# Homograph norms: An alternative approach to determining meaning dominance

# NANCY R. GEE AND SHELLY L. HARRIS

SUNY Fredonia, Fredonia, New York

In the present study, we provide a new technique for the collection of homograph norms that reduces subjectivity in the determination of meaning dominance by allowing participants rather than experimenters to indicate to which meaning or meanings the associates were related. To evaluate the effectiveness of this new technique, a subset of homograph norms were used in three separate experiments, demonstrating that (1) when presented with additional meaning categories; (2) overall, the participants were most familiar with primary meanings, followed by secondary, tertiary, and quaternary meanings; and (3) the meaning categories provided to the participants during norms collection were appropriate, since the two meanings provided for each homograph by the participants were consistent with the original data. Finally, in a fourth experiment, we compared the results of this new technique with a parallel set collected in Australia. The high degree of similarity in the results provides validity for this procedure. The homograph norms discussed in this article may be downloaded from http://brm.psychonomic-journals.org/ content/supplemental.

The English language is complicated by the fact that many words are spelled the same but have different meanings (Cramer, 1970). For example, the word *bat* can refer to the winged mammal or to several different pieces of sporting equipment. Devoid of context, it is impossible to distinguish which meaning should be applied. These ambiguous words, which have the same spelling but different meanings, are referred to as *homographs*, and they have a long history of study in the psychology of language. Often, these words are not only spelled the same, but also sound the same, which can create confusion for teachers and experimenters and may even present a challenge during conversation (Cartwright & Nickerson, 1979).

Some linguists further distinguish *polysemes*, words with the same spelling and overall meaning but a different sense, from those of other homographs. At times, polysemous words can be relatively easily identified as such. *Milk* for instance, in its verb form means to extract a liquid and in noun form means the liquid extracted. In general, however, the discriminations necessary to call out polysemes from other homographs are subtle and subjective, and they will therefore not be the focus of this research.

Although homographs can complicate language comprehension, they also have a wide range of beneficial applications. They have been used as a means to map and understand the organization of lexical knowledge (Nelson, McEvoy, & Schreiber, 2004) and are considered by some researchers to be ideal stimuli (Wollen, Cox, Coahran, Shea, & Kirby, 1980) because once a baseline of meaning frequency is established, the context can be manipulated at encoding. This makes them particularly useful in recall and recognition tasks (e.g., Carroll & Simington, 1986; Gee, 1997; Hutchison & Balota, 2005). Other conventional uses of ambiguous words (not limited to homographs that are written, but also those presented auditorily, known as *homophones*) include cross-modal priming experiments (e.g., Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney, 1979), word recognition studies dealing with the structure of the lexicon (e.g., Gerard & Scarborough, 1989; Titone & Salisbury, 2004), and eye movement experiments (e.g., Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Sereno, Pacht, & Rayner, 1992). Homographs have also been employed in research outside of these traditional uses in order to examine the cognitive consequences of a depressed mood (Hertel & El-Messidi, 2006) and to show that syllable stress is the primary determinant of word color in synesthesia (Simner, Glover, & Mowat, 2006). Additionally, they have been used to investigate characteristics of a variety of populations, including children with autism, where homographs were used to examine deficits in contextual processing (Hala, Pexman, & Glenwright, 2007).

Before homographs can be adequately utilized for research purposes, a baseline of meaning frequency must be established. In the absence of context, some meanings of homographs are more likely to occur than others. Although some homographs are relatively balanced (i.e., their meanings occur with about the same frequency), others can be considered polar (i.e., one meaning occurs

N. R. Gee, nancy.gee@fredonia.edu





much more often than the others). If homographs are to be used in studies in which a biasing context is not provided, it is imperative to have accurate estimates of which meaning is most likely to be thought of by the participant for a given homograph. Researchers have developed a variety of methods to provide empirical norms of meaning frequency, but here we will focus on the most common method: free association. In this technique, researchers select words that they determine to have at least two distinct meanings, usually from a dictionary. These words are then presented individually to participants who are asked to provide the first word that comes to mind in response to that word. The individual word associates are then separated into categories on the basis of their presumed meaning. Experimenters typically categorize the associates into separate meaning categories individually and then as a group in order to reach a final consensus about which associates should be classified into which categories. However, not all responses fit neatly into a single meaning classification, so additional meaning categories might be created or multiple categories might be combined on the basis of the judgment of the experimenters in order to accommodate the data provided by the participants. For example, in Nelson's homograph norms (Nelson, McEvoy, Walling, & Wheeler, 1980), when participants were presented with the homograph bill they reported both "dollar" and "statement," which Nelson et al. (1980) ultimately grouped together under the meaning money. Although this decision seems like a logical choice, it represents a judgment call made by the consensus of the group on the basis of their perceptions of which meaning the participants intended by their responses.

Gawlick-Grendell and Woltz (1994) pointed out that although typical, group consensus is not the only method used to categorize participant responses. One of the different methods to accomplish this task is grouping responses according to meaning categories determined prior to the experiment, as was done by Cramer (1970). Cramer provided a table that lists the homographs used in her study along with two meanings for each. For example, the homograph court is given, with Meaning 1 being court of law and Meaning 2 being royal palace. Once word associates were collected from the participants, three independent judges decided whether the word associates to each homograph referred to Meaning 1, Meaning 2, or neither. Those associates that did not fall into one of the two predetermined meaning categories were not reported at all. For the homograph *court*, the average number of Meaning 1 (court of law) associates as determined by the experimenter was reported as 73.3, and the average number of Meaning 2 (royal palace) associates reported was 14.6.

Whereas Cramer (1970) used a combination of predetermined meanings and word associates to calculate meaning frequencies, Warren, Bresnick, and Green (1977) relied on definitions that they collected from their participants. Their method involved clustering responses by grouping participant definitions by similarity after the experiment concluded. They presented a list of homographs with a corresponding list of the subsequent definitions that were reported by the participants and the frequency of each definition. Twilley, Dixon, Taylor, and Clark (1994) used free association responses that were categorized by two raters working collaboratively, with a third rater reviewing all of their decisions. If an associate was ambiguous with regard to meaning category and it could be reasonably placed under more than one, it was classified under a category titled "unclear." This classification was also used for associates that the researchers deemed to be incorrect and those that had no association at all.

