Psychological norms for simple three-line graphic shapes

By: William A. Kealy, Albert D. Ritzhaupt, and Cameron Spears

Kealy, W. A. Ritzhaupt, A. D., & Spears. C. (2007). Psychological norms for simple three-line graphic shapes. *Behavior Research Methods*, 39(4), 1012-1019.

Abstract:
There is a longstanding tradition in psychological research for norming lists of words that are used in experimental studies. The present study extends this practice to graphic imagery by obtaining norming data on 24 simple abstract graphic shapes composed of three straight-line segments. The attributes obtained in the norming procedure were the shapes' familiarity, describability, associability, availability, and potential for word association. Results from rating data indicate significantly different, yet reliable, responses by participants to the various shape configurations. Multidimensional scaling analysis of shape ratings identified two underlying dimensions of perceived differences: the continuity of a shape's linear direction and the consistency or regularity of its interior angles. By contrast, performance in generating word associations for figures appeared to be linguistically driven, with initial responses related to the similarity of shapes to letters of the alphabet. The norms and the computer program used to collect them can be downloaded from www.psychonomic.org/archive.

Article:
Throughout the history of psychology, there has been a venerable tradition of establishing norms for experimental materials used in research. The overwhelming majority of these efforts have been directed to achieving norms for verbal material, whether fictitious (Ebbinghaus, 1913) or factual (Jenkins, 1970; Nelson, McEvoy, & Schreiber, 1998; Thorndike & Lorge, 1944).

In sharp contrast with the many efforts to norm linguistic information, there have been relatively few attempts to create comparable norms for graphic images. To illustrate this point, of the 81 databases available in the Psychonomic Society Archive of Norms, Stimuli, and Data (2005), only 7 pertain to graphic images. Here, as in most other instances, the normalized sets of images consist of pictures of everyday objects (Alario & Ferrand 1999; Nishimoto, Miyawaki, Ueda, Une, & Takahashi, 2005; Snodgrass & Vanderwart, 1980). One recent exception to this was a study by McDougall, Curry, and de Bruijn (1999), in which 239 icons and symbols were rated on a variety of traits, including concreteness, complexity, familiarity, and meaningfulness.

Our work was a departure from most of the previous attempts to norm graphic images, in that the stimuli used in the present study were abstract three-line shapes. A primary motivation for the present study arose from our previous research (Bunch, Seyferth, Williams, Haun, & Kealy, 2004) on spatial mnemonics (see Bellezza, 1983, 1986), which suggests that some abstract shapes are more easily associated with superimposed target words than others. Hence, an important purpose of our present research was to gather psychological norms on a large set of abstract shapes that would allow replication of this earlier work while controlling for intervening variables such as a shape's capacity for verbal association.

In formulating our approach to systematically establishing norms for abstract shapes, we were guided by the previous efforts of Clark and Paivio (2004) in creating norms for 2,311 words. Specifically, like these researchers, we sought measures that reflected the stages of processingrepresentational, referential, and associative-suggested by dual coding theory (Paivio, 1971, 1986; Sadoski & Paivio, 1994). Representational processes elicited by a word or image involve activation of an internal representation of the stimulus presented. By contrast, referential and associative processes involve internal generation of alternative mental representations that are, respectively, either of a different type (e.g., images eliciting words) or the same type (e.g., images eliciting images).

Since our overarching concern was the use of abstract shapes in the context of spatial mnemonics, a rationale existed for identifying tasks that called upon all three stages of processing. We reasoned that both a shape's capacity for being mentally represented and its facility for prompting related words and images might conceivably contribute to its use as a mnemonic aid. Hence, we derived tasks for the study that evoke representational, referential, and associative processes. In the case of representational processes, participants rated the ease with which a given image could be described (i.e., describability), since this trait probably corresponds well with image identification and recognition. For the same purpose, another task had participants rate the familiarity of a given shape. However, as Clark and Paivio (2004) point out, familiarity probably influences referential and associative processes as well, since familiar images are more likely to mentally prompt, respectively, meaningful words and pictures than unfamiliar images. Our study also explored associative processing through an availability task, whereby participants rated how well a given image brought other images to mind. Finally, we developed two tasks related to referential processing. One had participants rate the number of word associations (i.e., associability) prompted by a particular image, whereas another had participants actually generate word associations for a given image by typing as many words as they could during a 30-sec period.

