requirement. All were native English speakers (determined by the same questionnaire that was used in Experiment 1), and none had participated in Experiment 1.

Materials. The 207 homophone sets from the spelling tasks in Experiment 1 were used in this experiment. However, to prevent the experimental session from being too lengthy, and to accommodate the five homophone sets for which we examined more than two alternative spellings (*pair* [pear, pare], *you* [ewe, yew], *vain* [vein, vane], *sight* [cite, site], and *air* [heir, ere, err]), the words were randomly assigned to four different lists, with the constraint that only one of the spellings for each homophone set appeared on a given list. Items within a list were also randomized. All participants were tested in the second year of the project and were randomly assigned to one of the four lists, with the constraint that 50 participants served in each list.

Procedure. The participants were seated before a computer monitor and were told that they would see a number of trials on which a word preceded by a trial number would be presented. For each trial, they were instructed to write the first word that came to mind next to the trial number on the answer sheet. Items were presented centered on the 13th line of the computer monitor and remained there until the next item was presented. The intertrial interval was 7 sec, and participants were told at the beginning of the experiment that they would be given 7 sec to provide an answer for each trial. After the experimenter was sure that the instructions were understood, the participant was left alone to complete the word association trials.

Results and Discussion

Responses to the alternative forms in each homophone set were scored with respect to their appropriateness in meaning. For each form, responses were classified as belonging to one of three categories: related to the presented form ("correct"), related to an alternative spelling of the homophone ("alternative"), or "neither." The "neither" response category included omissions, rhyme associates

that were unrelated in meaning (e.g., *dead* as an associate of *bread*), and another spelling as the associate (e.g., *bred* as the associate to *bread*). Be aware that a number of the homophones presented were homographic for one or both spellings. For the most part, associates were relatively easy to classify, because the alternative forms were quite distinct in meaning. In a few instances, we scored a word as appropriate to the homophone presented, even though it might also have been a strong associate to another member of a homophone set. For example, both of the words in the *mail–male* set elicited the response *man* from some of the participants. Our scoring therefore was biased toward classifying responses as appropriate.

Online Appendix C presents the proportions of correct, alternative, and other responses for each homophone set. For the dominant spelling, the mean proportion of appropriate (correct) responses was .89, whereas the mean proportion of appropriate responses to the second spelling was .69. Consistent with this difference was the large proportion of alternative responses in general. However, a closer examination of Appendix C indicates that the distribution of alternative responses is highly skewed. Whereas, on average, word associations to the alternative meaning occurred 9.2% of the time, the median rate of such responses was only 4.0%.

EXPERIMENT 3

Editing of Homophones in a Sentence Context

The goal of the third experiment was to examine the influence of sentence context on the processing of visually presented homophones. To accomplish this, we compared sensibility judgments for sentences that contained a contextually appropriate word form from a homophone set with sentences that contained the inappropriate word form. To the extent that the participants as a group knew the meaning of a particular spelling of a homophone, and thus could correctly accept the use of that spelling in an appropriate sentence context (e.g., "Two of each animal went aboard the *ark*"), we predicted that they would (1) reject the use of that spelling in a different, inappropriate context (e.g., "The *ark* of the circle was wide") and (2) reject the use of an alternative spelling in the context appropriate to the original spelling (e.g., "Two of each animal went aboard the *arc*").

The design of the present experiment was similar to one previously used to investigate the role of phonological encoding during reading for children (Doctor & Coltheart, 1980). The main conclusion from that prior research was that although very young children rely on phonological encoding, this reliance decreases with age; older children found sentence frames containing inappropriate homophones less acceptable than did young children. A major difference between the Doctor and Coltheart study and the present one is that their participants were pretested to ensure that every one knew the meaning of each form of the homophones employed. To provide an additional normative baseline, this participant selection criterion was eliminated in the present experiment. Excluding this baseline knowledge requirement provides a test of the predictive power of the other normative measures (spelling, cued

spelling, word association, and frequency in print), albeit the test is limited by the choice of sentences.