Gawlick-Grendell and Woltz (1994) used a combination of the aforementioned methods to sort homograph norms. They not only clustered definitions that they collected from participants on the basis of meaning similarity but also verified their decisions by direct comparison to the meanings listed in a dictionary. If the experimenters felt that their meaning categories were too permissive or restrictive after consulting a dictionary, they would either expand or collapse categories accordingly. In order to get around inevitable disagreements on how the responses were categorized, published data have at times included not only homographs and meaning categories but also raw responses and frequencies of each (e.g., Nelson et al., 1980; Perfetti, Lindsay, & Garson, 1971).

Although there are a variety of different methods for categorizing participant responses in homograph norms, the typical measurement used to report meaning frequency is called meaning dominance. It is the ratio of the responses that were grouped under one particular meaning category over all possible responses. Meaning dominance values for homograph norms have been previously published, but their usefulness has been limited because of the small number of meanings given (Gawlick-Grendell & Woltz, 1994). In the majority of studies, methodologies have been utilized in which participants are asked for the first one or two word responses that come to mind. Gawlick-Grendell and Woltz pointed out that some researchers only reported the two meanings of homographs with the highest frequency, even though many homographs have more than two meanings. They provided examples of studies in which their results were reported in this manner (e.g., Cramer, 1970; Geis & Winograd, 1974; Gorfein, Viviani, & Leddo, 1982; Kausler & Kollasch, 1970). Associates to individual homographs are frequently used as retrieval cues or priming items for homographic words, and in some applications, less dominant meanings are crucial. Therefore, reporting less dominant meanings in homograph norms is very important for later uses in research.

In the present study, we employed an innovative technique for determining meaning dominance in homographs. In previous studies, individual responses to homographs were classified into a meaning on the basis of a general agreement among a small number of researchers (e.g., Cartwright & Nickerson, 1979; Cramer, 1970; Gawlick-Grendell & Woltz, 1994; Geis & Winograd, 1974) whereby a given response was assigned to only one meaning. Gee (1997) pointed out that this is potentially problematic, because some associates can reasonably be classified into more than one category of meaning. For instance, "wings" was given as a response to the homograph fly (Gee, 1997). This response could logically be classified into either the meaning to fly in the air or the meaning *insect*. The participants could have easily intended that "wings" was related to the to fly in the air meaning, they could have been referring to the insect meaning of the homograph fly, or both. The researchers who make the decisions about which meaning was intended by the participants cannot know which meaning the participants had in mind, so they make the decision on the basis of their own experiences. By contrast, the methodology in the present study requires the participants to classify associates into one or more of the multiple meanings of each homograph.

Under the present methodology, we presented participants with a homograph along with a set of word associates taken from the Nelson, McEvoy, and Schreiber (1998) norms. The participants were also presented with meaning categories for each homograph that were drawn from a dictionary (Gove et al., 1993). A small group of researchers agreed on a brief description or word that best represented each meaning. For example, consider the homograph *right*. Two distinct meanings of this word were elucidated by agreement of three researchers: *direction* and *correct*. In Figure 1, the homograph *right* is included in the top left corner, with the two meaning categories across the top and the associates listed directly below and spaces available for the Xs to be assigned by the participants.

The participants were asked to indicate which meaning categories the associates were related to by marking an X in the appropriate box or boxes. In this manner, the participants (not the researchers) were able to indicate which meaning category or categories a given response was associated with. This technique allowed the participants the flexibility to classify associates into more than one meaning category, eliminating the subjectivity of the experimenter's opinion regarding which meaning was intended.

# **EXPERIMENT 1**

In the first phase of data collection, participants who reported that English was their native language were presented with a packet of tables. Each table contained one homograph, the possible meaning categories for that homograph, and the individual associates to that homograph as determined by single response association norms (Nelson et al., 1998). The participants were asked to determine which of the meaning categories each individual response was related to, and they were informed that each individual response could be related to more than one meaning.

# Method

**Participants**. The participants were 2,257 undergraduate students from SUNY Fredonia. They received extra credit for an intro-

RIGHT	DIRECTION	CORRECT
WRONG		
LEFT		
GO		
ARM		
NORTH		
ON		

Figure 1. Example homograph form from Experiment 1.

ductory psychology course, their primary language was English, and all of the participants were at least 18 years old. No one participated in the experiment more than twice, and if they did participate twice, they received a different set of homographs each time. Data were collected over a period of 10 years, and 100 participants were scored for each of 197 different homographs.

Materials. The set of homographs chosen for inclusion in this experiment were selected because they and their associates appeared in an available online word association, rhyme, and wordfragment norms database (Nelson et al., 1998). In selecting words from this database, we made an attempt to use words that also appeared in other previously published homograph norms, and we were able to find 72 such homographs (Azuma, 1996; Cramer, 1970; Gawlick-Grendell & Woltz, 1994; Geis & Winograd, 1974; Gilhooly & Logie, 1980; Gorfein et al., 1982; Kausler & Kollasch, 1970; Nelson et al., 1980; Perfetti et al., 1971; Twilley et al., 1994; Warren et al., 1977; Wollen et al., 1980). Additionally, we selected 125 other homographs from the Nelson database that we were not able to find in the other published homograph norms. A group of three experimenters evaluated the meanings listed in previous research, when available, in conjunction with a dictionary (Gove et al., 1993), to determine the meaning categories that were used in the experiment. See Appendix A for a list of the 197 homographs used in this experiment.

**Procedure**. Each homograph was placed in a separate table, along with its respective list of word associates and meaning categories. Approximately 20 homographs were randomly selected to be included in any given response booklet. Multiple response booklets were constructed such that the pages within each booklet were presented in different random orders. The response booklets and pages were created as needed to fill out the 100 participant requirement for each homograph. When 100 data sheets had been completed for a given homograph, that table was no longer included in any of the new response booklets.

As is shown in Figure 1 for the homograph *right*, the homographs appeared in bold type in the upper left, and their meaning categories were printed across the page on the same line. The associates were printed directly beneath the homograph.

