

A cross-culturally standardized set of pictures for younger and older adults: American and Chinese norms for name agreement, concept agreement, and familiarity

CAROLYN YOON and FRED FEINBERG
University of Michigan, Ann Arbor, Michigan

TING LUO
Chinese Academy of Sciences, Beijing, China

TREY HEDDEN, ANGELA HALL GUTCHESS, HIU-YING MARY CHEN, and JOSEPH A. MIKELS
University of Michigan, Ann Arbor, Michigan

SHULAN JIAO
Chinese Academy of Sciences, Beijing, China

and

DENISE C. PARK
University of Illinois at Urbana-Champaign, Urbana, Illinois

The present study presents normative measures for 260 line drawings of everyday objects, found in Snodgrass and Vanderwart (1980), viewed by individuals in China and the United States. Within each cultural group, name agreement, concept agreement, and familiarity measures were obtained separately for younger adults and older adults. For a subset of 57 pictures (22%), there was equivalence in both name agreement and concept agreement, and for an additional subset of 29 pictures (11%), there was nonequivalent name agreement but equivalent concept agreement, across all culture-by-age groups. The data indicate substantial differences across culture-by-age groups in name agreement percentages and number of distinct name responses provided. We discovered significant differences between older and younger American adults in both name agreement percentages (67 pictures, or 26%) and concept agreement percentages (44 pictures, or 17%). Written naming responses collected for the entire set of Snodgrass and Vanderwart pictures showed shifts in both naming and concept agreement percentages over the intervening decades: Although correlations in name agreement were strong ($r = .71, p < .001$) between our younger American samples and those of Snodgrass and Vanderwart, name agreement percentages have changed for a substantial proportion (33%) of the 260 pictures; moreover, 63% of the stimuli for which Snodgrass and Vanderwart reported concept agreement now appear to differ. We provide comprehensive comparison statistics and tests for both the present study and prior ones, finding differences across numerous item-level measures. The corpus of data suggests that substantial differences in all measures can be found across age as well as culture, so that unequivocal conclusions with respect to cross-cultural or age-related differences in cognition can be made only when appropriate stimuli are selected for studies. Data for all 260 pictures, for each of the four groups, and all supporting materials and tests are freely archived at http://agingmind.cns.uiuc.edu/Pict_Norms. The full set of these norms may be downloaded from www.psychonomic.org/archive/.

Pictorial stimuli have long been used in neuropsychological investigations, as well as in research studies in-

volving memory and other cognitive processes. Different attributes of a picture, such as object or picture familiarity (Lachman & Lachman, 1980), or word frequency (Oldfield & Wingfield, 1965), are known to correlate with cognitive measures and to affect both memory and retrieval processes.

Picture norms of everyday objects, in particular, are widely used by researchers in cognitive psychology. For example, such pictures have been used in studies of how images and visual-spatial representations differ from

This study was supported by Grant R01 AG015047 from the National Institute on Aging. We thank Dick Nisbett, Phaythoune Chothmounethinh, Suzanne Hambricht, Ping Hu, Alice Huang, Serena Salloom, Cat Woods, and Vickie Yeung. Correspondence concerning this article should be addressed to Carolyn Yoon, University of Michigan Business School, 701 Tappan Street, Ann Arbor, MI 48109-1234 (e-mail: yoonc@umich.edu).

verbal or abstract representations in memory (e.g., Kirsner, Milech, & Stumpf, 1986; Snodgrass, 1984). They have also been used to investigate the effects of picture priming on implicit and explicit memory (e.g., McDermott & Roediger, 1994; Mitchell & Brown, 1988; Rajaram, 1996). In cognitive neuroscience, researchers have employed picture norms to examine the nature of representational systems underlying visual memory in normal adults, as well as those with deficits and impairments (e.g., Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000; Nyberg, Cabeza, & Tulving, 1996; Stark & Squire, 2000). In studies of cognition and aging, picture norms have helped elucidate differences in visual perception and memory across the life span (e.g., Park, Smith, Morrell, Puglisi, & Dudley, 1990; Parkin & Streete, 1988). An extensive index of norms and ratings appearing in journals of the Psychonomic Society for the last 40 years appears in Proctor and Vu (1999).

Despite researchers' extensive reliance on picture norms, validation data on the norms have been relatively scarce since Snodgrass and Vanderwart (1980) introduced their set of 260 pictures standardized for name agreement for young American adults. Snodgrass and Yuditsky (1996) compared voice naming times and keypress naming times for 250 of the original Snodgrass and Vanderwart pictures and found them to be similarly sensitive indicators of naming difficulty. Although Snodgrass and Yuditsky reported the percentages of pictures that were correctly generated by study participants, they did not attempt to validate their data against the Snodgrass and Vanderwart norms (which were based on written responses). More recently, Bates and colleagues (Bates et al., 2003; Székely et al., 2003) have, as part of their Center for Research in Language project, collected timed picture naming data among younger adults in seven languages/countries, including English and Chinese. They collected data for 520 pictures (176 of them from Snodgrass & Vanderwart) in each country and found evidence of cross-linguistic similarities and differences across the different languages. Székely et al. further compared the data for younger Americans collected by Bates et al. with the data of Snodgrass and Vanderwart, and Snodgrass and Yuditsky on a subset of 161 overlapping pictures, and concluded that there were comparable levels of name agreement across the three studies. However, because Székely et al. drew primarily on analyses of summary statistics aggregated across overlapping items to conclude that there was comparability, their results may have masked any differences existing at the item level. Thus, one of the goals of the present study was to test and identify pictures for which the norms have changed substantially for younger Americans since Snodgrass and Vanderwart reported them in 1980.

The second goal was to ascertain the applicability of such picture norms to an older population. In prior studies that have utilized pictorial stimuli to investigate age differences, researchers appear largely to have presumed that younger and older adults do not differ in terms of the names they assign to objects.¹ In this study, we in-

vestigate the validity of this assumption and suggest that there may be cohort differences in naming certain objects.

Our third and final goal was to identify pictures suitable for use in studies investigating age and/or cross-cultural differences in cognition. The recent emergence of research interest in cross-cultural differences in cognition, particularly between East Asian and Western cultures, underscores the need for culture-invariant stimuli. Although pictures of everyday objects are a potentially useful source of stimuli for studies comparing East Asians and Americans (Park, Nisbett, & Hedden, 1999), some objects are likely to vary in familiarity across cultures. For example, certain animals and vegetables that are indigenous to the U.S. (e.g., raccoons, and asparagus) are virtually unknown in East Asia, except perhaps through foreign media or literary depictions. Moreover, there are objects for which naming responses may vary for different age groups within a culture (e.g., older American adults, compared with their younger counterparts, may have greater experience with tools such as chisels and pliers). It is therefore important that researchers develop normative data to guide the selection of pictorial stimuli for studies of cross-age and -cultural differences between East Asians and Americans.² Comparisons of picture naming data from younger American and younger Taiwanese adults conducted by Bates et al. (2003) provide a useful resource for designing cross-cultural studies. However, although it may be possible to use these data as a basis for implementing studies anywhere Mandarin is spoken, orthographic differences (e.g., a completely distinct pictographic system) call into question their application in mainland China and in experiments requiring written rather than vocal responses.