Method

Participants. A total of 200 students enrolled in introductory psychology courses at the University of Texas, Arlington, participated as partial fulfillment of a course requirement. All were native English speakers (determined by the same questionnaire used in Experiments 1 and 2), and none of them had participated in Experiment 1 or 2.

Materials. The 207 homophone sets from Experiments 1 and 2 were used in this experiment. For each of the spellings of the homophones, a sentence was constructed that was appropriate only to that spelling. Each of these sentences had a corresponding "inappropriate" sentence that was generated by replacing the correctly spelled homophone by an alternatively spelled homophone (e.g., "I'd like *some* more cake" vs. "I'd like *sum* more cake").² The sentence frames were randomly divided into four booklets, with the constraint that the same sentence frame and the same word form from a homophone set (e.g., *some* or *sum*) did not occur within the same booklet. For each booklet, the numbers of sentences containing appropriate and inappropriate word forms were equated.

To minimize the likelihood that participants would discover the purpose of the study, we created 60 additional filler items that contained 30 correct sentences and 30 incorrect sentences. The correct sentences all included a homograph (e.g., "As an actor, he was an awful *ham*"), whereas the incorrect sentences contained erroneous words (e.g., "The circulatory system has tiny blood vessels called *capulets*"). The same set of 60 filler sentences was included in each booklet. In addition to obscuring the manipulation, these filler sentences provided a means of determining whether the four groups of participants (one group per booklet) were equivalent in their baseline sensibility judgments. Finally, the experimental and filler sentences within each booklet were randomized, and then each was assigned a number.

Procedure. The participants were tested in small groups in a classroom setting. The four test booklets were distributed randomly, with the restriction that no booklet type was given out $N+1$ times until all four booklets had been assigned to N participants. Participants were instructed to read each sentence to determine whether it made sense. If participants thought a sentence did not make sense, they were asked to circle the number in front of that sentence, as well as one word within the sentence that made the sentence nonsensical. This enabled the investigator to determine whether the participant was identifying the correct problem or demonstrating some idiosyncratic judgment of sensibility.

Results and Discussion

We first examined performance on the filler items to determine whether participants from each of the four groups (one per booklet) were equivalent in terms of baseline sensibility judgments. Correct acceptances averaged .762, and the means for Groups 1–4 were .755, .781, .759, and .754, respectively. Correct rejections averaged .552, and the group means were .578, .529, .553, and .550, respectively. There were no significant differences among the group means for correct acceptances or correct rejections. Thus, we can be confident that the following analysis of the homophone sentences was not unduly influenced by a group bias.

The correct acceptance scores and correct rejection scores for the homophone sentences are reported in online Appendix D. Scores for correct acceptances were calculated by dividing the number of appropriate sentences that had not been circled by the total number of participants.

Correct rejections occurred when both an inappropriate sentence and its corresponding inappropriate homophone had been circled. Correct rejection scores were thus calculated by dividing the total number of correct rejections by the total number of participants.

As stated above, there were two main predictions: If on average the participants knew the meaning of a particular spelling of a homophone, and thus correctly accepted the use of that spelling in an appropriate sentence context (e.g., "Two of each animal went aboard the *ark*"), they should have (1) rejected the use of that spelling in a different, inappropriate context (e.g., "The *ark* of the circle was wide") and (2) rejected the use of an alternative spelling in the context appropriate to the original spelling (e.g., "Two of each animal went aboard the *arc*"). In other words, there should have been a positive correlation between the correct acceptance scores and each of the two types of correct rejection scores. Although correct acceptance scores were not correlated with correct rejection scores when the same spelling was used in a different context [$r(400) = .04$], they were when a different spelling was used in the same context [$r(400) = .40$]. Thus, even though participants are sensitive to the contextual appropriateness of alternative forms of a homophone in the same sentence context, this does not necessarily mean that they are sensitive to the contextual appropriateness of using the same form of a homophone across alternative sentence contexts. More generally, these findings suggest that sentence context contributes to sensibility judgments when participants are aware of the distinctions in meaning among alternative spellings of a homophone, but that it contributes little when they are not aware of this distinction.