The participants were given instructions that defined homographs and detailed the various meanings of the words, as well as the associate words that were given to each homograph in the previous normative study. An example was drawn on the chalkboard of the homograph right (see Figure 1). It was explained to the participants that the word *right* is a homograph that has two distinct meanings: a direction meaning and a correct meaning. They were shown that each of the associates of the homograph word could be related to one or both meanings of the word right. For example, the associate word wrong could be related to both the direction and the correct meaning. The participants were instructed to classify the related words into as many meaning categories as they considered appropriate by placing an X into the box next to the word and below the meaning category that they classified it into. The participants were shown an example of how this would work if a person assumed that the response wrong was related to both the *direction* and the *correct*  meanings of the word *right*. The experimenter placed an X in both boxes next to the associate *wrong* on the table.

Groups of participants completed the response booklets in various classrooms around campus. Each booklet contained approximately 20 homographs, and all of the participants took less than 60 min to complete their booklet.

# **Results and Discussion**

The norms (homographs, meaning categories, associates, summary classification responses, and meaning dominance values) for all 197 homographs are provided as supplemental materials that can be downloaded from http:// brm.psychonomic-journals.org/content/supplemental. Table 1 details the Pearson product-moment correlation coefficients that were calculated between dominance values calculated for our norms and previously published values using the method outlined in Nelson et al. (1980). In comparing our dominance values to those in previous studies, note that this new technique does not calculate meaning dominance values as they have been traditionally defined. In previous studies, associates have only been allowed to be classified into a single meaning at most, whereas in the present experiment, associates may be included in more than one category of meaning. Dominance value in this procedure was, therefore, computed as the proportion of associates across all participants that have been classified under one meaning category relative to the total number of classifications made across all categories, where associates may be included in more than one category of meaning.

Correlations were performed between the present experiment and each previous study, individually. For the overlapping homographs, the dominance values of all meanings were compared. For example, there were 100 homographs normed in Cramer (1970), 9 of which were identical to homographs in the present experiment. Among those 9 homographs, there were 17 meanings that were compared between that Cramer and the present experiment. The overall correlation between other studies with the present method was .70, which is rather remarkable, considering the many different methodologies used (N = 461).<sup>1</sup>

Meaning dominance values from the data collected in our experiment were compared with those from 12 other previous sets of homograph norms (Azuma, 1996; Cramer, 1970; Gawlick-Grendell & Woltz, 1994; Geis & Winograd, 1974; Gilhooly & Logie, 1980; Gorfein et al., 1982; Kausler & Kollasch, 1970; Nelson et al., 1980; Perfetti et al., 1971; Twilley et al., 1994; Warren et al., 1977; Wollen et al., 1980). This analysis revealed that in some cases, the number of meanings listed for each homograph influenced the dominance values calculated. When you consider that dominance values represent a proportion of total responses, this certainly makes sense. For example, the homograph *star* in our experiment had three meanings—*sky*, celestial, and celebrity—and the dominance values were .44, .41, and .15, respectively. There is arguably no true distinction between *sky* and *celestial* (except perhaps to astronomical lexicographers), and had they been merged, the dominance for the combined term would have been significantly more pronounced. When we compared the meanings of *star* to those in Nelson et al. (1980) in which the homograph was also used, we found that they listed only two meanings for that word: *heavenly body* (.87) and *prominent performer* (.13). If we were to merge the *sky* and *celestial* meaning categories (.44 + .41 = .85) into one category, similar to their *heavenly body* category, our results closely match those of Nelson et al. (1980).

In considering these results, we wanted to address a number of questions: First, did we use the appropriate number of meaning categories (Experiment 2), and were those meanings familiar to our participants (Experiment 3)? Second, did we use appropriate meaning categories for each homograph as part of our methodology (Experiment 4)? Third, would our results generalize for use in another English-speaking country (Experiment 5)? Each of these questions will be addressed below.

# **EXPERIMENT 2**

On the basis of examples like that of the homograph star described above, we decided to examine the impact of the number of meaning categories provided in our methodology in greater detail. In Experiment 2, we varied the number of meaning categories provided to participants for individual homographs. First, we selected the subgroup of homographs from our study that had the largest amount of disagreement in terms of meaning dominance values with previous studies. Second, we systematically altered the number of meaning categories that we presented to the participants for each of these homographs. We always presented at least two meaning categories, including the primary and secondary meanings for each homograph. Other conditions included the top three meanings or all four meanings. This procedure would allow us to examine the extent to which dominance values differed on the basis of the number of meaning categories presented to the participants, and it also allowed us to examine the validity of the dominance values that we calculated in the original data collection.

#### Method

**Participants.** Thirty-three SUNY Fredonia undergraduate students participated in the experiment for extra credit in one of their psychology courses. Their primary language was English, and all of the participants were at least 18 years old.

**Materials**. The homographs were selected on the basis of how many primary meanings differed between those from our study and those from 12 other studies. There were 72 homographs that overlapped between our study and at least one previously published study (refer to Table 1 for the specific number of overlapping homographs by study). As Appendix B shows, there were 17 homographs for which a different primary meaning from that in the present experiment was reported in at least two other studies and 20 additional homographs for which a different primary meaning was reported in one other study. In order to manipulate the number of meaning categories in this experiment, it was necessary to choose the 15 homographs that had four meanings out of the subset of 37. These homographs with four meanings are indicated with asterisks in Appendix B.

and correlations between the rresent and other rorms								
Norms	Number of Items Normed	Overlapping With Present Data Set	R	n				
Cramer (1970)	100	9	.66	17				
Kausler & Kollasch (1970)	40	3	.38	8				
Perfetti et al. (1971)	109	13	.70	28				
Geis & Winograd (1974)	50	9	.45	18				
Warren et al. (1977)	20	2	.81	4				
Gilhooly & Logie (1980)	387	22	.52	44				
Nelson et al. (1980)	314	42	.76	85				
Wollen et al. (1980)	117	17	.67	42				
Gorfein et al. (1982)	96	15	.78	30				
Gawlick-Grendell & Woltz (1994)	120	19	.81	46				
Twilley et al. (1994)	566	49	.73	101				
Azuma (1996)	110	14	.63	38				

Table 1
Available Homograph Norms, Number of Items Evaluated
Number of Items Overlapping With the Present Norms,
and Correlations Between the Present and Other Norms

There was one homograph per page of the booklets. Three different types of packets were made: homographs with two meaning categories, homographs with three meaning categories, and homographs with four meaning categories. The order of the meanings presented across the top was consistent with that in Experiment 1.