METHOD

Participants

A total of 67 undergraduate education majors at a university in the southeastern United States took part in the study, earning extra course credit for their contribution. Although there were no between-subjects treatments involved, we randomly assigned participants to workstations, to minimize the chance of socializing during the sessions.

Materials

Using a 2 × 2 square matrix as a design template, we generated all possible shapes composed of three continuous straight-line segments whereby the vertices of each fell on the intersection points of the underlying grid. Additionally, none of the images thus constructed resulted in a closed shape. The fourth shape from the top left (see Figure 1) illustrates our procedure for generating shapes. In this manner, as Figure 1 indicates, we created a total pool of 24 shapes.

For the study, we developed a computer program that instructed participants about the research session procedures, provided practice items, and presented both the experimental stimuli and the mechanism for making responses.

To aid participants on the four rating tasks, we developed a computer interface that presented them with an onscreen slide for which they used the mouse to drag a short vertical bar (the slider) to the left or right of a rectangular box (the slide). Movement of the slider between the bipolar adjectives produced continuous mathematical values to the 11th decimal place that corresponded with the slider's location at the moment the Tab key was pressed. An illustration of the rating interface is shown in Figure 2.

For the study, we used a university classroom containing 25 computer workstations running the Microsoft Windows XP Professional operating system. Each computer was equipped with a 15-in. (38.1-cm) flat panel liquid crystal display configured at a video resolution of 1,024 × 768 pixels. We installed the completed programs on 12 computers distributed throughout the room, leaving at least one nonfunctioning computer between the workstations used by the participants.

Procedures

As participants arrived for an experimental session, they were assigned to computer workstations showing a blank gray screen. Participants were briefed on the general purpose and tasks of the study, received answers to any existing procedural questions, and began the program.

Participants then saw the words "Familiarity Ratings" and, just above the title, the words "Part 1 of 5." In the center of the screen appeared two columns of text, in black, 18-point Arial Narrow bold lettering, with the left-hand column stating:

Shapes can differ in their apparent degree of familiarity to viewers. Some shapes may seem more commonplace than others because they appear in the objects, signs, and images encountered in everyday life. Other shapes, by contrast, may seem relatively unfamiliar, novel, and unique. For example, consider the following two shapes:

Immediately below this column appeared two shapes, one a perfect circle and the other a truncated ellipse rotated to a 45° angle. The right-hand column continued:

The shape on the left, a circle, is commonly seen, whereas the one on the right is relatively more unusual. In a moment, you'll be rating 24 simple shapes, one at a time, on each one's degree of familiarity. Very familiar shapes are those that seem common to you while very unfamiliar shapes are those that you think are unusual.

As soon as they had completed reading the text, participants clicked the button at the bottom center of the screen marked "Click to Continue," to go to the next screen. After they had done this, all but the titles at the top of the screen vanished and the computer presented two new columns of text. This new screen informed participants that they would be making their ratings by using the mouse to "click and drag" a slider bar similar to the one that appeared at the bottom of their screens below the text.

The instructions told participants to practice moving the slider bar back and forth. The text further stated that when actually making a rating, they were to press the Tab key to record the position of the slider, scoring a rating between 1 (very unfamiliar) and 7 (very familiar). Below the text appeared an example of the rating interface they would be using. At the left and right ends of the rectangle, respectively, were the words very unfamiliar and very familiar, printed in black 18-point Arial bold capitals. Just below the left and right edges of the rectangle were the numbers "1" and "7" in blue 36-point Arial bold lettering. A button centered at the very bottom of the screen prompted participants to start the first task when ready.