INTERCORRELATIONS AMONG HOMOPHONE MEASURES

We next report the intercorrelations among the measures of homophone dominance discussed above. The measures included were (1) proportion of responses from the uncued spelling task that reflected the more common (i.e., dominant) meaning of an aurally presented homophone (Experiment 1); (2) proportion of responses from the cued spelling task that were appropriate to the meaning of a visually presented cue for an aurally presented homophone (Experiment 1); (3) proportion of word associate responses that were appropriate to the meaning of the visually presented form of a homophone (Experiment 2); (4) proportion of correct acceptances for sentences that contained a contextually appropriate form of a visually presented homophone (Experiment 3); and finally, two measures of frequency in print: (5) the natural log of the frequency in print (Kučera & Francis, 1967) of a form of the homophone and (6) the natural log of the frequency in the HAL database (Balota et al., 2007). The intercorrelations among these measures are presented in Table 2.³

Inspection of Table 2 indicates that the word association responses were highly correlated with both the cued spelling responses ($r = .78$) and the correct acceptances ($r = .77$). The latter two measures were also highly correlated ($r = .74$). Furthermore, these three measures were

Table 2
Intercorrelations Among Measures of Homophone Dominance From Experiments 1–3 and Two Measures of Frequency in Print

	Cued Spell	Word Association	Sentence Acceptability	Ln K–F Freq.	Ln HAL Freq.
Uncued spell	.506	.494	.410	.324	.345
Cued spell		.776	.741	.514	.529
Word association			.774	.449	.501
Sentence acceptability				.440	.502
Ln K–F freq.					.759

Note—K–F, Kučera and Francis (1967).

more strongly correlated with each other than any one of them was with the uncued spelling measure. Perhaps the most important observation to note is that neither of the print frequency measures was as highly correlated with the other measures as the latter measures were with each other. Nevertheless, when we conducted a regression analysis attempting to predict correct acceptance scores for the correct acceptances from the sentence task, we found that print frequency did contribute significantly as a predictor variable, raising the amount of variance accounted for by the word association scores from .60 to .65.

We also examined the correlations between the measures from the present study and the White and Abrams (2004) word association norms for the comparable age group. The two studies have 82 words in common. As shown in Table 3, the word association scores from the White and Abrams sample correlated more highly with the present uncued spelling norms (.72) than did the word association scores in the present sample (.56). However, the White and Abrams word association scores were less successful in predicting performance on the sentence criteria (.43) than were the word association scores in the present sample (.54).

In the previous description of our study, the reader may have noted that no mention has been made of the reliability of the norms. This is not uncommon with respect to norms for ambiguous words, because it is inherently difficult to define reliability with respect to word associations or spellings of ambiguous words. Most crucially, the response to the word might reflect any of the following possibilities: a participant might have been aware that the stimuli had more than two possible meanings and/or spellings, might have misread (or misheard) the stimulus item, or might have failed to answer. These alternatives do not exist on any obvious continuum, such that a score could be assigned for the purpose of computing a cor-

relation. Performance across words should also be relatively dependent on an individual's lifetime experience, including age of acquisition and any recent exposure to particular words included in the stimulus set (Gorfein, 2001; Gorfein & Brown, 2007; Simpson & Kang, 1994). As such, many investigators have adopted the procedure of reporting correlations of their items' dominance scores with the results from other samples in the literature, as we have done in Tables 1 and 3. These comparisons indicate a modest degree of correlation across geographic locations and times. In the one task that could be scored as correct and incorrect (sentence sensibility judgments), we computed Cronbach's alpha for each of the four test lists, and all values obtained exceeded .89. It is mathematically true that significant correlations among variables cannot exist in the absence of reliability of the measures.