**Procedure**. We used the same procedure here as was used in Experiment 1. Again, all of the participants took less than 60 min to complete their booklet.

#### **Results and Discussion**

A 2 (primacy: primary vs. secondary meaning)  $\times$  3 (number of meaning categories: 2, 3, or 4) mixed-model ANOVA was conducted on the data. It revealed a significant main effect of primacy [ $F(1,42) = 27.60, R^2 = .99$ ]. Primary meanings had a significantly higher meaning dominance (M = .44, SD = .11) than secondary meanings (M = .35, SD = .09). A significant main effect for number of meaning categories was also obtained [ $F(2,42) = 253.84, R^2 > .99$ ]. Multiple comparisons revealed that the homographs with two meaning categories had the highest meaning dominance value (M = .50, SD = .004), followed by three meaning categories (M = .31, SD = .03) and four meaning categories (M = .31, SD = .03). There was no significant interaction of meaning dominance  $\times$  number of meaning categories [ $F(2,42) = 0.88, R^2 = .19$ ].

Given the disparity that we found in the literature with regard to specific dominance values obtained for individual homographs, we needed to determine whether the number of meaning categories that the participants were exposed to had an impact on the meaning dominance values that we calculated. Our analysis revealed-not surprisinglya main effect for primacy, which demonstrated that dominance values for the primary meanings were consistently higher than the secondary meanings. This analysis also revealed that as the number of meaning categories provided to the participants increased, the dominance values tended to decrease. Again, this finding is not surprising, since dominance is calculated as a proportion of total responses, and this effect is inherent in the way that norms are computed. When there are more options for responses (i.e., more boxes to check), the meaning dominance values decrease. Even in this new technique, in which the participants were free to indicate multiple meanings for a given response, they still tended to choose meanings with the same relative frequency. In other words, this analysis revealed that regardless of the number of meaning categories presented to the participants, meaning dominance values for primary and secondary meanings were not functionally altered or reordered, but rather the change that we found was one of overall magnitude.

Although this effect does not create a problem for researchers who are simply interested in the order of dominance, it does pose a conundrum for those interested in controlling for or manipulating dominance values. If an experimenter is interested in selecting a homograph with a primary meaning that has a relatively low dominance value, it will naturally be a homograph with more meanings than if the desired primary meaning has value that is relatively high. Again, this is because dominance is calculated as a proportion.

Earlier, we asked whether we used the appropriate number of meaning categories in our study. The answer to this question seems to be that it does not matter. When we altered the number of meaning categories presented to the participants for a given homograph, the primacy of the meanings remained stable, but the magnitude of the overall meaning dominance values decreased as the number of meaning categories increased. This finding elucidates the impact of the way meaning dominance values are calculated. They represent a ratio of responses for a given meaning relative to the total number of responses. Providing more meanings simply provides more opportunities for responses, thus decreasing the overall ratio. We found that increasing the number of meaning categories did nothing to alter the ordering of those meanings in terms of their primacy.

#### **EXPERIMENT 3**

Once data collection had begun on the homograph norms (Experiment 1), we were committed to our original

method of meaning selection, and we could not deviate from our predetermined protocol for meaning selection. Because we had selected the meanings of each homograph, we were concerned about the extent to which our participants were familiar with those meanings. Therefore, we decided to conduct an experiment using the Gilhooly and Logie (1980) methodology for determining participant familiarity with the meanings.

#### Method

**Participants.** One hundred seventeen undergraduate students from SUNY Fredonia participated in this experiment. They received extra credit for an introductory psychology course, their primary language was English, and all of the participants were at least 18 years old.

**Materials**. All 37 homographs that had conflicting order of dominance values with those of previous studies (refer to the Materials section of Experiment 2 for further explanation) were included in this experiment. The booklet was formatted similarly to that in Azuma (1996). In order to evaluate the appropriateness of the meaning categories provided to the participants in the original norms collection, we used those exact meanings for this experiment.

Procedure. Each participant was given one booklet with a random ordering of homographs (with their respective meanings), one per page, to complete. All of the participants received homographs with the same random order of meanings. The participants were given instructions in written form. These instructions indicated that we were particularly interested in exploring characteristics of words that are meaningfully related to homographs. Furthermore, we provided them with a homograph and its meaning, and they were asked to rate how familiar they were with each of those meanings. The example right was used with the meaning direction and the meaning correct. The example showed the 7 circled for the direction meaning along with an explanation that the experimenter was "very familiar with the *direction* meaning," so they (i.e., experimenter) circled 7. The example showed the 6 circled for the *correct* meaning along with an explanation that they experimenter was "slightly less familiar with the correct meaning" so they circled 6. It was indicated that the familiarity scale was set up such that if a meaning is not at all familiar to the participants, they should rate it a 1, and if it was very familiar to them, they should rate it a 7, and all other levels of familiarity would be rated between those two endpoints.

The participants completed their booklets in groups, in classrooms. All of the participants took less than thirty minutes to complete their booklet.

#### **Results and Discussion**

A one-way within-subjects ANOVA<sup>2</sup> was conducted on the average familiarity ratings for each meaning. There was a significant main effect of primacy of meaning  $[F(3,48) = 7.22, R^2 = .96]$ . Linear contrasts [F(1,16) = $21.96, R^2 > .99]$  revealed that the highest familiarity ratings were assigned to primary meanings (M = 6.39, SD =0.52), followed by secondary (M = 5.58, SD = 0.92), then tertiary (M = 5.23, SD = 1.00) meanings, and the lowest familiarity ratings were assigned to quaternary meanings (M = 5.17, SD = 0.99).