Immediately after participants clicked the button to begin the rating task, everything on the screen disappeared except the slider bar annotated with the word and numerical labels. Above the slider bar, and at the center of the screen, appeared the first of 24 shapes, presented in random order, for rating. The height of each shape was almost exactly one third the height of the screen. Figure 2 is a cropped screen capture that illustrates the typical rating screen layout.

After participants completed rating all 24 items, the screen erased completely and the title of the next rating assignment appeared at the top of the screen accompanied by an orientation to the new task similar to the one that they had read earlier. In this manner, participants viewed an introduction with an example for each of the rating dimensions prior to each rating activity.

When the final rating task was completed, the screen erased to present a new title, "Word Associations," with the words "Part 5 of 5" printed above. The two columns of text below the title told participants that the purpose of this section was to identify the items or objects associated with each of the 24 shapes. Each shape, they were told, would be presented one at a time for 30 sec; during that period, they were to type as many items as possible that they thought were associated with the image shown. The instructions explained that they should type the items as they came to mind, and not in any particular order. As soon as the 30-sec period expired, the computer program informed participants that the next shape would replace the shape on the screen. As participants progressed to the next screen, they viewed an example of the task they were being asked to perform and the small clock-like icon at the top of the screen that indicated the passing time. Once participants clicked the button to begin the word association task, the screen depicted in Figure 3 appeared. As each successive shape appeared in random order on the left side of the screen, participants typed as many word associations as they could, pressing the Enter key after every word to reposition the cursor for the next entry.

When the last shape disappeared from the screen, the computer debriefed participants on the session and provided contact information for the experimenters if any questions arose. Each participant left the room upon completing the program and logging off the computer. A typical session lasted approximately 30 min.

RESULTS AND DISCUSSION

Out of concern for typical rater errors of leniencyseverity and central tendency (Saal, Downey, & Lahey, 1980), we first performed a boxplot analysis of rating variance for each participant on the four rating dimensions. This identified four outliers, whose ratings on the describability dimension showed extraordinarily high variance: 8.04.8.09,7.87, and 7.95, compared with a mean variance of 3.44. Examination of scores by these participants revealed a preponderance of "1" and "7" ratings across each rating category. Consequently, these participants were excluded from the remainder of the analysis.

Reliability Estimates

For each of the five constructs we measured for the shapes, we calculated the reliability for the participants on the 24 items. The Cronbach's alphas for familiarity, associability, describability, availability, and number of word associations were, respectively, .88, .89, .87, .85, and .95. This indicated that there was a high degree of internal consistency in the responses of participants for each of the traits measured.

Comparison of Ratings

Means and standard deviations for ratings-within a range of 1 to 7- for the 24 normed shapes on four of the five dimensions scored (familiarity, associability, describability, and availability) appear in Table 1. We averaged the means and standard deviations across all four rating categories to derive an overall rating for shapes, using this metric to vertically arrange the shapes from highest rated ("Z," Shape 1) to lowest rated (Shape 24). Generally, shapes that formed capital letters (i.e., "Z," "U," and "N") received the highest ratings, whereas the three-sided "bracket" shapes with one side bent (Shapes 22, 23, and 24) were rated the lowest by participants. Overall, participants gave shapes higher ratings on the describability dimension ($M = 4.54$, $SD = 1.27$) than for associability ($M = 3.80$, $SD= 1.46$).