To test the validity of the present norms, we examined the relationship between the present measures (cued and uncued spelling, word association, and sentence acceptability) and three measures provided in the English Lexicon Project database (Balota et al., 2007): lexical decision accuracy, lexical decision reaction time, and naming time. Using multiple regression, we partialled out word length, orthographic neighborhood size (orthographic *N*), and HAL frequency (all obtained from the English Lexicon Project database) from the present measures. The resulting intercorrelations are reported in Table 4. Note that, although the intercorrelations among the present measures are slightly reduced when word length, orthographic *N*, and HAL frequency are partialled out, the patterns are consistent with those reported in Table 2, where these lexical variables were not statistically controlled for. In particular, the word association, cued spelling, and correct acceptance scores are strongly correlated. Furthermore, these three measures are more strongly correlated with each other than any one of them is with the uncued spelling scores.

Table 3
Intercorrelations Among Measures From Experiments 1–3 and Word Association for the Young Group in White and Abrams (2004)

	Cued Spell	Word Association	Sentence Acceptability	White–Abrams (Young)
Uncued spell	.597	.563	.421	.715
Cued spell		.699	.659	.652
Word association			.539	.586
Sentence acceptability				.428

Note—Only the 82 homophones that were common to the present experiments and White and Abrams (2004) were included in the analyses.

Table 4
Intercorrelations Among Measures of Homophone Dominance From Experiments 1–3 and
Lexical Decision and Naming Data From the English Lexicon Project

	Cued Spell	Word Association	Sentence Acceptability	Lexical Decision RT	Lexical Decision Accuracy	Naming
Uncued spell	.388	.396	.270	–.329	.323	–.174
Cued spell		.684	.631	–.618	.618	–.348
Word association			.680	–.518	.605	–.244
Sentence acceptability				–.503	.671	–.317
Lexical decision RT					–.662	.456
Lexical decision accuracy						–.329

Note—Ln HAL frequency, orthographic *N*, and word length were partialled out from the analyses. RT, reaction time.

We next consider the power of the present variables in predicting the lexical decision and naming time data from the English Lexicon Project (Balota et al., 2007). Table 5 contains the proportions of variance accounted for by each of the present measures: cued spelling, sentence acceptability, word association, and uncued spelling (once again, with word length, orthographic *N*, and HAL frequency partialled out from the analysis). As illustrated in Table 5, the present variables appear to be successful at predicting the lexical decision data, accounting for .39 of the variance in the accuracy data and .28 of the variance in the reaction time data, although they are less successful at predicting the naming time data, accounting for only .11 of the variance. Assuming that the lexical decision task places a greater emphasis on the meaning of the stimulus (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), the superior predictive power of our measures for the lexical decision data relative to the naming time data is not surprising. An additional noteworthy pattern from Table 5 is that the cued spelling, sentence acceptability, and word association measures all appear to be stronger predictors than the uncued spelling measure. This pattern is evident with the lexical decision accuracy and reaction time data, and also with the naming time data. Thus, the same measures that are strongly correlated with each other (see Tables 2 and 4) also appear to be the strongest predictors of the lexical decision and naming time data.

GENERAL DISCUSSION

A primary purpose of normative studies of ambiguity has been to provide the investigator with knowledge

of the materials that will lead to improved research designs. The five dominance measures described here provide this information for 207 homophone sets. Choosing an appropriate measure of homophone dominance will obviously depend on investigators' goals and on design considerations. For example, whereas the norming data from Experiment 1 will be particularly useful when the homophones are presented in the auditory domain, the data from Experiments 2 and 3 will be more appropriate when the printed form is presented. Furthermore, investigators who intend to examine the influence of context on homophone processing would do well to focus on the data from the cued spelling task in Experiment 1 and the sentence comprehension task in Experiment 3.

