

Pictorial Naming Specificity across Ages and Cultures: A Latent Class Analysis of Picture Norms for Younger and Older Americans and Chinese

Carolyn Yoon^{a,b} Fred Feinberg^a Angela H. Gutchess^c

^aUniversity of Michigan Business School, ^bInstitute for Social Research, and ^cDepartment of Psychology, University of Michigan, Ann Arbor, Mich., USA

Key Words

Picture norms · Naming specificity · Latent class analysis

Abstract

Background: Cognitive aging research, as well as cross-cultural research, often relies on pictorial stimuli to address how perceptions of common objects vary by population group. **Objective:** We investigate *naming specificity* – the degree of detail elicited for object labels – across age (younger-older) and culture (American-Chinese). **Methods:** Using latent class analysis techniques, we segregated picture-item responses into multiple specificity levels. The analysis was conducted for 260 pictures, across four groups of younger (aged 17–25 years) and older (aged 59–76 years) adults in the US and mainland China. **Results:** Overall, three naming specificity classes were supported. Age differences were modest compared to those across culture. In particular, Chinese groups showed far greater variation, naming more items both with high and with low specificity than age-matched American counterparts. Our results differed from prior studies using familiarity and latency measures. Moreover, approximately 28% of commonly-used picture items differed across all four groups. **Conclusion:** These results highlight the need to select appropriate pictorial stimuli for studies involving populations diverse in age and cultural background.

Copyright © 2006 S. Karger AG, Basel

Introduction

Studies of memory and cognitive processes have long relied on pictorial stimuli, typically simple, abstract line drawings of common objects. Attributes such as object or picture familiarity [1] are known to correlate well with cognitive measures, and affect both memory and retrieval processes. Cognitive psychologists have made broad use of pictorial stimuli to study, for example: how images and visual-spatial representations differ from verbal or abstract representations in memory [2, 3]; the effects of picture priming on implicit and explicit memory [4–6]; the nature of representational systems underlying visual memory in normal and impaired adults [7–9], and differences in both visual perception and memory across the lifespan [10, 11].

Among the first to attain wide usage, Snodgrass and Vanderwart's [12] norms have only recently been re-examined using other young American adult samples [13–15], though several studies have provided cross-cultural and/or cross-age validation for them and various supersets ([16, 17] for French; [18, 19] for Spanish; [20] for Italian; [15] for Chinese; [21] for comparisons spanning seven languages).¹ Across these and other studies, numerous

¹ Whereas a few researchers have collected picture norms on young adult samples in East Asia, access is hampered by publication only in their native languages [22, 23]. We thank Joan Gay Snodgrass for bringing these studies to our attention.

subject-specific covariates have been examined in relation to naming and imagery. An abridged list would include agreement and latency for naming, as well as familiarity, variability and complexity for the images themselves. An increasingly clear portrait has thus emerged of the relative suitability of various pictorial stimuli for research across age and cultural groups.

A presumption typically underlying research using pictorial stimuli is that various groups, in particular younger and older adults, do not differ in terms of the specificity of names they assign to objects. Here, we investigate the validity of this assumption and suggest that there may be pronounced cohort differences, specifically using the most commonly used picture norms for older versus younger and American versus Chinese, those of Snodgrass and Vanderwart [12]. To this end, we apply latent class techniques to parcel pictorial stimuli into endogenously determined classes, and rigorously test whether these derived classes differ across culture-by-age groups. That is, we adopt a 'bottom-up' approach, with specificity classes determined by the corpus of pictorial naming data alone, not by a priori notions of which pictures are indeed more specific, and to whom.

We introduce *naming specificity* – the degree of hierarchical detail elicited in object label descriptions – to identify pictorial stimuli suitable for investigating age and/or cross-cultural differences in cognition. The recent emergence of research interest in cross-cultural differences in cognition, particularly so East-Asian and Western, underscores the need for culture-invariant stimulus materials. Although pictures of everyday objects are potentially useful stimuli for comparing East-Asians and Americans [24], some objects are likely to vary in terms of how specifically they are perceived across cultures. For example, certain animals and vegetables indigenous to the US (e.g. raccoon, asparagus) may be recognizable to East-Asian subjects in terms of an appropriate category, but named at a relatively superordinate level, such as 'small mammal' or 'vegetable'.