In this article, we report, for both American and Chinese groups, name agreement, concept agreement, and familiarity measures on the 260 Snodgrass and Vanderwart (1980) pictures, for younger adults and older adults separately. We compare the norms for younger American adults in the present study with those reported by Snodgrass and Vanderwart, Snodgrass and Yuditsky (1996), and Bates et al. (2003), and identify pictures for which there are significant cross-study differences in naming and familiarity measures.

We also compare norms for younger and older Americans, as well as those for younger and older Chinese, and indicate which pictures reflect low name and/or concept agreement in each culture. The present article, to the best of our knowledge, is the first to provide separate norms for older adults despite the fact that pictured materials are commonly used in patient populations in which older adults are disproportionately represented. The resulting older adult norms thus provide a valuable resource for researchers needing to construct appropriate materials to test older patients and seeking to investigate age-related differences in cognition.

Finally, we assess name and concept agreement across all culture-by-age groups and identify a subset of 57 pictures for which there is across-the-board equivalence.

Our data complement those of Bates et al. (2003) and are expected to be useful to a broad range of researchers, from those concerned with processes underlying visual memory in general to those interested in cross-cultural variation.

METHOD

Participants

One hundred thirteen younger adults (17–25 years) from the University of Michigan and 103 community-dwelling older adults (60–75 years), constituting the American cultural group, were recruited for testing in Ann Arbor, Michigan. One hundred younger Chinese students (18–23 years) recruited from three universities (Beijing Normal University, Capital Normal University, and Aeronautics and Space University) and 100 older Chinese adults (59–76 years) from Beijing were tested at the Institute of Psychology, Chinese Academy of Sciences. Summary age statistics for all groups appear in Table 1.

Comparisons of sample characteristics revealed that education levels and self-reported health ratings across the groups were generally equivalent (see Table 1 for means). All four samples were well-educated with mean (and median) levels of secondary (college or university) education. All groups rated their general health as being better than average (greater than 3 on a 5-point scale). All younger participants were college students. Eighty-three percent of the older American adults and 91% of older Chinese adults were retirees, with no difference in the average number of years since retirement ($M = 7.6$ and 7.3 years for Americans and Chinese, respectively). The remaining older adults reported working either part time or full time in various professional occupations. Specific information on other sample characteristics (e.g., gender, ethnicity, marital status) appears in Appendix C of the Archived Materials.

Stimuli and Procedures

All 260 standardized pictures developed by Snodgrass and Vandewart (1980) were tested in the present study. Experimental sessions were conducted in groups of 10 to 25 participants. At the start of each session, participants were given general instructions, followed by several practice examples. Chinese participants were given both verbal and written instructions in Mandarin, whereas the Americans were provided with equivalent instructions in English.

Each of the 260 pictures was projected, one at a time, in random order. The pictures were black outline drawings on a white background, projected on a screen within a slide presentation. The slide projector was situated 12 ft from the screen, and participants sat next to and behind the projector, at a distance from which they could clearly see the pictures. Every object projected on the screen was centered and the pictures were displayed by a standard digital projector at 800×600 pixels resolution. Lighting in the room was slightly dimmed, but bright enough for participants to read the instructions and comfortably write their responses.

Each picture was presented for 8 sec followed by a 2-sec pause. Participants were instructed to write down “the first name of the object that comes to mind” for every picture presented. They were further instructed to respond with an “X” if they thought they had encountered the object before but did not know its name (i.e., what it is called), or could not remember the object’s name (don’t know name—DKN). If they did not know what the object was, they were to respond with an “O” (don’t know object—DKO).

Participants were also asked to judge the familiarity of each object featured in a picture immediately after having written that object’s name.³ In the space provided to the far right on the same line as their written response, participants were instructed to rate the object for familiarity, based on the degree to which they come into contact with or think about the object. Specifically, they rated each pictured object for familiarity on a 5-point rating scale, with 1 indicating *the object is not at all familiar to me* (e.g., never come into contact with the object) and 5 indicating *the object is highly familiar to me* (e.g., very frequently come into contact with the object).

Participants recorded their responses in individual paper booklets and were instructed to respond to every picture and not to leave any blanks. After being presented with 140 pictures, they were given a 5-min break. After the break, they viewed the remaining 120 pictures.

Following the main task, participants were administered two short questionnaires, one pertaining to demographic information and the other to health status. Upon completion of the questionnaires, participants were debriefed, thanked for their participation, and dismissed. Younger Americans received credit in the form of partial fulfillment of their subject pool requirements in a psychology course. All other groups received monetary compensation for their participation in the study. The entire session lasted approximately 75 min.

Analyses

Written name responses were first coded for all 260 pictures and then compiled for each culture-by-age group. Files containing these naming data for each of the four groups—Younger American, Older American, Younger Chinese, and Older Chinese—are freely archived at http://agingmind.cns.uiuc.edu/Pict_Norms.

The resulting name responses formed the basis of all subsequent analyses, yielding the following measures: dominant name, name agreement, concept agreement, and naming errors. Mean familiarity ratings were also calculated for each culture-by-age group. We provide descriptions of each of these measures next, and present main findings in the subsequent section.

Name response. Different approaches for coding the name response data were assessed through discussions with psycholinguists knowledgeable about both languages and cultures.⁴ We then established a final set of guidelines for counting different instances of names to ensure consistency and reasonableness across both American and Chinese cultures. For comparison purposes, the data were coded in a manner consistent with prior research. First, all name responses were recorded, with any misspellings (e.g., homonyms or errors) corrected. Second, in cases in which two or more responses were

Table 1
Sample Characteristics for Americans and Chinese, by Age Groups

Culture	Age Group	N	Age (Years)		Education (Years)		Health	
			M	SD	M	SD	M	SD
American	Younger	113	18.77	1.05	13.07	0.75	3.62*	0.72
	Older	103	66.47	4.24	15.88	2.52	3.86*	0.91
Chinese	Younger	100	20.09	1.04	14.01	0.75	3.31	0.73
	Older	100	64.68	3.38	16.73	1.35	3.24	0.67

Note—Health status was assessed on a 5-point scale (1 = much worse than average, 2 = worse than average, 3 = average, 4 = better than average, 5 = much better than average). *Means within culture differ at $p < .05$.