Although the design issues mentioned above should inform investigators in their selection of a dominance measure, we urge investigators to also consider the intercorrelations among the measures reported throughout this article. In particular, although print frequency measures have often been employed as a measure of dominance, the intercorrelations reported in Table 2 suggest that this approach can be seriously misleading, because the print frequency measures are not as highly correlated with the present measures as the latter measures are with each other. The subjective familiarity judgments reported by Kreuz (1987) fare much better as predictors of dominance. For the 66 homophone pairs common to the studies of Galbraith and Taschman (1969), Gorfain and O'Brien (1985), Kreuz (1987), and the present study, the obtained correlation of the ratio of the rated familiarity measure for each of the two forms of a homophone obtained by Kreuz (dominant familiarity/secondary familiarity) and the aver-

Table 5
Proportions of Variance Accounted For by the Present Measures in Lexical Decision and
Naming Data From the English Lexicon Project

<i>R</i> ²	Lexical Decision Accuracy	Lexical Decision RT	Naming Time
Cued spelling, sentence acceptability, word association, and uncued spelling	.39	.28	.11
Cued spelling alone	.29	.26	.09
Sentence acceptability alone	.33	.16	.07
Word association alone	.27	.18	.05
Uncued spelling alone	.08	.08	.02

Note—Ln HAL frequency, orthographic *N*, and word length were partialled out from the analyses. RT, reaction time.

age dominance for the uncued spelling tasks (Galbraith & Taschman, 1969; Gorfein & O'Brien, 1985; and the present study) is .59. Notably, the correlation using Kučera and Francis (1967) frequencies is .29. Investigators employing homophones might consider the subjective familiarity ratings as a substitute for printed word frequencies when selecting or counterbalancing items.

For the purposes of selecting the most sensitive measure of homophone dominance, perhaps the best course of action is to consider multiple measures. In particular, it is instructive to examine the proportions of errors that occurred in the cued spelling task (aurally presented homophone with a visual cue to meaning) alongside the proportions of errors that occurred in the free association task (visually presented homophone). For many homophone sets, the proportion of errors is consistent across the two measures. For example, .50 of the sample produced *flare* as the spelling in response to the visually presented cue *knack* and the aurally presented homophone. Likewise, as indicated in Appendix C, .40 of the sample produced a word that was consistent with *flare* when presented with the visual form *flair*. In stark contrast, note that in the cued spelling task, .50 of the sample produced *ball* when the cue was *cry*, but that in the word association task, only .02 of the sample (1 participant) produced a word that was consistent with *ball* when the cue was *bawl*. The important point is that for some homophone sets, although a given proportion from the cued spelling task might in and of itself indicate that many participants are not familiar with a low-frequency form, the corresponding proportion obtained from the word association task might suggest otherwise.

As with all studies of word identification, when selecting materials for studies of homophone processing, investigators should take into consideration the time period in which the norming data were collected. One reason for this is that over time, one or more forms of a homophone can become homographic for a new meaning. A comparison of the older norms (Galbraith & Taschman, 1969; Gorfein & O'Brien, 1985) with the present norms provides at least two demonstrations of this. First, whereas *rap* was essentially unambiguous for the meaning "knock" in the older norms, it is unambiguous for a new meaning ("type of music") in the present norms. And second, whereas *phone* was not an associate to *cell* in the older norms, it is the dominant associate now. On a related topic, for some homophones it is reasonable to expect that the time of year during which the norming study was conducted could alter the baseline relative dominance among alternative forms. This could account for the unexpectedly high proportion with which participants provided the spelling *claus* (.50) in the uncued task of Experiment 1, which was conducted just before Christmas break.

In conclusion, the present study indicates that the same measure employed to obtain norms for homographs (i.e., word association to the printed form of the word) is also useful for nonhomographic homophones. Thus, investigators should find the norms supplied by White and Abrams (2004), as well as those presented here, useful. Finally, the test of any set of norms is its utility. The Gorfein and O'Brien (1985) norms have served one of the authors'

purposes in selecting the words for a number of studies (Gorfein et al., 2007; Gorfein & DeBiasi, 1997; Gorfein & Walters, 1989). It is our hope that others will find these more extensive norms useful as well.