The methodology and analysis in the present paper attempt to help guide selection of pictorial stimuli for studies of age and cross-cultural differences, and thereby enrich the normative data available to researchers. In particular, we examine the pattern of responses across age, culture, or any two culture-by-age groups, in order to assess the extent to which: (1) any of the culture-by-age groups tend more towards naming specificity than the others; (2) the pictures themselves fall into natural *classes* within any of the culture-by-age groups; and (3) the natural classes vary across groups.

Naming Specificity

Several examples drawn from the data help illustrate distinctions between naming specificity and image or concept familiarity. (Note that, throughout, we use 'class' to refer to sets of pictures, and 'group' to sets of participants.) Consider, as per Snodgrass and Vanderwart's [12] numbering, 'rocking chair' (item 188) and 'barrel' (item 18): both were highly familiar to all subjects, and did not differ strongly on any main agreement measures [15]. However, Chinese subjects were remarkably more specific in naming the (fairly generic) depiction of a barrel presented. Whereas American subjects limited themselves primarily to the set [barrel, keg, wheel barrow, beer barrel, wood barrel], Chinese subjects, particularly older ones, offered not only 'barrel' (桶), but 'subordinate' variants [wooden barrel (木桶), wine barrel (酒桶), bamboo barrel (竹桶), nail barrel (钉子桶), water barrel (水桶), small barrel (小桶)]; highly specific connotations [old style wine barrel (老式酒桶), wooden beer barrel (啤酒木桶)]; as well as some (fairly rare) questionable item names [drum (鼓), commode (便桶) and wine cup (斗)].

We stress that it is not the number of different items generated, but their relative specificity, which differs markedly. Let us compare 'barrel' and 'rocking chair'; for both, despite similarities in familiarity and other commonly used measures, degree of naming specificity diverges sharply across cultures. For 'barrel', the Chinese generated a greater number of distinct items and were more specific in their object naming. 'Rocking chair', however, displays the opposite pattern, with the Chinese producing many more responses overall, but being notably less specific: though many did produce 'rocking chair' (e.g. 摇摇椅), a far larger percentage produced the less specific 'chair' (椅 or 椅子) than did their American counterparts [and the number of less overtly 'accurate' responses was higher, including couch (躺椅) and vine chair (藤椅)]. We in fact found no systematic relationship between naming specificity and sheer number of responses recorded for a group, nor with that group's overall level of familiarity with the picture, item or concept. As we shall see, latent class analyses suggest that inferences about pictorial suitability garnered from naming specificity do not generally accord with those based on the major measures used in prior research.

Table 1. Age, education and health characteristics for Americans and Chinese, by age group

Culture	Age	n	Age years		Education years		Self-rated health status	
			mean	SD	mean	SD	mean	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 ^a	0.73
	older	100	64.68	3.38	16.73	1.35	3.24 ^a	0.67

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 in the same column that do not share superscripts differ at $p < 0.05$.

Methods

Participants

One hundred and thirteen younger adults (17–25 years) from the University of Michigan and 103 older adults (60–75 years), comprising the American cultural group, were recruited for testing in Ann Arbor, Mich., USA. The younger Americans were recruited from a subject pool of undergraduate students taking an introductory psychology course, while the older Americans were active, well-educated and community-dwelling volunteers recruited through a local community center. One hundred younger Chinese adults (18–23 years) and 100 older Chinese adults (59–76 years) were tested at the Institute of Psychology, Chinese Academy of Sciences, in Beijing, China. The younger Chinese were students recruited from three universities (Beijing Normal University, Capital Normal University, and Aeronautics and Space University), and the older Chinese were recruited from a pool of retired staff and faculty at the National Academy of Sciences, Beijing, China. Summary comparison statistics for all groups (table 1) indicate general concordance in terms of sample characteristics.