We further studied these descriptive statistics, and in particular the distribution of ratings, by constructing a stemand-leaf plot of mean ratings for the 24 graphic shapes that participants rated on four dimensions (see Figure 4). Unlike a traditional stem-and-leaf plot, however, this display substituted numerical values of the shapes' means with small icons representing the shapes themselves. This revealed for each of the four criteria, a distribution of ratings that was right-skewed with clusters of more than half of the scores centering at roughly 3.4 on the scale-almost a full point below the mean. The plot also suggests a second clustering of ratings at the high end of each scale that contributed to the elevated means relative to the median scores. The source of these suggested bimodal distributions, however, is difficult to discern, because they are the product of variability among both the shapes rated and the raters themselves. Given the high rater reliability on each rating dimension, however, it is more likely that the unusual distributions of ratings reflect perceptions by all raters of two distinct classes of shapes-one rated low and the other rated high across all four categories.

To examine the influence of both shape and rating category on the rating behavior of participants, we subjected the data to a 4 (category; i.e., the four criteria on which the shapes were rated) × 24 (shape) repeated measures

ANOVA. The analysis showed significant main effects for both category [$F(3,248) = 12.03$, $p < .01$, $d = 1.0$] and for shape [$F(23, 5,704) = 204.01$, $p < .01$, $d = 1.0$]. ANOVA also identified a significant [$F(69, 5,704) = 3.77$, $p < .01$, $d = 1.0$] category × shape interaction. Essentially, these findings validate our judgment from the descriptive statistics that participants perceived both the differences between the various rated shapes as well as differences between rating categories. The origin of the significant interaction between category and shape is evident in Figure 4, where, for example, the shape rated highest on associability (i.e., the upside-down "U") was rated a full point below the highest-rated shape for familiarity and describability (the letter "Z").

Correlation Between Rating Constructs

Given the aforementioned high reliability of rater behavior within the four constructs rated we examined the relationship between constructs by first calculating, for each one, mean ratings for every one of the 24 shapes. Using these mean ratings, we then computed Pearson correlations on all possible pairs of constructs. This yielded six R values that ranged between .94 and .98 and represented significant correlations ($p < .05$). We also identified significant Pearson correlations between the mean number of words generated for each shape during the word association task and their corresponding ratings for each of the four constructs. Not surprisingly, higher correlations were noted between word associations prompted by the shapes and their ratings for associability ($R = .84$) and availability ($R = .85$), compared with their correlations with familiarity ($R = .75$) and describability ($R = .72$) ratings. Explicably, the acts of generating word associations from shapes and rating shapes on their degree of associability involve the same referential processes. Similarly, we expected high correlations between familiarity and describability rating tasks that are primarily representational in nature. Other highly correlated rating categories, such as familiarity and describability ($R = .96$), are less clearly explained and probably reflect, as Clark and Paivio (2004) have noted activities that involve more than one stage of processing.

Multidimensional Scaling Analysis

Examination of the stem-and-leaf plot (Figure 4) suggested that participants perceived naturally forming categories or groups within the pool of 24 shapes. To explore this further, with the possibility of identifying underlying factors by which participants based their perceptual judgment, we performed a multidimensional scaling (MDS) analysis of the rating data. We applied an Alscale procedure that created Euclidean distances from the mean ratings across the four rating categories. Since the stress values and the squared correlations for the 2-dimensional and 3-dimensional solutions were almost identical, we analyzed the former (see Figure 5) for the sake of parsimony.

Four compact groups of shapes, labeled Group I, Group II, Group III, and Group IV, emerged from the analysis. Initially, we speculated that rating differences globally represented the distinction between shapes that were either alphabetical or nonalphabetical in form. However, this fails to explain why the "XT' and "C" shapes in Group II are set apart from the "N" and "Z" shapes in Group IV. An alternative explanation is that the former shapes and their mirror images are all "trifold" shapes formed by a straight line bent twice at right angles in the same direction. By contrast, the latter shapes of Group IV are all "accordion" folds, resulting from bending each successive line segment in the opposite direction. This distinction also corresponds to shapes in Groups I and III, the clusters of shapes immediately to the left of Groups II and IV, respectively. The former groups, however, differ from the latter, in that the two interior angles of each shape within Groups I and III are unequal, whereas the angles in each shape within Groups II and IV are the same. Consequently, each shape in Groups II and IV is identical to the shape created by its horizontal and vertical mirror image, with the result that there are only half as many shapes in these groups combined as there are in Groups I and III.