AUTHOR NOTE

The authors acknowledge the contribution of Mona Lisa Byers, Larry Carter, Kenya Malcolm, Melody Paulk, and Anthony Pedrazine to the data collection and scoring of this project. We thank James Erickson for the use of laboratory facilities and statistical advice, as well as Vincent Brown and Ira Bernstein for statistical help. D.S.G. is professor emeritus from Adelphi University. Correspondence relating to this article may be sent to D. S. Gorfein, 3804 Summercrest Drive, Fort Worth, TX 76109 (e-mail: dgorfein@yahoo.com).

REFERENCES

- BALOTA, D. A., CORTESE, M. J., SERGENT-MARSHALL, S. D., SPIELER, D. H., & YAP, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, **133**, 283-316.
- BALOTA, D. A., YAP, M. J., CORTESE, M. J., HUTCHISON, K. A., KESSLER, B., LOFTIS, B., ET AL. (2007). The English Lexicon Project. *Behavior Research Methods*, **39**, 445-459.
- BRITTON, B. K. (1976). Semantic encoding stability and context. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 69-75.
- BUBKA, A., & GORFEIN, D. S. (1989). Resolving semantic ambiguity: An introduction. In D. S. Gorfein (Ed.), *Resolving semantic ambiguity* (pp. 3-12). New York: Springer.
- CHACE, K. H., RAYNER, K., & WELL, A. D. (2005). Eye movements and phonological parafoveal preview: Effects of reading skill. *Canadian Journal of Experimental Psychology*, **59**, 209-217.
- CRAMER, P. (1970). A study of homographs. In L. Postman & G. Keppel (Eds.), *Norms of word association* (pp. 361-382). New York: Academic Press.
- DOCTOR, E. A., & COLTHEART, M. (1980). Children's use of phonological encoding when reading for meaning. *Memory & Cognition*, **8**, 195-209.
- EICH, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition*, **12**, 105-111.
- FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- GALBRAITH, G. G., & TASCHMAN, C. S. (1969). Homophone units: A normative and methodological investigation of the strength of component elements. *Journal of Verbal Learning & Verbal Behavior*, **8**, 737-744.
- GARTMAN, L. M., & JOHNSON, N. F. (1972). Massed versus distributed repetition of homographs: A test of the differential-encoding hypothesis. *Journal of Verbal Learning & Verbal Behavior*, **11**, 801-808.
- GEIS, M. F., & WINOGRAD, E. (1974). Norms of semantic encoding variability for fifty homographs. *Bulletin of the Psychonomic Society*, **3**, 429-431.
- 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.
- GORFEIN, D. S. (2001). On the consequences of meaning selection: An overview. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 3-8). Washington, DC: American Psychological Association.
- GORFEIN, D. S., & BROWN, V. R. (2007). Saying no to inhibition: The encoding and use of words. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 103-124). Washington, DC: American Psychological Association.
- GORFEIN, D. S., BROWN, V. R., & DEBIASI, C. (2007). The activation-selection model of meaning: Explaining why the son comes out after the sun. *Memory & Cognition*, **35**, 1986-2000.
- GORFEIN, D. S., & DEBIASI, C. (1997, November). *Word versus picture priming of two implicit memory tasks*. Poster presented at the Annual Meeting of the Psychonomic Society, Philadelphia.
- GORFEIN, D. S., & O'BRIEN, J. (1985). *The Adelphi norms for homophone spelling*. Unpublished manuscript.
- GORFEIN, D. S., VIVIANI, J. M., & LEDDO, J. (1982). Norms as a tool for the study of homography. *Memory & Cognition*, **10**, 503-509.