Stimulus Materials and Procedures

All 260 standardized pictures developed by Snodgrass and Vanderwart [12] – black outline drawings on a white background – were included, projected on a screen within a slide presentation. Experimental sessions were conducted 10–25 participants at a time. Chinese participants were given both verbal and written instructions in Mandarin; Americans were provided with equivalent instructions in English. Each picture was projected, singly in random order, for 8 s, followed by a 2-second pause. Every object projected on the screen was centered and approximately 4 inches squared. The pictures were displayed by a standard digital projector, situated 12 feet from the screen, at 800 × 600 resolution. Lighting in the room was slightly dimmed, but bright enough for participants to read the instructions and comfortably write out their responses.

Subjects were instructed to write down 'the first name of the object that comes to mind' for each presented picture, and to respond with an 'X' if they thought they had encountered the object before but didn't know, or could not remember, the object name, or with an 'O' if they had not encountered it (these 'do not know'

and 'cannot name' responses are included in the project archive (<http://agingmind.cns.uiuc.edu/Specificity>). Familiarity ratings were also elicited (see [15] which additionally analyzed name and concept agreement).

Subjects recorded their responses in individual paper booklets, and were instructed to provide a response for each picture. A 5-min break occurred between the first 140 picture items and the remaining 120. Finally, two short questionnaires were administered, eliciting demographic and health-related information, after which participants were debriefed, thanked for their participation and dismissed. The entire session lasted approximately 75 min. Younger Americans received credit towards subject pool requirements in a psychology course, and all other groups received monetary compensation for their participation in the study.

Specificity Coding

The 260 pictures tested were ordered alphabetically and numbered accordingly, with all measures discussed below compiled separately for each culture-by-age group. Data files for all 260 pictures and each of the four groups – younger American, older American, younger Chinese, and older Chinese – are freely archived at the project archive site (<http://agingmind.cns.uiuc.edu/Specificity>), as well as cross-referenced to comparison figures across groups on name agreement, concept agreement, familiarity, latency and other measures mentioned herein.

Various approaches to coding the name response data were evaluated in conjunction with psycholinguists knowledgeable about both languages and cultures. A final set of guidelines for counting different instances of names was established to ensure consistency and reasonableness across both American and Chinese name responses. For comparison purposes, the data were coded in a manner consistent with prior research. First, all name responses were recorded, with any obvious misspellings (e.g. homonyms) corrected. Second, when two or more responses were given, the first was retained (e.g. 'house' for 'house, home'). Third, quantifiers or articles accompanying name responses were removed (e.g. 'two', 'the', 'a', 'an'). Finally, any elaborations (e.g. 'index finger' and 'finger') or non-trivial abbreviations were each retained as separate name responses; this was crucial in coding for specificity.

Table 2. Distributions of naming specificity for ‘arm’ for Americans and Chinese, by age group

Group		Valid responses	Specificity				
culture	age		very high (2)	high (1)	moderate (0)	low (-1)	very low (-2)
American	younger	113	0	3	110	0	0
	older	103	0	4	91	8	0
Chinese	younger	100	0	9	78	3	10
	older	96	3	22	66	0	5

Each participant’s name responses for all 260 pictures were initially coded for nine specificity levels, on a –4 to 4 scale, which (ordinally) correspond to the standard levels of categorization (e.g. superordinate, basic, and subordinate). Consistent with extant findings in the categorization literature [25, 26], basic-level names (relative to each picture; see below) were coded as moderate in specificity and assigned a numerical score of 0. Responses that reflected greater detail (e.g. subordinate level) exhibited higher specificity and were assigned a score of 1 or greater, depending on the hierarchical level of detail subjects produced. Responses at more general levels of abstraction (e.g. superordinate level) were analogously coded as –1 or lower.

This initial nine-point specificity coding was performed for all responses (American and Chinese) for each of the 260 pictures by two independent judges fluently bilingual in English and Mandarin Chinese. Because less than 1% of overall responses fell into the ‘–4, –3’ or ‘3, 4’ categories, these were merged with the ‘–2’ and ‘2’ categories, respectively, yielding a five-point ordinal specificity scale (with empirical cell counts statistically consistent with the data model presented in the following section). Inter-rater reliability scores for the resulting five-point scale were over 98%; any remaining inconsistencies were resolved via discussion. Full codings for all picture items across each culture-by-age group, comprising over 4,000 unique picture item responses, is available from the project site.