Not surprisingly, our results demonstrated that the participants' level of familiarity with the meaning categories used in the original norms collection matched the primacy of the meaning itself, such that they were most familiar with primary meanings, followed by secondary meanings, followed by tertiary meanings, followed by quaternary meanings. Additionally, our results demonstrate that on a scale of 1–7, where 7 indicated the highest level of familiarity, our participants' lowest average rating of familiarity (quaternary meaning) was over 5. It is apparent from these data that our participants were sufficiently familiar with the meaning categories that we used in the normative data collection (Experiment 1).

#### **EXPERIMENT 4**

The process of developing a set of homograph norms typically involves a subjective component in the determination of the meaning categories that are appropriate for a given homograph. In most cases, a set of judges determines which categories are appropriate and which associates from the participants should be classified as intending which meaning. Although in the present experiment we utilized a technique in which the participants categorized the associates produced by other participants, which eliminates the experimenters' making that determination, the meanings provided for each homograph were determined in advance of the experiment by a set of judges. It is possible that the meaning categories that were selected were inappropriate, simply because they were determined by a set of judges rather than by the participants. To investigate this possibility, we again took a subset of the homographs used in our homograph data collection experiment and asked the participants to generate two meanings for each word. To clarify their intentions, we also asked them to generate a sentence for each of the two meanings. This procedure was intended to allow us to determine whether the meanings that we selected were consistent with the meanings that our participants considered appropriate.

#### Method

**Participants**. One hundred seventeen undergraduate students from SUNY Fredonia participated in this experiment. They received extra credit for an introductory psychology course, their primary language was English, and all of the participants were at least 18 years old.

**Materials.** The same 37 homographs used in Experiment 3 were included in this experiment.

**Procedure.** The participants completed booklets that contained a random ordering of homographs with one homograph per page. The participants were informed that we were interested in exploring words that have more than one meaning. In response to each homograph, the participants were asked to provide two meanings and a sentence for each meaning. They were then shown an example of the homograph *right* and the definitions *direction* and *correct* with short sentences.

Again, the participants completed their booklets in groups, in classrooms. They took less than 30 min to complete their response booklet.

#### **Results and Discussion**

Two independent raters determined the different meanings for each homograph on the basis of the responses given by the participants. We examined interrater reliabilities separately for the first meanings and for the second meanings provided by the participants. Many researchers include idiosyncratic responses (e.g., Azuma, 1996; Gawlick-Grendell & Woltz, 1994; Nelson et al., 1980; Perfetti et al., 1971) in their analyses, so we also calculated interrater reliabilities with and without the category of idiosyncratic responses (denoted as any definitions provided by participants with a frequency of less than 2). For the first meaning provided by the participants, our raters had an r = .91 (n = 174) with the number of idiosyncratic responses included and an r = .89 (n = 137) without that number included. For the second meaning provided by the participants, our raters had an r = .83 (n = 192) with idiosyncratic responses included and an r = .80 (n = 156) without idiosyncratic responses included.

In order to directly compare the meanings that we found in this experiment with the meanings that we used in the Experiment 1 procedure, we again carried out a correlation analysis. Each independent rater determined the number of participants who generated a given meaning for each homograph. We averaged these numbers across raters to determine the number of participants who responded with each meaning. For example, for the homograph brush, Rater 1 indicated that 67 participants responded with the tool meaning, and rater 2 indicated that 73 participants generated the same meaning. The average of these two values is 70. We then correlated these average values with the meaning dominance values from the original homograph norms data collection procedure. This correlation provides a measure of the degree of relationship between the meanings generated in this experiment with the actual meanings used in the original normative experiment (Experiment 1).

On the basis of this experiment, it appears that the appropriate meanings were used during the norms collection procedure. Even taking into consideration the completely different methodologies (participants categorizing previously collected word associates compared with participants providing two definitions), it is noteworthy that the correlation of .41 between the two procedures was significant (n = 92). This shows that although the meanings that were agreed on among the experimenters while consulting a dictionary were not a perfect match to what the participants would have defined as meanings of the homographs, the meanings that were provided to the participants in which to categorize the word associates appeared to be appropriate for the purposes of norms collection.

One of the interesting comparisons between the two methodologies is the difference in number of meanings. We performed an independent means *t* test to determine whether there was a significant difference in the number of meanings generated by these two different approaches and found that the difference was not statistically significant [t(72) = -1.88, p = .06]. Although this difference was not significant, it did approach significance, and on average, the method used in the norms collection generated slightly fewer meanings (M = 3.30, SD = 0.78) than the method utilized in the present experiment (M = 3.81, SD = 1.47). It is possible that when the raters made the meaning determinations for the original norms collection with the aid of a dictionary, they focused on the most distinct meanings and therefore limited the overall number of meanings used for each homograph, whereas when the participants in this experiment freely generated two meanings without the aid of a dictionary, they listed whatever came to mind regardless of how distinct that particular meaning might be, just so they could fill in their response sheets. Another possibility is that some of the words might actually have new or even culturally relevant meanings, since the English language is constantly evolving, and the meanings used in the original norms procedure were generated several years earlier and on the basis of other even older homograph norms.

## **EXPERIMENT 5**

We wanted to examine the generalizability of our results, so we opted to collect data in another country on a subset of the homographs that we used in Experiment 1. Previous researchers have compared word associates in English, French, German, and Italian (Rosenzweig, 1961) and subsequently compared word associations between college students in Nigeria with those from the United States (Ekpo-Ufot, 1978). These studies revealed a significant amount of similarity in the cross-cultural comparisons. The students from Nigeria, who had English as a second language and had many different primary languages among them, had 42% identical associates (out of 90 comparisons) that overlapped with those from the United States for primary, secondary, and tertiary meanings. We chose to draw a cross-cultural comparison with Australia, where the primary language is English, to establish whether the homograph norms could also be used satisfactorily in another English-speaking country. If societal factors and popular culture have a large impact on the determination of meanings for homographs, we would expect little generalizability, but on the other hand, if the multiple meanings of homographs are more fundamental to the language, we would expect to see similar results for homographic data collected in another English-speaking country.

#### Method

**Participants.** The participants were 100 undergraduate students enrolled at Adelaide University in South Australia. All of them were native English speakers, over 18 years old, and participated for extra credit in an introductory psychology course.