As Figure 5 shows, the shapes on the far left are distorted and graphically compact versions of those on the far right. Shapes at the top half of the display, meanwhile, consist of shapes made by two successive bends of a line in the same direction; those in the bottom half have bends in opposite directions. Accordingly, Figure 5 suggests two dimensions by which participants may have perceived the shapes they rated: continuity (y-axis Dimension

1), representing shapes having either a broken or continuous line direction, and consistency (x-axis Dimension 2), referring to shapes having either irregular or uniform interior angles. From this, we concluded that perceived differences among shapes are probably due to more factors than simply whether or not their designs are alphabetic. Nevertheless, shapes are probably also distinguished by their similarity to letters of the alphabet. To study this, we examined the word associations prompted by alphabetic and nonalphabetic shapes.

Word Associations Cued by Shapes

In the study, participants generated as many word associations as possible for each of the 24 shapes shown during a 30-sec period This resulted in 4,747 short-answer responses from 63 participants (after outliers were removed). From this pool, we deleted 60 entries that were unreadable fragments of words. We partitioned the remaining valid entries into four categories, depicted in Table 2, following a levels model of mental processing (Craik & Lockhart, 1972): (1) associations reflecting the surface features of shapes (letters, Roman numerals, mathematical symbols, etc.), (2) responses that were phonologically based on a letter suggested by a given shape (e.g., "zebra," "Zorro," and "zipper," prompted by the "Z" shape), and (3) associations at the deeper semantic level that were further divided into nouns and nonnouns. Using an Excel spreadsheet, we arranged all the words generated so that the 24 columns corresponded to the shapes, whereas the rows represented participants. The number of rows assigned to a given participant equaled the largest number of words that person produced for a shape, with every response occupying its own cell. Working separately, two researchers then gave each cell a different color, depending on the category of the word it contained: surface feature, phonological, noun, or nonnoun. In situations where the identification of a word as a noun or nonnoun was ambiguous, we based our decisions on the context created by the participants other responses. If word associations produced by a participant for a given shape were, for instance, "jump," "climb," and "hold" then the response "rock" would be considered a nonnoun (a verb), whereas if the responses were, respectively, "cliff," "ladder," and "cup," the word "rock" would be interpreted as a noun. Once completed the color coding of the two spreadsheets was compared revealing that the two judges agreed on the category assignment of participants' responses 93% of the time.

Intuitively, we anticipated that shapes resembling letters would produce the greatest number of word associations. However, the data revealed that the highest number of word associations were prompted by shapes S02 and S04, both yielding an average of 4.6 words per person. These were clearly interpreted as nonwords, based on the relatively modest number of responses at the surface or phonological levels. Shape SOI ("Z"), by contrast, evoked roughly three times the number of surface and phonologically based associations, while prompting only about one third the number of nouns. A similar outcome was evident among the word associations cued by the "N" shape (S03).

Our earlier analysis showed that word association productivity correlated significantly with each of the rating dimensions. This is clearly evident from the fact that the same shapes rated highest were also among those that evoked the greatest number of word associations. Nevertheless, the nature of the word associations from highly rated shapes differed considerably. Despite its alphabetic appearance, the "U" shape (S04) elicited the highest number of nouns, and these varied widely in semantic content (e.g., "magnet," "wastebasket," "beaker"). From examples such as these, we hypothesize that nonalphabetic shapes may possess an inherently richer potential for generating associations than shapes that are readily identified as letters.