It is crucial to note that coding is always relative to a particular picture item. For example, the response ‘chair’ would be coded at the superordinate level for the picture ‘rocking chair’ (item 188), but at the basic level for the picture ‘chair’ (item 53). Were this not the case, the codings would largely reflect hierarchical interrelations between the pictures themselves, a feature carefully and deliberately built-in by Snodgrass and Vanderwart [12]. Arguably, then, ‘basic’ level responses are correct responses, so that other levels suggest either more or less detail than warranted by the itself. And further, systematic deviations from the scale center suggest that a group perceives that picture item differently from how researchers may intend them to.

Analyses

We wish to understand whether, and how, naming specificity differs across the four culture-by-age groups. Data consist of specificity scores for each of the 260 pictures, that is, how many subjects in each group fell into one of five specificity categories: very high, high, moderate, low, very low. For example, counts for ‘arm’ (item 7), appear in table 2.

Observations thus consist of cell counts for an ordered categorical variable (specificity) on a 5-point scale, for four culture-

by-age groups and 260 pictures. Our goal is to determine whether the overall ‘pattern’ of responses – in a sense to be made rigorous below – differs across groupings of interest. Specifically, we address three sets of issues:

- 1 Does the pattern of response differ across age, culture, or any two culture-by-age groups?
- 2 Do any of the culture-by-age groups tend more towards naming specificity than the others?
- 3 For any culture-by-age group, do the pictures themselves fall into natural classes, and do any such classes vary across groups?

Addressing these questions requires a data model, one which describes the pictures and groups in a parsimonious fashion and which allows for clear statistical inference. Because we are interested in classification, we appeal to discrete latent class methods, a form of finite mixture model [27]. And, because we are modeling an ordered categorical variable in a parsimonious manner, we make use of the rank-order binomial model in particular. There is a broad literature on latent class and mixture model methods, and we direct the reader unfamiliar with their use to the primary literature (see [28] for a continually updated bibliography).

Here, we use Rost’s [29] rank-order binomial model, a standard tool in the area. It is especially parsimonious, describing the entire distribution of count data (i.e. across the five specificity levels, ‘very high’, ..., ‘very low’, coded 2, ..., –2) through a single parameter, π . This parameter represents ‘how far along the scale’ the underlying mean of the ordinal cell counts lies. In terms of visualization, we can view each of the lines in table 2 as a set of tosses of a (not necessarily fair) coin four times. For example, we have $n = 103$ responses for the ‘American, older’ group, and (ordered) cell counts of ‘0, 4, 91, 8, 0’. The rank-order binomial model would ask what single coin flip probability, π (of, say, ‘heads’), would be most likely to yield 0 ‘no heads’, 4 ‘one heads’, 91 ‘two heads’, etc., if we performed a set of four flips, $n = 103$ times. For a thorough introduction to the model, including likelihoods and estimation, see [30].

Thus, the latent class rank-order binomial model seeks out distinct classes of objects which can each be (parsimoniously) described by the same value of the parameter π . It allows for defensible statements of the sort ‘For the older Chinese group, there are three classes of pictures, with the following mean specificity parameters: π_1, π_2, π_3 ’. This in turn yields concrete inferences about how many distinct picture-classes there are in each culture-by-age group, which pictures fall into each, and how these differ across groups. The methodology therefore affords unambiguous responses to each of the three issues raised at the outset and is, to our knowledge, the only suitable method for doing so.

In accordance with the data model, the probability that the response of subject i for picture item j is k_{ij} (on an ordinal scale of 1, ..., K) is given by:

$$\Pr(\text{Observe } k_{ij}) = \binom{K-1}{k_{ij}-1} \pi_s^{k_{ij}-1} (1-\pi_s)^{K-k_{ij}}$$

This is a conditional probability, assuming we know that the picture item falls into class s , with parameter π_s . In reality, we do not know this with certainty, and the resulting heterogeneity (in pictures) is captured by (discrete) latent classes, with class-specific (mean scale location) parameters π_s . Note that these parameters can differ across culture-by-age groups or be constrained not to, although we have avoided subscripting them to reflect this; we take this issue up at length below.