given, only the first name for the picture was counted (e.g., *house* for *house*, *home*). Third, any articles, quantifiers, prefixes, or suffixes accompanying name responses were not retained (e.g., *two*, *the*, *a*, *an*). In addition, any elaborations (e.g., *index finger* and *finger*) were counted as separate name responses. Because written Chinese is not amenable to abbreviations the way English is, the following criterion was applied to American responses only: Commonly accepted abbreviations were counted as separate names (e.g., *TV* and *television*), whereas uncommon abbreviations were written out (e.g., *spinning wheel* instead of *sp. wheel*). Finally, for Chinese responses only, a special criterion was imposed to accommodate conventions specific to the Chinese language. In Chinese, certain responses are deemed identical in their colloquial usage (e.g., 大象 and 象 [elephant]; 老虎 and 虎 [tiger]), even though to someone with limited familiarity with Chinese, they may appear to differ. All such responses were considered individually and discussed extensively with psycholinguists to determine how they should be coded. On the basis of these discussions, Chinese name responses were merged where appropriate and thereafter counted as the same.⁵

Dominant name. A dominant name was empirically determined and defined as the name most frequently assigned to a picture by participants within each culture-by-age group. Because the 260 pictures developed by Snodgrass and Vanderwart (1980) were originally normed on younger American adults, dominant name responses that did not match the picture names assigned by Snodgrass and Vanderwart for this group were separately identified; there were 11 such pictures for the younger American group.⁶ For the calculation of name agreement scores, we chose to retain the new dominant names for each picture as the “correct” response. Comparative data for these 11 pictures, including dominant names and agreement percentages from both Snodgrass and Vanderwart and the present study, appear in Appendix D of the Archived Materials; these have also been tagged in all relevant project Web site files.

Name agreement. Two scores for name agreement were computed for each picture. The first score, termed *name agreement percentage*, was calculated by taking the proportion of all responses (i.e., including all DKNs and DKOs and excluding all nonresponses [no answer, NA]) to a picture representing the most frequent name response within each group. This imposed the rather stringent constraint that only responses constituting dominant names counted toward the score. Hence, for the percentage measure, a higher value signifies greater name agreement.

The second score, called the *H* statistic, is a common measure of entropy or *disorder*, and was calculated by including all responses (including the above three categories of naming failures). The *H* statistic is useful in that, unlike the name agreement percentage score, it captures information about the *distribution* of names across participants. The *H* statistic was computed for each picture for each of the four culture-by-age groups according to the standard formula (Shannon & Weaver, 1949; Snodgrass & Vanderwart, 1980):

$$H = -\sum_{i=1}^k p_i \log_2(p_i),$$

where *k* is the total number of different names given each picture for all participants and *p_i* is the proportion of participants giving the *i*th name. Consistent with the manner in which *H* statistics were computed in previous norming studies, the DKNs and DKOs were excluded. For the *H* statistic measure, a lower value signifies greater name agreement; for example, when all subjects within a group supply the same name, the value is 0. A higher value indicates that a greater number of alternative names were supplied. For example, if two objects are both given their dominant name by 75% of the participants, but one is given a single other name and the second is given five other names, their name agreement percentages would be equivalent, but the first would have a lower *H* value.

Concept agreement. As stated earlier, the criterion for defining name agreement was quite conservative. Although this definition of

name agreement is useful for studies involving measures of picture naming responses, it is less useful for cognitive tasks (e.g., picture recall) in which synonyms for the dominant name may be scored as correct responses. Therefore, in accordance with what prior researchers of picture naming have used (e.g., Bates et al., 2003; Snodgrass & Vanderwart, 1980; Snodgrass & Yuditsky, 1996), we calculate a more liberal measure called *concept agreement* that captures the extent to which individuals think of a synonym for the dominant name or an equivalent concept when viewing a picture. Different name responses that represent the same basic concept were combined to create a score called *concept agreement percentage*. For instance, among Americans, both “airplane” and “plane” were included in deriving the concept agreement score, but were treated as separate name responses for the name agreement score responses.⁷ For the Chinese as well, separate names were counted as representing the same basic concept. For example, because it is very common for Chinese to use terms drawn from different dialects to name the same object, a participant might respond “扣子,” “钮扣,” “纽,” and so forth, when presented with a generic picture of a button. In addition, Chinese morphology and phrasing is such that a term can consist of different combinations of one, two, or even multiple written characters or words that are loosely related.

Naming difficulty and failures. For every picture, the number of naming failures (NA) and difficulties (DKO and DKN) were calculated for each culture-by-age group.

Familiarity. As noted above, participants used a 5-point rating scale to indicate their degree of familiarity with each object. When a participant did not know the object depicted (DKO) or did not provide a name response (NA), a familiarity rating was not generated. Such occurrences were therefore not included in computing means.

Statistical tests. With the picture as a unit of analysis to compare measures across culture and age, responses to 260 pictures were analyzed. For familiarity ratings, analyses of variance (ANOVAs) were performed to compare the groups. By contrast, the critical analysis for determining which pictures were perceived equivalently entailed comparison of cell counts (or, equivalently, proportions); we thus conducted likelihood ratio tests on name and concept agreement percentages on all groups and subsets of comparison groups. Sample sizes and expected cell counts were large enough in all cases to ensure the appropriateness of the asymptotic chi-square distribution.

In each case, the tests were to be carried out on a $2 \times 2 \times 2$ contingency table, in which the factors were culture (American or Chinese), age (younger or older), and name response (yes or no). Collapsing the last dimension from cell counts into proportions yielded a 2×2 table of probabilities. Seven tests, falling into three types, were conducted for each of the 260 picture items: (1) no effects at all, so that all four cell probabilities are equal (one test possible), (2) no effect of the Age or Culture factors (two tests possible), and (3) two specific cell probabilities are equal (six tests possible, but “diagonal” comparisons excluded, leaving four). Each test posits a type of equality across sets of cells; this in turn imposes a set of parametric restrictions, relative to allowing each cell probability to take on its own value, in which case the “null model” entails four degrees of freedom. For example, setting all four cell probabilities equal (Case 1 above) requires a single number to specify the common probability, as opposed to four separate probabilities in the ‘unrestricted’ case, and so entails losing three degrees of freedom. Thus, the three types of test require three, two, and one parametric restriction respectively, giving the degree-of-freedom differences for the standard (chi-square-distributed) test of log-likelihood differences.