An important accomplishment of this study was its demonstration of the feasibility of obtaining norms for abstract shapes. The study also presented a methodology for doing so based on the recent work of Clark and Paivio (2004) in establishing norms for words. Although the number of participants in the present study is modest in comparison with typical attempts at establishing psychological norms, we believe that this work nevertheless represents a useful starting point for several avenues of future research. We are especially eager to use the normed shapes for our continued research on spatial mnemonics. This normed data will facilitate better control of extraneous variables introduced by the shapes used for mnemonic displays. Additionally, the data reported here will allow us to study the effect of shapes with varying degrees of cuing potential on the efficacy of spatial mnemonics that are based on these forms. Another research possibility, and one that we are currently exploring, is to extend the scope of the present work to that of four-line shapes. Additional work is also required to identify the classes of objects found in the word associations for different shapes. Ideally, it may be possible to determine the relative salience of words that are evoked by a particular shape. There are many other research questions and opportunities, currently not envisioned by us, that could conceivably employ the data and materials resulting from this study. We hope other researchers may benefit from these possibilities.

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 (Kealy), and the publication year (2007).

FILE: KEALY-BRM-2007.zip

DESCRIPTION: The compressed archive file contains five files:

Table 1 .txt, containing tab-delimited data from Table 1.

Tablel.xls, containing the above information in Excel spreadsheet format.

Table2.txt, containing tab-delimited data from Table 2.

Table2.xls, containing the above information in Excel spreadsheet format.

FIGN0RM1024x768.a4p, containing the editable Authorware program used to collect the norms.

REFERENCES

ALARIO, E-X., & FERRAND, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. Behavior Research Methods, Instruments, & Computers, 31,531-552.
BELLEZZA, F. S. (1983). The spatial-arrangement mnemonic. Journal of Educational Psychology, 75, 830-837.
BELLEZZA, F. S. (1986). A mnemonic based on arranging words on visual patterns. Journal of Educational Psychology, 78,217-224.
BUNCH, J. M., SEYFERTH, T., WILLIAMS, D., HAUN, D., & KEALY, W. A. (2004, April). Symmetrical or not, are "spatial mnemonics " credible aids for learning? Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.
CLARK, J. M., & PAIVIO, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. Behavior Research Methods, Instruments, & Computers, 36, 371-383.
CRAIK, F. I., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior, 11,671-684.
EBBINGHAUS, H. (1913). Memory: A contribution to experimental psychology (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College, Columbia University. (Original work published 1885. Retrieved October 11,2005 from psychclassics.yorku.ca/Ebbinghaus/index.htm)
JENKINS, J. J. (1970). The 1952 Minnesota word association norms. In L. Postman & G. Keppel (Eds.), Norms of word association (pp. 1-38). New York: Academic Press.
MCDOUGALL, S. J. P., CURRY, M. B., & DE BRUUN, O. (1999). Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningftilness, familiarity, and semantic distance for 239 symbols. Behavior Research Methods, Instruments, & Computers, 31,487-519.

NELSON, D. L., MCEVOY, C. L., & SCHREIBER, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved October 02,2005 from www.usf.edu/FreeAssociation/.
NISHIMOTO, T, MIYAWAKI, K., UEDA, T., UNE, Y., & TAKAHASHI, M. (2005). Japanese normative set of 359 pictures. Behavior Research Methods, 37,398-416.
PAIVIO, A. (1971). Imagery and verbal processes. New York: Holt, Rinehart & Winston.
PAIVIO, A. (1986). Mental representations: A dual coding approach. New York: Oxford University Press.
SAAL, F. E., DOWNEY, R. G., & LAHEY, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. Psychological Bulletin, 88,413-428.
SADOSKI, M., & PAIVIO, A. (1994). A dual coding view of imagery and verbal processes in reading comprehension. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), Theoretical models and processes of reading (4th ed., pp. 582-601). Newark, DE: International Reading Association.
SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Learning, Memory, & Cognition, 6, 174-215.
THORNDIKE, E. L., & LORGE, I. (1944). The teacher's word book of 30,000 words. New York: Teachers College, Columbia University.