**Materials**. A random subset of 34 homographs of the original 197 used in Experiment 1 was used.

**Procedure**. The same procedure was used for data collection here as was used in Experiment 1.

#### **Results and Discussion**

The number of responses indicated for each homograph meaning was tabulated for the Australian participants and provided side by side with the original norms collection data from U.S. participants (refer to the supplemental materials). For example, 98 U.S. participants indicated that for the homograph *bat*, the response "ball" is related to the *baseball* meaning, whereas 99 Australia participants

indicated that the response "ball" was related to the *base-ball* meaning. Each number in the corresponding cell represents the number of students who indicated that a particular response was related to a particular meaning. Two correlational analyses were completed, one for responses based on each of the possible meaning hierarchies (i.e., primary, secondary, tertiary, and quaternary) and one for meaning dominance values that compared homographs with two, three, and four meanings across the two countries.

Correlational analysis of participant responses. The Pearson product moment correlation coefficients were calculated for all responses to each of the possible meanings for each homograph across the two countries (primary, r = .97, n = 483; secondary, r = .96, n = 483; tertiary, r = .97, n = 324; and quaternary, r = .98, n =93). All of the correlations were found to be significant. There appears to be an impressive degree of similarity with regard to the number of responses given to each meaning of homographs across this sample of U.S. and Australian participants. For instance, for the homograph coat, the meaning hierarchy for the U.S. and Australia data is identical; the primary meaning is *jacket*, followed by fur, hanger, and paint. The dominance values for the four meanings in the U.S. were .38, .29, .20, and .12, compared with the very similar Australia dominance values of .39, .28, .19, and .14.

Correlational analysis of meaning dominance values. Meaning dominance values were calculated by taking the total number of responses for each meaning and dividing by the total number of responses for that particular homograph. For the U.S. data on the homograph company, the meaning dominance values were .56 for business and .44 for *friends*. The Australia data for *company* shows meaning dominance values that were the same as the U.S. data. The amount of agreement across these two samples in terms of the dominance values calculated was so strong that, with one exception, the ordering of the dominance values across meanings for all of the homographs was identical for the two countries. The single homograph that had a different order of dominant meanings was trip, and it had the same primary meaning across the two samples (vacation) and the same dominance value for that meaning (.54). It was the second and third meanings that produced a slight difference in dominance values, resulting in a different ordering of meanings. As can be seen in the supplemental materials, the second and third meanings in the U.S. were *fall* (dominance value = .24) and *hal*lucinate (.22), but in Australia, these were reversed such that *hallucinate* was second (.25) and *fall* was third (.20). Although the ordering of the meanings for this single homograph is different on the basis of meaning dominance values, the actual values for the secondary and tertiary meanings are all very close in terms of absolute value (range = .20 - .25).

Correlations were also calculated for the meaning dominance values across the two samples for homographs with two (r = .99, p < .01, n = 12), three (r = .99, p < .01, n = 16), or four (r = .99, p < .01, n = 6)

meanings, and they were all extremely large correlations, indicating a near-perfect relationship between these two samples. The results of this comparison experiment show that not only does the hierarchy of meanings stay stable among homographs (with one exception), but that the actual dominance values are nearly identical across these two samples.

Note that although the cross-cultural agreement between the U.S. and Australia was notable in this experiment, differences in culture, economy, ethnicity, or other factors can nevertheless produce different dominance results even within geographically near regions that share the same language. Arthur (1971) compared student samples for homographs collected in Kingston, Ontario, Canada, with those previously normed in Minnesota by Palermo and Jenkins (1964). Of the 20 homographs included in the comparison, the researcher found 19 that had at least one significant difference in response frequency on the basis of region. The existence of such variations may advocate for a more refined analysis of regional norms, in which undertaking the present technique could be beneficial.

## **GENERAL DISCUSSION**

The technique presented in Experiment 1 for collecting meaning dominance data on homographs provides an innovative method that avoids requiring researchers to make subjective decisions about participants' intentions regarding meaning selection. This technique allows the participants to make the meaning selections, eliminating the subjectivity of a small number of researchers categorizing word associates. The technique also allows the flexibility to classify those associates into more than one meaning. Four follow-up experiments answered a number of questions that we had about this new technique.

In Experiment 2, we found that the number of meaning categories that we provided had no impact on the primacy of meaning dominance results. Although increasing the number of meaning categories presented to the participants did reduce the overall magnitude of the meaning dominance values calculated, the hierarchy of those meanings was unchanged.

In Experiment 3, we found that when we examined our participants' familiarity with the meanings that we provided, they were most familiar with the primary, then secondary, then tertiary, and finally quaternary meanings. However, even their lowest average familiarity rating (for quaternary meanings) was high (above 5 on a scale of 1-7). This indicated that the meaning categories that we chose to use in our norms collection were, in fact, adequately familiar to our participants.

In Experiment 4, we found that by allowing the participants to generate their own meanings for individual homographs they produced roughly the same meanings that we used. Although not statistically significant, they did produce slightly more meanings than we used in our original data collection. The purpose of our homograph data collection procedure was to eliminate the subjective component of having a small group of judges determine what meaning the participant intended by his or her response and instead allowing a large number of participants to indicate that meaning. Even with the differences in the methods (i.e., having participants provide two definitions per homograph, as was the case in Experiment 4, vs. having the meanings provided for them, as in our homograph data collection procedure) there was a significant correlation between the meanings that they produced in Experiment 4 and those that we used in our normative data collection procedure. Thus, the results of Experiment 4 provide evidence that the meaning categories used in the normative data collection were appropriate for the homographs used in the experiment.

In Experiment 5, we investigated the extent to which our findings would generalize to those in another Englishspeaking country—specifically, Australia. We found an extremely high level of agreement across our two samples (U.S. and Australian). It appears that the use of this particular technique for collecting homograph norms does generalize to at least one other English-speaking country.

The technique presented here for the development of homograph norms allows participants to make judgments with regard to which meaning of a homograph was intended by previous participant responses. It removes the subjective nature of enlisting a small group of researchers to determine which meaning or meanings a particular participant intended by his or her associates. The four follow-up experiments answer a number of questions about the procedure itself and provide the element of external validity, in that the results gathered in the U.S. were replicated in considerable detail by a sample of participants in Australia.