RESULTS AND DISCUSSION

Table 2 contains summary statistics for the main dependent measures—*H* statistics, name agreement percentages, and concept agreement percentages. For each culture-by-age group, the mean and standard deviation,

Table 2
Summary Statistics of Main Dependent Measures for Americans and Chinese, by Age Groups

Culture	Age Group	<i>H</i> statistic		Name Agreement (%)		Concept Agreement (%)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
American	Younger	.57	.60	87	16	89	15
	Older	.67	.60	85	16	89	13
Chinese	Younger	1.41*	.97	67*	25	74*	24
	Older	1.74*	1.13	62*	25	69*	25

Note—*Means within culture differ at $p < .05$.

along with the median, 25th (Q1) percentile, 75th (Q3) percentile, range, and skew information for all 260 pictures are also provided.⁸

Validation of Picture Norms for Younger American Adults: Comparisons With Snodgrass and Vanderwart (1980) and Bates et al. (2003)⁹

To gauge the extent to which picture norms have remained constant among younger American adults over the last 25 years, the normative data from Snodgrass and Vanderwart (1980) and Bates et al. (2003) were compared with results obtained in the present study. In general, we found name agreement to be similar across the three samples. Whereas the mean *H* statistics were .56 in the Snodgrass and Vanderwart study and .67 in the Bates et al. study,¹⁰ it was .57 in our study.

Using a two-tailed *z* test of proportions to compare the name agreement percentages in this study with those of Snodgrass and Vanderwart (1980) and of Bates et al. (2003), we found a number of significantly different name agreement percentages at $p < .10$ (86 pictures [33% of the 260 pictures] for Snodgrass and Vanderwart, and 53 pictures [30% of 179 overlapping picture items] for Bates et al.).¹¹ For example, during the last 25 years, whereas the name agreement percentage increased from 36% to 76% for *dresser*, it decreased from 60% to 22% for *frying pan*.

Correlation between name agreement percentages in the Snodgrass and Vanderwart (1980) study and the present one provided further support that naming responses had changed for a significant proportion of the pictures. While the correlation for all 260 pictures suggested a strong relationship ($r = .71, p < .001$), the correlation for the subset of 86 pictures with significantly different name agreement percentages indicated a far weaker relationship ($r = .34, p < .002$), as verified by a likelihood ratio test. As would be expected, correlation for the remaining 174 pictures was high ($r = .94, p < .001$). Correlation between name agreement percentages for the 179 overlapping items in the Bates et al. (2003) study and the present study was also significant ($r = .50, p < .001$).

Of the 35 pictures for which Snodgrass and Vanderwart (1980) reported concept agreement data, 22 of them (63% of 35) also had concept agreement percentages that were significantly different from ours ($p < .10$). Further in-

spection of the data suggested that differences were due to actual shifts in frequencies of naming responses rather than changes in *H* statistics across the two studies. Correlation analysis of the mean concept agreement percentages for the 35 pictures indicated no correlation in the responses for the two samples ($r = -.18, n.s., p > .20$). This appears to be due largely to the lack of a strong correlation in concept agreement for the subset of 22 pictures with significant changes in concept agreement ($r = -.41, p < .10$). For the remaining 13 pictures with stable concept agreement, we found a positive correlation ($r = .67, p < .01$).

Comparisons of Picture Norms for Older Versus Younger American Adults

Next, we sought to identify pictures with equivalent name agreement and/or concept agreement percentages for younger versus older American adults by comparing responses for the two age groups. Contrary to what appears to be a common assumption when using pictorial stimuli to investigate age differences in cognition, we suggest that a substantial proportion of the pictures are potentially unsuitable for such studies. Comparisons of naming responses for 260 pictures by younger and older American adults revealed a total of 67 pictures (26%) with low name agreement and 44 pictures (17%) with low concept agreement. Per Bates et al. (2003), who reported that lower word frequency tends to produce lower name agreement, we tested whether there were differences in word frequencies for pictures with low versus equivalent name or concept agreement. Indeed, mean word frequencies for pictures with low name or concept agreement ($M = 2.43$) were lower than those for pictures with equivalent name or concept agreement ($M = 2.83, p < .05$).

Concordant with the results above for naming agreement for all 260 pictures, correlation between younger and older American age groups for the entire set of pictures indicated a significant positive relationship ($r = .59, p < .001$; see Table 3A). However, also as expected, correlation analyses on subsets of the data indicated a lack of relationship for the 67 pictures with low name agreement ($r = -.04, n.s., p > .20$) as well as the 44 with low concept agreement ($r = -.14, n.s., p > .20$). By contrast, correlations between younger and older American age groups were high for the remaining 193 pictures (74%

Table 3
Correlations for 260 Pictures, Across All
Culture-by-Age Groups

	Younger American	Older American	Younger Chinese	Older Chinese
A. Mean Name Agreement Percentages				
Younger American	1.00			
Older American	.59*	1.00		
Younger Chinese	.37*	.22*	1.00	
Older Chinese	.39*	.28*	.88*	1.00
B. Mean Concept Agreement Percentages				
Younger American	1.00			
Older American	.57*	1.00		
Younger Chinese	.41*	.28*	1.00	
Older Chinese	.38*	.33*	.89*	1.00
C. Mean Familiarity Ratings				
Younger American	1.00			
Older American	.93*	1.00		
Younger Chinese	.39*	.38*	1.00	
Older Chinese	.32*	.27*	.73*	1.00

*Correlation is significant at the .01 level (two-tailed).

of 260) with high name agreement ($r = .87, p < .001$), and 216 pictures (83%) with high concept agreement ($r = .78, p < .001$). Although this clearly suggests that a large number of pictures are, in fact, suitable for use in studies of aging and cognition, researchers nonetheless should exercise caution when selecting pictorial stimuli.

A list of 42 pictures yielding both low name and concept agreement in comparisons of naming responses of younger and older adults appear in Appendix E of the Archived Materials. As would be expected, correlation analyses between the two American age groups, run on the 42 pictures, for name agreement and concept agreement responses also indicated a weak relationship ($r = -.14, n.s., p > .20$, for name agreement; and $r = -.08, n.s., p > .20$, for concept agreement). The age differences appeared not to stem from any one identifiable cause.

First, half of the 42 pictures yielded lower name and concept agreement for older, compared with younger, adults. For a majority of these 21 pictures, the difference may well be attributable to how the objects are depicted. For instance, some of the pictured objects (e.g., *cloud, moon, watermelon*) were relatively difficult to make out or discern, evidenced by low naming percentages for all adults, but differentially lower naming percentages and high naming failures for the older adults. Analyses of naming errors and failures are discussed further below.

For the remaining half of the 42 pictures, which yielded higher name and concept agreement for older adults, two explanations appeared to account for the apparent age differences. Some of the pictures featured objects that, despite present-day commonality for all age groups, looked somewhat old-fashioned or dated (e.g., *ashtray, doll, jacket*). These objects may have thus been more readily recognizable in the depicted forms to the older than to the younger group. Other pictures, despite being depicted in an up-to-date style, were of objects that younger adults would encounter less frequently (e.g., *ar-*

tichoke, chisel, spinning wheel). Analysis of familiarity ratings for these pictures was generally consistent with such explanations: Older adults ($M = 3.92$) generated higher familiarity ratings for these objects than younger adults ($M = 3.22$) [$t(214) = 7.30, p < .001$], although the correlation of familiarity responses generated by the two groups for the 42 pictures was relatively high ($r = .80, p < .001$). We further discuss familiarity responses below.