- GORFEIN, D. S., & WALTERS, M. F. (1989). When does "soar" become "sore"? Some comments on the chapter of Simpson and Kellas. In D. S. Gorfein (Ed.), *Resolving semantic ambiguity* (pp. 57-62). New York: Springer.
- JACOBY, L. L., & WITHERSPOON, D. (1982). Remembering without awareness. *Canadian Journal of Psychology*, **36**, 300-324.
- KAUSLER, D. H., & KOLLASCH, S. F. (1970). Word associations to homographs. *Journal of Verbal Learning & Verbal Behavior*, **9**, 444-449.
- KAUSLER, D. H., & KOLLASCH, S. F. (1973). Continued associations to homographs. *Bulletin of the Psychonomic Society*, **1**, 17-18.
- KREUZ, R. J. (1987). The subjective familiarity of English homophones. *Memory & Cognition*, **15**, 154-168.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- NELSON, D. L., McEVoy, C. L., WALLING, J. R., & WHEELER, J. W. (1980). The University of South Florida homograph norms. *Behavior Research Methods & Instrumentation*, **12**, 16-37.
- PERFETTI, C. A., LINDSAY, R., & GARSON, B. (1971). *Association and uncertainty: Norms of association to ambiguous words*. Pittsburgh, PA: University of Pittsburgh, Learning Research & Development Center.
- POLLATSEK, A., LESCH, M., MORRIS, R. K., & RAYNER, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 148-162.
- SIMPSON, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin*, **96**, 316-340.
- SIMPSON, G. B., & KANG, H. (1994). Inhibitory processes in the recognition of homograph meanings. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 359-381). San Diego: Academic Press.
- TWILLEY, L. C., DIXON, P., TAYLOR, D., & CLARK, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition*, **22**, 111-126.
- WHITE, K. K., & ABRAMS, L. (2004). Free associations and dominance ratings of homophones for young and older adults. *Behavior Research Methods, Instruments, & Computers*, **36**, 408-420.
- WINOGRAD, E., & GEIS, M. F. (1974). Semantic encoding and recognition memory: A test of encoding variability theory. *Journal of Experimental Psychology*, **102**, 1061-1068.
- YATES, J. (1978). Priming dominant and unusual senses of ambiguous words. *Memory & Cognition*, **6**, 636-643.

NOTES

1. At the suggestion of an anonymous reviewer, we have included the HAL frequencies, since they are more up to date and have been shown to be better predictors of performance in such tasks as lexical decision and word naming (Balota et al., 2007).

2. Sentences were constructed in such a way that the experimenters judged the intended meaning of each homophone to be clear in its con-

text. Although we cannot be certain that the homophones were identical with respect to their predictability in the sentence frames, it is reasonable to assume that any differences in predictability would have had a negligible impact on the particular measure that was used. That is, the task was to decide whether or not the sentence made sense (rather than, for example, a sensibility measure that used a continuous scale, or a reaction time measure).

3. The number of items varied somewhat from cell to cell, but the *N* was quite large, ranging from 399 to 414. Therefore, we chose to present the values of all scorable items for each correlation.

ARCHIVED MATERIALS

The following materials associated with this article may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, www.psychonomic.org/archive.

To access these files, search the archive for this article using the journal name (*Behavior Research Methods*), the first author's name (Gorfein), and the publication year (2008).

FILE: Gorfein-BRM-2008.zip

DESCRIPTION: The compressed archive file contains eight files:

Gorfein (APP-A).txt, containing the proportions of participants who provided particular spelling for one of the auditorily presented homophones in the uncued task of Experiment 1.

Gorfein (APP-A).xls, containing the above information in Excel spreadsheet format.

Gorfein (APP-B).txt, containing the proportion of correctly spelled words under each cue condition in Experiment 1, as well as the proportion of times an alternative spelling was given for each word, and the proportion of times participants provided a spelling that did not match any form in the homophone set.

Gorfein (APP-B).xls, containing the above information in Excel spreadsheet format.

Gorfein (APP-C).txt, containing the proportions of correct, alternative, and other responses for each of the visually presented homophone sets in Experiment 2.

Gorfein (APP-C).xls, containing the above information in Excel spreadsheet format.

Gorfein (APP-D).txt, containing correct acceptance scores and correct rejection scores for homophone sentences in the sentence acceptability task of Experiment 3.

Gorfein (APP-D).xls, containing the above information in Excel spreadsheet format.

AUTHOR'S E-MAIL ADDRESS: dgorfein@yahoo.com.

(Manuscript received July 2, 2007;
revision accepted for publication November 23, 2007.)