#### **AUTHOR NOTE**

The authors acknowledge the assistance from the following people, without whom data collection would not have been possible. Thanks to Tanya Ross, Beth Conrad, Marie Manna, Julie Adzimo, Nicole Fisher, Kimberly Merritt, Daniel Krawcyzk, Amy Moore, Cindy Hadzega, Carly Bicheller, Jessica Corrier, Sara Nease, Stephanie Bielecki, Denelle Newcomb, Laurie Whittermore, Meghan Ryan, Amanda Ross, and Kelly Guiderelli. This work was partially supported by a Scholarly Incentive Award from SUNY Fredonia. Correspondence concerning this article should be addressed to N. R. Gee, Department of Psychology, SUNY Fredonia, Fredonia, NY 14063 (e-mail: nancy.gee@fredonia.edu).

#### REFERENCES

- ARTHUR, A. Z. (1971). A comparison of Queen's and Minnesota Kent– Rosanoff word association norms. *Canadian Journal of Behavioural Science*, 3, 291-297. doi:10.1037/h0082271
- AZUMA, T. (1996). Familiarity and relatedness of word meanings: Ratings for 110 homographs. *Behavior Research Methods, Instruments,* & Computers, 28, 109-124.
- CARROLL, M., & SIMINGTON, A. (1986). The effects of degree of learning, meaning, and individual differences on the feeling-of-knowing. *Acta Psychologica*, 61, 3-16. doi:10.1016/0001-6918(86)90018-1
- CARTWRIGHT, D. S., & NICKERSON, C. A. (1979). An empirical thesaurus: Meaning norms for ninety common words. *Modern Language Journal*, 63, 442-447.
- CRAMER, P. (1970). A study of homographs. In L. Postman & G. Keppel (Eds.), Norms of word association (pp. 361-370). New York: Academic Press.

- DUFFY, S. A., MORRIS, R. K., & RAYNER, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory & Language*, 27, 429-446.
- EKPO-UFOT, A. (1978). Word associations: A comparative study among college students in Nigeria and the United States. *Journal of Cross-Cultural Psychology*, 9, 455-468. doi:10.1177/002202217894005
- GAWLICK-GRENDELL, L. A., & WOLTZ, D. J. (1994). Meaning dominance norms for 120 homographs. *Behavior Research Methods, In*struments, & Computers, 26, 5-25.
- GEE, N. R. (1997). Implicit memory and word ambiguity. *Journal of Memory & Language*, 36, 253-275. doi:10.1006/jmla.1996.2481
- GEIS, M. F., & WINOGRAD, E. (1974). Norms of semantic encoding variability for fifty homographs. *Bulletin of the Psychonomic Society*, 3, 429-431.
- GERARD, L. D., & SCARBOROUGH, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 305-315. doi:10.1037/0278-7393.15.2.305
- GILHOOLY, K. J., & LOGIE, R. H. (1980). Meaning-dependent ratings of imagery, age of acquisition, familiarity, and concreteness for 387 ambiguous words. *Behavior Research Methods, Instruments, & Computers*, **12**, 428-450.
- GORFEIN, D. S., VIVIANI, J. M., & LEDDO, J. (1982). Norms as a tool for the study of homography. *Memory & Cognition*, 10, 503-509.
- GOVE, P. B., ET AL. (EDS.) (1993). Webster's third new international dictionary of the English language, unabridged. Springfield, MA: Merriam-Webster.
- HALA, S., PEXMAN, P. M., & GLENWRIGHT, M. (2007). Priming the meaning of homographs in typically developing children and children with autism. *Journal of Autism & Developmental Disorders*, 37, 329-340. doi:10.1007/s10803-006-0162-6
- HERTEL, P. T., & EL-MESSIDI, L. (2006). Am I blue? Depressed mood and the consequences of self-focus for the interpretation and recall of ambiguous words. *Behavior Therapy*, **36**, 259-268. doi:10.1016/ j.beth.2006.01.003
- HUTCHISON, K. A., & BALOTA, D. A. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous critical lures. *Journal of Memory & Language*, 52, 1-28. doi:10.1016/j.jml.2004.08.003
- KAUSLER, D. H., & KOLLASCH, S. F. (1970). Word associations to homographs. Journal of Verbal Learning & Verbal Behavior, 9, 444-449. doi:10.1016/S0022-5371(70)80086-X
- NELSON, D. L., MCEVOY, C. L., & SCHREIBER, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved September 28, 1999, from http://web.usf.edu/ FreeAssociation/.
- NELSON, D. L., MCEVOY, C. L., & SCHREIBER, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36, 402-407.
- NELSON, D. L., MCEVOY, C. L., WALLING, J. R., & WHEELER, J. W. (1980). The University of South Florida homograph norms. *Behavior Research Methods, Instruments, & Computers*, **12**, 16-37.
- PALERMO, D. S., & JENKINS, J. J. (1964). Word association norms: Grade school through college. Minneapolis: University of Minnesota Press.
- PERFETTI, C. A., LINDSAY, R., & GARSON, B. (1971). Association and uncertainty: Norms of association to ambiguous words. Pittsburgh: University of Pittsburgh, Learning and Research Development Center.
- RAYNER, K., & DUFFY, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191-201.
- ROSENZWEIG, M. R. (1961). Comparisons among word association responses in English, French, German, and Italian. *American Journal of Psychology*, 74, 347-360.
- SEIDENBERG, M. S., TANENHAUS, M. K., LEIMAN, J. M., & BIEN-KOWSKI, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14, 489-537. doi:10.1016/0010 -0285(82)90017-2
- SERENO, S. C., PACHT, J. M., & RAYNER, K. (1992). The effect of meaning frequency on processing lexically ambiguous words: Evidence

from eye fixations. Psychological Science, 3, 296-300. doi:10.1111/ j.1467-9280.1992.tb00676.x

- SIMNER, J., GLOVER, L., & MOWAT, A. (2006). Linguistic determinants of word colouring in grapheme-colour synaesthesia. Cognition, 42, 281-289. doi:10.1016/S0010-9452(08)70353-8
- SWINNEY, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning & Verbal Behavior, 18, 645-659. doi:10.1016/S0022-5371(79)90355-4
- TITONE, D. A., & SALISBURY, D. F. (2004). Contextual modulation of N400 amplitude to lexically ambiguous words. Brain & Cognition, 55, 470-478. doi:10.1016/j.bandc.2004.02.073
- 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.
- WARREN, R. E., BRESNICK, J. H., & GREEN, J. P. (1977). Definitional dominance distributions for 20 English homographs. Bulletin of the Psychonomic Society, 10, 229-231.