A list of 25 pictures with nonequivalent name agreement, but equivalent concept agreement, across the two age groups appears in Appendix F of the Archived Materials. Accordingly, analyses comparing the two groups, run on the 25 pictures, indicated that name agreement responses were not correlated ($r = -.14, n.s., p > .20$) and that concept agreement responses were positively correlated ($r = .89, p < .001$). Interestingly, this set of pictures consisted primarily of objects for which older adults exhibited lower name agreement than did younger adults. The age difference seemed to reflect a shift in the usage of dominant names for certain objects (e.g., *tie* vs. *necktie, couch* vs. *sofa, bike* vs. *bicycle*, for younger vs. older adults, respectively). As such, familiarity ratings by the two groups for these pictures were highly correlated ($r = .91, p < .001$).

Comparisons of Picture Norms Across Culture and Age Groups

Next, we analyzed the data for all study participants in order to identify which pictures were equivalent in terms of name and/or concept agreement across both culture and age. Correlations of name agreement and concept agreement percentages among all culture-by-age groups can be found in Tables 3A and B. Chi-square tests on all 260 pictures revealed that 57 of them (22%) had equivalent levels of name and concept agreement percentages ($p > .10$) across all four culture-by-age groups. Results of the correlation analyses for the 57 pictures are presented in Table 4. Summary statistics associated with the 57 pictures, presented by culture and age, appear in Appendix A of the Archived Materials; also included is a listing of name responses with their frequency counts in parentheses.

Additional analyses yielded another 29 pictures (11%) with equivalent concept agreement but nonequivalent name agreement percentages; these appear in Appendix B of the Archived Materials. Results of the correlation analyses for the 29 pictures can be found in Table 5. Notwithstanding some differences in mean familiarity ratings and H statistics across the groups, normative data for pictures identified in Appendices A and B of the Archived Materials thus comprise a suitable set of picture stimuli for research studies investigating cross-cultural and/or age-related differences in cognition.

Comparisons with Snodgrass and Vanderwart (1980), Snodgrass and Yuditsky (1996), and Bates et al. (2003)

We next compared our results with those of Snodgrass and Yuditsky (1996)¹² and Bates et al. (2003), in addition

Table 4
Correlations for 57 Pictures With Equivalent Name and Concept Agreement, Across All Culture-by-Age Groups

	Younger American	Older American	Younger Chinese	Older Chinese
A. Mean Name Agreement Percentages				
Younger American	1.00			
Older American	.90*	1.00		
Younger Chinese	.82*	.78*	1.00	
Older Chinese	.83*	.77*	.76*	1.00
B. Mean Concept Agreement Percentages				
Younger American	1.00			
Older American	.90*	1.00		
Younger Chinese	.76*	.65*	1.00	
Older Chinese	.79*	.70*	.68*	1.00
C. Mean Familiarity Ratings				
Younger American	1.00			
Older American	.98*	1.00		
Younger Chinese	.69*	.71*	1.00	
Older Chinese	.71*	.74*	.86*	1.00

*Correlation is significant at the .01 level (two-tailed).

Table 5
Correlations for 29 Pictures With Equivalent Concept Agreement and Nonequivalent Name Agreement, Across All Culture-by-Age Groups

	Younger American	Older American	Younger Chinese	Older Chinese
A. Mean Name Agreement Percentages				
Younger American	1.00			
Older American	.55*	1.00		
Younger Chinese	n.s.	n.s.	1.00	
Older Chinese	n.s.	n.s.	.42**	1.00
B. Mean Concept Agreement Percentages				
Younger American	1.00			
Older American	.89*	1.00		
Younger Chinese	.68*	.72*	1.00	
Older Chinese	.67*	.64*	.62*	1.00
C. Mean Familiarity Ratings				
Younger American	1.00			
Older American	.94*	1.00		
Younger Chinese	.65*	.59*	1.00	
Older Chinese	.58*	.54*	.88*	1.00

*Correlation is significant at the .01 level (two-tailed). **Correlation is significant at the .05 level (two-tailed).

to Snodgrass and Vanderwart (1980). This enabled us to identify objects for which name agreement remained equivalent across time, age, or culture, and the relative degrees of consistency reflected across these variables.

In order to do this, we analyzed all 179 pictured objects included across the four studies. Table 6 contains the mean name agreement values for the 179 pictures that could be calculated on the basis of item-level data available from the three prior published studies. Correlations among the data in these studies were also calculated (see Table 7).

Overall, name agreement appears to be more consistent across age and time than across culture. Mean val-

ues for the 179 overlapping items indicated that name agreement between younger and older Americans in our study was high, with nearly identical mean name agreement scores for the younger ($M = .89$) and older group ($M = .87$). Name agreement was also equivalently high for the 179 items in the Snodgrass and Yuditsky (1996) experiments ($M = .94$ for Experiment 1; $M = .96$ for Experiment 2), and for the Bates et al. (2003) study ($M = .89$). By contrast, mean name agreement values were more modest for the Chinese groups in our study ($M = .65$ for younger Chinese; $M = .60$ for older Chinese; [$F(1,356) = 2.43, p = .12$ for the cross-age difference]). Bates et al. found the name agreement scores for Chinese to be the lowest of the seven languages they included in their study. We obtained directionally consistent results in our study with lower name agreement among younger Chinese, compared with younger Americans. Interestingly, the mean naming agreement values for the 179 overlapping items was significantly higher for the younger Chinese adults in the Bates et al. study ($M = .77$) compared with those in our study [$M = .65; F(1,356) = 20.08, p < .001$]. The reasons for this difference are unclear and could not be tested directly, because Bates et al. did not provide item-level naming responses or H statistics in their database. However, we found that for younger Chinese adults in the Bates et al. study, naming agreement percentages for 113 out of 179 overlapping picture items (63%) were significantly different from our study ($p < .10$). We speculate that the divergence may be due, in part, to the differences in samples and procedures between the two studies. Bates et al. collected vocal responses from Mandarin Chinese speakers in Taipei, Taiwan, while we collected written responses from Mandarin-speaking Chinese individuals in Beijing, China. Thus, there may be regional differences in naming responses and differences between spoken and written responses. Furthermore, the smaller sample size ($n = 50$) in the Bates et al. study compared with ours ($n = 100$) may have lowered the overall number of idiosyncratic responses generated by participants, on which the asymptotic accuracy of the H statistic critically depends.

Finally, correlation analyses provided further support for consistency in naming agreement across age and time relative to culture. Consistent with the correlation for all 260 pictures, there was high correlation between naming agreement values for younger Americans in our study and in Snodgrass and Vanderwart's (1980) study for the subset of 179 overlapping picture items ($r = .63, p < .001$). Interestingly, this correlation across time was higher than the correlation between younger and older Americans in the present study ($r = .51, p < .001$). If we were to consider age as a proxy for time, it makes sense that the strength of the relationship for naming agreement should be greater for groups separated by 25 years as opposed to an average age difference of 45 years. Compared with these values, correlations between Americans and Chinese within each age group were lower ($r = .28, p < .001$, for younger adults; $r = .17, p < .02$,

Table 6
Comparison of Results of the Present Study With Those of Snodgrass and Vanderwart (1980), Snodgrass and Yuditsky (1996), and Bates et al. (2003) for 179 Overlapping Items

Variable	Present Study				SV	SY—Exp. 1	SY—Exp. 2	Bates	Bates
	Younger	Older	Younger	Older					
	American	American	Mainland Chinese	Mainland Chinese	Younger American	Younger American	Younger American	Younger American	Younger Taiwanese
<i>H</i> statistic	0.502	0.632	1.364	1.693	0.516	NA	NA	NA	NA
Name agreement (%)	89	87	65	60	88	NA	NA	89	77
Concept agreement (%)	91	90	70	67	91*	94	96	NA	NA
Familiarity	3.55	4.08	4.53	4.57	3.34	NA	NA	NA	NA
Valid responses (%)	98.7	99.1	93.0	91.1	98.3	NA	NA	96.1	89.3
Invalid responses (%)	1.0	0.6	6.8	4.0	NA	NA	NA	1.5	6.1
No responses (%)	0.3	0.4	0.2	5.0	NA	NA	NA	2.3	4.6

Note—Present study = Yoon et al. (2004). SV = Snodgrass and Vanderwart (1980). SY = Snodgrass and Yuditsky (1996; Experiments 1 and 2). Bates = Bates et al. (2003). *This percentage was based on a subset of 35 pictures for which Snodgrass and Vanderwart (1980) reported concept agreement. NA = Not available.

for older adults), lending further support to the notion that consistency in naming responses is lower across culture than across age or time.¹³

Familiarity

ANOVAs performed on the effects of culture and age on familiarity ratings revealed a significant main effect of age [$F(1,1036) = 15.26, p < .0001$], with older adults reporting higher levels of picture familiarity than did younger adults (see Table 8 for summary statistics). We also found a significant main effect of culture [$F(1,1036) = 320.96, p < .0001$], with the Chinese group, somewhat unexpectedly, indicating higher levels of familiarity than did the American group. The main effects were qualified by a significant interaction effect [$F(1,1036) = 32.34, p < .0001$]. Specifically, whereas older American adults reported significantly higher levels of picture familiarity than did their younger American counterparts [$F(1,518) = 55.25, p < .0001$], the age difference in familiarity for the

Chinese group was only marginally significant [$F(1,518) = 3.41, p < .07$].

Among both the younger and older age groups, the Chinese reported significantly higher familiarity ratings than did the Americans [$F(1,518) = 245.17, p < .0001$, within the younger group; $F(1,518) = 119.97, p < .0001$, within the older group]. We speculate that the Chinese participants' familiarity ratings may reflect a systematically positive bias in the use of 5-point scales rather than an underlying difference in familiarity strength. Lending informal support to this explanation were the high rates of naming errors accompanied by high familiarity scores for some of the pictures. For example, only 45 out of the 200 Chinese participants correctly identified the picture for *stove* (Picture 219), but the mean familiarity score was nonetheless very high ($M = 4.57$).

Although we are unable, on the basis of these data, to ascertain the significance of the differences in familiarity ratings between the Chinese and Americans, we sug-

Table 7
Correlations of Name Agreement Percentages From the Present Study and Those From Snodgrass and Vanderwart (1980), Snodgrass and Yuditsky (1996), and Bates et al. (2003), for 179 Overlapping Items

		Present Study				SV	SY		Bates	
		Younger	Older	Younger	Older		Exp. 1	Exp. 2	Younger	Younger
		American	American	Chinese	Chinese	Younger American	Younger American	Younger American	Younger American	Younger Taiwanese
Present study	Younger American	1.00								
	Older American	.51*	1.00							
	Younger Mainland Chinese	.28*	.12	1.00						
	Older Mainland Chinese	.30*	.17**	.87*	1.00					
SV	Younger American	.63*	.67*	.26*	.28*	1.00				
SY, Exp 1	Younger American	.44*	.17**	.39*	.38*	.36*	1.00			
SY, Exp 2	Younger American	.50*	.19**	.36	.37*	.41*	.81*	1.00		
Bates	Younger American	.71*	.38*	.28*	.29*	.62*	.39*	.39*	1.00	
Bates	Younger Taiwanese	.40*	.28*	.64*	.59*	.42*	.39*	.40*	.41*	1.00

*Correlation is significant at the .01 level (two-tailed). **Correlation is significant at the .05 level (two-tailed). Present study = Yoon et al. (2004). SV = Snodgrass and Vanderwart (1980). SY = Snodgrass and Yuditsky (1996; Experiments 1 and 2). Bates = Bates et al. (2003).

Table 8
Summary Statistics of Familiarity Ratings for Americans and Chinese, by Age Groups

Culture	Age Group	Familiarity	
		<i>M</i>	<i>SD</i>
American	Younger	3.51*	0.94
	Older	4.04*	0.69
Chinese	Younger	4.52	0.46
	Older	4.58	0.49

Note—Familiarity ratings were assessed on 5-point scales (1 = *the object is not at all familiar to me*, 5 = *the object is highly familiar to me*). Means within culture differ at $p < .05$.

gest that in most cases this may not be a cause for serious concern. For instance, a review of the mean familiarity ratings data in Appendix B of the Archived Materials shows that a majority of the pictures, for which we obtained lower ratings for Americans than for Chinese, are items (e.g., *cat, fish, ladder, peanut, scissors, star*) that are presumably of high familiarity to the American samples even if they are not frequently encountered in everyday life.

In addition, correlational analyses of familiarity ratings for all the pictures across the four culture-by-age groups indicated high correlations between age groups within each culture ($r = .93$ for Americans and $r = .73$ for Chinese), as well as significant, albeit lower, correlations for all other paired groups (see Table 3C). Interestingly, we also found a strong correlation between the mean familiarity ratings of the younger American cohort in this study, and those reported by Snodgrass and Vanderwart (1980) for all 260 pictures ($r = .92$, $p < .001$). Moreover, correlations remained high, regardless of whether the pictures had high or low name and/or concept agreement. However, we found a significant difference in mean familiarity ratings across all 260 pictures for the Snodgrass and Vanderwart (1980) study compared with the present one [$M = 3.3$ vs. $M = 3.5$, respectively; $t(259) = 9.01$, $p < .001$]. Although this indicated that familiarity levels may have diverged over the last 2 decades, we found that for half (130) of the pictures, the mean familiarity ratings for the two groups were not significantly different. Moreover, no systematic differences in familiarity were detected across the sets of pictures yielding either equivalent or nonequivalent levels of name and concept agreement. Thus, the more

likely explanation is that the increase in average familiarity ratings for the entire set of 260 pictures occurred because of the experimental procedure we used. We suggest that having participants name an object, followed immediately by having them rate its familiarity, may have served to enhance perceptions of familiarity with the object and thereby to inflate their judgments.