Wollen, K. A., Cox, S. D., Coahran, M. M., Shea, D. S., & Kirby,

R. F. (1980). Frequency of occurrence and concreteness ratings of homograph meanings. Behavior Research Methods, Instruments, & Computers, 12, 8-15.

#### NOTES

1. All effects in this study are significant at p < .05, unless otherwise indicated.

2. Because 1 of the 37 homographs had five meanings, it could not be used in the analysis. In addition, another homograph was dropped from the analysis because of an error on the response sheets that made the participant responses unusable.

#### SUPPLEMENTAL MATERIALS

The U.S. and Australian homograph norms discussed in this article may be downloaded from http://brm.psychonomic-journals.org/content/ supplemental.

Homographs Included in the Present Study								
ABSTRACT	CROSS	FRUIT	OCCUPATION	SMEAR				
ACTION	CUT	FRY	ODOR	SNAKE				
APPEAL	DARK	GAG	OFF	SOCKS				
APPLICATION	DEGREE	GATHER	OFFENSE	SOFT				
ARM	DESERT	GENTLE	ON	SOIL				
AVERAGE	DISHES	GET	OUTSIDE	SOLID				
AX	DOCTOR	GONE	PART	SORRY				
BAN	DRAIN	GRAND	PIGEON	SPEND				
BANG	DRINK	GREAT	POCKETBOOK	SPIKE				
BANK	DRIVER	HANDLE	POWER	SPIRIT				
BARK	DRUG	HAUNT	PRINCIPLE	SPORTS				
BEAR	DRUNK	HIP	RAT	SQUEEZE				
BLANK	EGG	HIT	RATE	STAY				
BLAST	ELABORATE	HOOK	RETURN	STEEL				
BOLD	ELEMENT	HOP	RIM	STICKER				
BOOTH	ENGAGE	HOT	RIOT	STIFF				
BRASS	ENGINEER	HOUND	ROCKS	STOCK				
BRUSH	ENTERTAIN	IRON	ROUGH	STOMACH				
BUCK	ESCAPE	JOG	RULE	SUPPORT				
BULB	EXTRA	KIND	RUNNING	SWIM				
BUM	EXTREMEITY	KING	RUST	THICK				
BUREAU	FACE	LABEL	SADDLE	THREAD				
CALF	FAST	LACE	SAVE	THRESHOLD				
CAP	FIELD	LAND	SCARF	THROUGH				
CARRY	FINISH	LIMB	SENSITIVE	TICKET				
CHECKING	FISH	LOST	SERIAL	TOGETHER				
CLEAR	FLICK	LOW	SERPENT	TOY				
CLICK	FLOWER	MASTER	SHOULDER	TRAIL				
CLOUD	FOG	MATE	SHOW	TRAVEL				
COMB	FOUNDATION	MIRROR	SIMPLE	WAG				
COUNT	FRAME	NEED	SKETCH	WAKE				
CRADLE	FREE	NIGHT	SLAY	WELFARE				
CROOKED	FROG	OBSERVE	SLIPPERY					

# **APPENDIX** A

Overlap in Primary Meaning Between the Present Study and Previous Research								1				
Homograph	Gee	1	2	3	4	5	6	7	8	9	10	11
APPEAL*	attraction					D						
"	attractive quality					D						
BARK	dog			S			S		S		S	D
BAT*	animal	D	D	D	D		D	S	D		D	
BEAM	light		S		S	D	D		S	D	S	
BEAR	animal			S			S	D			S	
BIT <sup>*</sup>	little		D		D					S	n/a	
"	bite		D		D					n/a	S	
BRASS*	musical instrument					D						
BRUSH <sup>*</sup>	hair					D						D
CALF	COW		S						S		S	D
CAP	cover						D					
COMPANY	business					D				S	S	
COURT	justice		D				S				S	
CROSS*	Christianity									D		D
"	figure									D		D
DEGREE*	diploma					D						
"	measure					D						
DESERT	dry						D					
DIAMOND	ring						D		D		D	
FIELD*	area						D					
FREE <sup>*</sup>	liberate										D	
GAG	cause to heave						D			D	D	
IRON*	metal				S	D		S	S		D	
LAND <sup>*</sup>	earth					S						D
LEAD	metal	S	S			S		D	D		D	
LIGHT	illumination			S			S	D			S	
LOW*	depressed									S	D	
"	inferior									n/a	D	
ORGAN	body part			D	D	D	D	D	D		S	
PART*	portion					S					D	
PICK	to choose							D			D	
POOL*	water						D			S	D	
PRINCIPLE	highest rank					D						
RATE	fixed charge					D				D		
SIGN <sup>*</sup>	symbol						S			D	D	
STAR	sky						D			D	D	
STORE	merchandise						S				D	
STRAW	drink						D				D	
VAULT	chamber						S			S	D	
WATCH	observe			D	D		D	D		D	D	D
VARD	lawn					D			D		S	

**APPENDIX B** 

Note—1, Kausler & Kollasch, 1970; 2, Cramer, 1970; 3, Perfetti, Lindsay, & Garson, 1971; 4, Geis & Winograd, 1974; 5, Gilhooly & Logie, 1980; 6, Nelson, McEvoy, Walling, & Wheeler, 1980; 7, Wollen, Cox, Coahran, Shea, & Kirby, 1980; 8, Gorfein, Viviani, & Leddo, 1982; 9, Gawlick-Grendell & Woltz, 1994; 10, Twilley, Dixon, Taylor, & Clark, 1994; 11, Azuma, 1996; D, different primary meaning; S, same primary meaning.

> (Manuscript received August 10, 2009; revision accepted for publication April 10, 2010.)