On the basis of the above results, we suggest that researchers utilizing these pictorial stimuli carefully consider all the measures reported for each picture as well as the experimental context to determine its appropriateness for inclusion in a particular study. Although differences in familiarity ratings across age, time, or culture appear to be less of a potential problem, it would seem prudent to consider *H* statistics in concert with name agreement and concept agreement percentages at the picture-item level. For instance, if naming responses are the measures of critical importance, it is important that *H* statistics be equivalent across the groups under examination. For studies involving picture recall, wherein multiple synonymous names for a particular object are acceptable, significant differences in *H* values or name agreement percentages may be of less concern, as long as the concept agreement percentages are equivalent.

Naming Errors

Finally, an analysis of naming failures and difficulties in each of the four culture-by-age groups revealed that both younger and older Chinese participants committed more naming errors than did their American counterparts (see Table 9). Younger Chinese participants indicated a relatively high rate of DKNs; that is, they recognized the object but could not name it. By contrast, older Chinese participants appeared to encounter naming difficulties with all three categories of error—NA, DKN, and DKO. A similar pattern of naming errors was found for the American and Chinese samples by Bates et al. (2003; see Table 6). The higher rates of naming errors among the Chinese are not surprising, given that Snodgrass and Vanderwart's (1980) pictures were developed in the U.S. It is therefore likely that, to Chinese individuals, some pictures do not clearly depict the concepts they were meant to represent. Furthermore, as previously mentioned, some objects (e.g., *artichoke, barn, chisel, and raccoon*) are either nonexistent or uncommon in China.

Table 9
Naming Failures and Difficulties for Americans and Chinese, by Age Groups

Culture	Age	NA (%)	DKO (%)	DKN (%)
American	Younger	0.30	0.30	0.80
	Older	0.30	0.30	0.30
Chinese	Younger	0.30	2.40	4.40
	Older	4.90	1.50	2.40

Note—The percentages were calculated by dividing the number of errors by the total number of responses possible for each culture-by-age group. NA, no answer; DKO, don't know object; DKN, don't know name.

Conclusions

In this paper, we provide cognitive researchers with current normative data for Snodgrass and Vanderwart's (1980) set of 260 pictures, for younger and older American and Chinese adults. These data complement and extend those stemming from the original study, as well as more recent ones by Snodgrass and Yuditsky (1996), Bates et al. (2003), and Székely et al. (2003); comprehensive item-level comparisons among these studies are available on the project Web site.

Analyses of the data yielded several noteworthy findings. First, we found substantial cross-cultural differences in name agreement and concept agreement for many of the pictures. For both younger and older groups, significant cross-cultural differences were obtained for a large number of the 260 pictures. Second, contrary to what has often been implicitly assumed in literature addressing picture naming across adult age groups, differences in name agreement or concept agreement were obtained for over 20% of the pictures within the American group; similar age-related differences were found within the Chinese group. Finally, in comparisons of Snodgrass and Vanderwart's (1980) picture norms and the current norms for younger Americans, we found that name agreement and concept agreement percentages had changed for a substantial proportion of the 260 pictures.

These findings highlight the importance of identifying subsets of pictures that are judged suitable not only for the particular goals of a study, but also the cultural or age groups being studied. Accordingly, we have provided a subset of pictures that constitute potentially suitable stimuli for use in research investigating the joint effects of culture and aging on cognitive functioning. We have also provided additional subsets of pictures that are appropriate for examining the separate effects of aging or culture on cognition. Because the patterns of variation across picture items, cultures, age groups, time, and studies are complex, researchers fashioning stimuli for use across various combinations of these different variables would be well advised to consider each potential pictorial stimulus on its own merits and assess it in the context of their overall research goals.

REFERENCES

BATES, E., D'AMICO, S., JACOBSEN, T., SZÉKELY, A., ANDONOVA, E., DEVESECOVI, A., HERRON, D., LU, C. C., PECHMANN, T., PLEH, C., WICHA, N., FEDERMEIER, K., GERDJIKOVA, I., GUTIERREZ, G., HUNG, D., HSU, J., IYER, G., KOHNERT, K., MEHOTCHEVA, T., OROZCO-FIGUEROA, A., TZENG, A., & TZENG, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*, 344-380.

BERMAN, S., FRIEDMAN, D., HAMBERGER, M., & SNODGRASS, J. G. (1989). Developmental picture norms: Relationships between name agreement, familiarity, and visual complexity for child and adult ratings of two sets of line drawings. *Behavior Research Methods, Instruments, & Computers*, *21*, 371-382.

CYCOWICZ, Y. M., FRIEDMAN, D., ROTHSTEIN, M., & SNODGRASS, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*, 171-237.

KIRSNER, K., MILECH, D., & STUMPFEL, V. (1986). Word and picture

identification: Is representational parsimony possible? *Memory & Cognition*, *14*, 398-408.

LACHMAN, R., & LACHMAN, J. L. (1980). Picture naming: Retrieval and activation of long-term memory. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Greenberg, & L. W. Thompson (Eds.), *New directions in memory and aging* (pp. 313-343). Hillsdale, NJ: Erlbaum.

MATSUKAWA, J. (1983). A study of characteristics of pictorial materials with Snodgrass and Vanderwart pictures. *Memoirs of the Faculty of Law and Literature, Shimane-Ken, Japan*: Shimane University Press.

MCDERMOTT, K. B., & ROEDIGER, H. L., III (1994). Effects of imagery on perceptual implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 1379-1390.

MITCHELL, D. B., & BROWN, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 213-222.

MITCHELL, K. J., JOHNSON, M. K., RAYE, C. L., MATHER, M., & D'ESPOSITO, M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology & Aging*, *15*, 527-541.

NYBERG, L., CABEZA, R., & TULVING, E. (1996). PET studies of encoding and retrieval: The HERA model. *Psychonomic Bulletin & Review*, *3*, 135-148.

OLDFIELD, R. C., & WINGFIELD, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273-281.

PARK, D. C., NISBETT, R., & HEDDEN, T. (1999). Aging, culture, and cognition. *Journal of Gerontology: Psychological Sciences*, *54*, P75-P84.

PARK, D. C., SMITH, A. D., MORELL, R. W., PUGLISI, J. T., & DUDLEY, W. N. (1990). Effects of contextual integration on recall of pictures by older adults. *Journal of Gerontology*, *45*, 52-57.

PARKIN, A. J., & STREETE, S. (1988). Implicit and explicit memory in young children and adults. *British Journal of Psychology*, *79*, 361-369.

PROCTOR, R. W., & VU, K-P. L. (1999). Index of norms and ratings published in the Psychonomic Society journals. *Behavior Research Methods, Instruments, & Computers*, *31*, 659-667.

RAJARAM, S. (1996). Perceptual effects of remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 365-377.

SEO, C. W. (1988). Studied similarities and differences in the processing of pictures and words using J. G. Snodgrass and M. Vanderwart's (1980) set of 260 black-and-white line drawings. *Korean Journal of Psychology*, *7*, 158-186.

SHANNON, C. E., & WEAVER, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

SHU, H., CHENG, Y. S., & ZHANG, H. C. (1989). *Chinese norms for name agreement, image agreement, familiarity and visual complexity of 235 pictures*. Unpublished manuscript.

SNODGRASS, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning & Verbal Behavior*, *23*, 3-22.

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: Human Learning & Memory*, *6*, 174-215.

SNODGRASS, J. G., & YUDITSKY, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *28*, 516-536.

STARK, C. E. L., & SQUIRE, L. R. (2000). Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. *Journal of Neuroscience*, *20*, 7776-7781.

SZÉKELY, A., D'AMICO, S., DEVESECOVI, A., FEDERMEIER, K., HERRON, D., IYER, G., JACOBSEN, T., & BATES, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*, 621-633.

NOTES

1. Norms for American children have, however, appeared in the literature for small subsets of pictures (e.g., Berman, Friedman, Ham-

berger, & Snodgrass, 1989; Cycowicz, Friedman, Rothstein, & Snodgrass, 1997).

2. Whereas a few researchers have previously collected picture norms on young adult samples in East Asia, access to these data is severely limited because they are published in their native languages (Matsukawa, 1983; Seo, 1988; Shu, Cheng, & Zhang, 1989). We thank Joan Gay Snodgrass for bringing these studies to our attention.

3. It should be noted that the procedures we used to collect familiarity ratings were different from those used by Snodgrass and Vanderwart (1980). We asked our study participants to write down the name of the object, and then immediately rate the object for familiarity. Thus, a total of 8 sec was provided to name a pictured object and rate its familiarity. By contrast, Snodgrass and Vanderwart had one group of participants identify the objects, and a separate group judged their familiarity.

4. We thank Yao Cui of the Institute of Psychology at the Chinese Academy of Sciences for his linguistic expertise.

5. In the data presentation found at the project Web site, we provide listings of all Chinese name responses that have been merged.

6. Snodgrass and Yuditsky (1996) also found in their picture naming study that for 8 of the 250 pictures they used, the dominant name response differed from those of Snodgrass and Vanderwart (1980). Interestingly, with the exception of *pocketbook* and *spool of thread*, which also emerged on our list of 11 pictures yielding different dominant names, those identified in the Snodgrass and Yuditsky study comprised a different set than those in our study. This may be due to the differences in their experimental procedure, which involved collecting vocal responses.

7. This is in accordance with the manner in which data were coded by Snodgrass and Vanderwart (1980).

8. The main dependent measures for each of the 260 pictures for each culture-by-age group can be obtained from the Archived Materials and at the project Web site.

9. We do not include the Snodgrass and Yuditsky (1996) study in the comparison, because they did not provide name agreement percentages based on dominant name responses at the picture-item level.

10. This mean *H* statistic reported by Bates et al. (2003) was calculated for all 520 pictures included in their study.

11. Comparisons between the normative data from Snodgrass and Vanderwart (1980) and the present study can be found in the Archived Materials and at the project Web site. We use a cutoff value of $p = .10$ throughout for conservatism in suggesting which pictures' measures may have changed, calling their use into question; one-tailed comparisons should therefore be interpreted as significant at .05. Our calculation for the number of overlapping items indicates 179 in total, as opposed to 176 in Bates et al. (2003).

12. Note that Snodgrass and Yuditsky (1996) provide the percentage of correct responses ("percentage of correct naming scored with a liberal criterion") rather than name agreement percentage.

13. Analogous results obtained for analyses using all 260 pictures.

ARCHIVED MATERIALS

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

To access these files or links, search the archive for this article using the journal (*Behavior Research Methods, Instruments, & Computers*), the first author's name (Yoon), and the publication year (2004).

FILE: Yoon-BRMIC-2004.zip.

DESCRIPTION: The compressed archive file contains:

Appendix_A.xls and Appendix_A.csv, containing 57 pictures with equivalent name and concept agreement, across all cultural and age groups;

Appendix_B.xls and Appendix_B.csv, containing 29 pictures with nonequivalent name agreement and equivalent concept agreement, across all cultural and age groups;

Appendix_C.xls and Appendix_C.csv, containing additional demographic characteristics for Americans and Chinese, by age groups;

Appendix_D.xls and Appendix_D.csv, containing items with dominant names different from Snodgrass & Vanderwart (1980) and their name agreement percentages for younger Americans;

Appendix_E.xls and Appendix_E.csv, containing 42 pictures with non-equivalent name and concept agreement for younger and older Americans;

Appendix_F.xls and Appendix_F.csv, containing 25 pictures with nonequivalent name agreement and equivalent concept agreement for younger and older Americans;

Tables 1–9 in the article proper are included as {Table_1.xls and Table_1.csv, ..., Table_9.xls and Table_9.csv};

pndata1.xls and pndata1.csv, containing main dependent measures (familiarity, *H* value, name agreement percentage, concept agreement percentage), and name responses included in name agreement and concept agreement percentages for 260 Snodgrass and Vanderwart (1980) pictures, for each culture-by-age group;

pndata2.xls and pndata2.csv, containing raw frequency counts of naming errors and all dominant and nondominant name responses for 260 Snodgrass and Vanderwart (1980) pictures, for each culture-by-age group;

pndata3.xls and pndata3.csv, containing comparisons of picture norming data for younger American adults from Yoon et al. (2004) and Snodgrass and Vanderwart (1980);

P_Value_Comparisons_1.xls and P_Value_Comparisons_1.csv, containing raw proportions for name and concept agreement in both English and Chinese, along with all possible pairwise two-sample tests of proportional differences, across four studies (Bates et al. [2003]; Snodgrass & Vanderwart [1980]; Snodgrass & Yuditsky [1996]; Yoon et al. [2004]).

LINK: http://agingmind.cns.uiuc.edu/Pict_Norms/

DESCRIPTION: All files above are downloadable from the project site at this URL.

AUTHOR'S E-MAIL ADDRESS: yoonc@umich.edu.

AUTHOR'S WEB SITE: <http://agingmind.cns.uiuc.edu>.

(Manuscript received July 3, 2003;
revision accepted for publication July 23, 2004.)