The sample likelihood is merely a weighted sum across these latent classes, with weights equal to the class sizes (proportions), which must sum to one. If there are S latent classes, with class sizes ' m_1, \dots, m_s ', then the likelihood for the entire set of sample data is given by:

$$\Lambda = \prod_{j(\text{items})} \left[\sum_{s=1}^S m_s \prod_{i(\text{subjects})} \binom{K-1}{k_{ij}-1} \pi_s^{k_{ij}-1} (1-\pi_s)^{K-k_{ij}} \right]$$

All parameters, ' π_s ' for $s = 1, \dots, S$ and ' m_s ' for $s = 1, \dots, (S-1)$, were estimated via maximum likelihood, given raw data of the type in table 2, including all 260 pictures. Optimization was accomplished through a constrained Newton-Raphson algorithm, with multiple start points to help rule out local optima. Convergence was quick and consistent in all cases. (Complete estimation procedures and results are available from the authors, and are included in abbreviated form at the project archive site, along with all 'raw' picture-item specificity cell counts: <http://agingmind.cns.uiuc.edu/Specificity>.)

Although the method by no means guarantees it, each of the four culture-by-age groups appeared to be best fit by the same number (three) of latent classes, based on the standard fit measure, BIC (Bayesian Information Criterion; see [31] for additional detail on how the optimal number of classes is determined). Analysis could, in theory, proceed using these four separate (culture-by-age group) solutions, each with three classes (parameters). However, this would mean that the three classes derived for each separate culture-by-age group would not be the same across groups. By way of analogy, this is similar to four universities each being asked to parcel its students into three classes based on academic achievement; there would be no guarantee that the students could be compared across universities, if the universities weren't equally selective, or even if the variance in student performance differed across them (irrespective of the mean).

To allow just this type of comparison across groups, we constrain the latent class solution – that is, the mean scale values ' π_s ' and the relative sizes of the classes ' m_s ' – to be the same across all four culture-by-age groups. Because this is a parametric restriction, standard likelihood-ratio tests allow a comparison of using the latent class approach on each group separately vs. constraining the solution so that they are all estimated jointly. Because we found that this restriction does not provide a globally inferior fit, the data can be very simply, yet appropriately, explained as consisting of three discrete classes, each with its own mean specificity level. In the remainder, we will for simplicity call these the

Table 3. Latent class picture naming specificity proportions for Americans and Chinese, by age group

Group		Latent class naming specificity, %		
culture	age	high	moderate	low
American	younger	7.3	97.7	0.0
	older	9.6	90.4	0.0
Chinese	younger	20.8	73.8	5.4
	older	21.5	71.9	6.5

'low', 'moderate' and 'high' naming specificity classes. It is important to note that this solution does not presume that the same proportions of low, moderate and high picture items falls into each of the four culture-by-age groups. We will find that this is in fact not empirically the case, by comparing class membership probabilities derived from the model.

Results

Our results support, overall, three (latent) naming specificity classes, as follows. The low specificity class comprises 2.2% of picture items overall, with $\pi_{\text{low}} = 0.666$; the moderate specificity class comprises 84.2% of items, with $\pi_{\text{moderate}} = 0.487$; and the high specificity class comprises 13.6% of items, with $\pi_{\text{high}} = 0.299$. Recall that larger values of π (on its intrinsic unit scale) reflect lower specificity, and that the specificity scale was -2 to 2 ; transforming these values using $(-2 + 4 * \pi)$ therefore implies that the low, moderate and high classes are 'centered' at 0.664, -0.052 and -0.804 on the five-point Specificity scale, respectively. Thus, the moderate class, with 84% of the items overall, is very nearly at the scale center. This is reassuring in light of prior work which considered familiarity and frequency, though not specificity, to suggest suitable categories for research studies on age and/or culture.

Naming Specificity Comparisons across Culture-by-Age Groups

Table 3 lists the proportion of items in each of the three latent classes by group, and table 4 presents various tests regarding picture agreement among them.

It is immediately apparent that the Chinese groups are more dispersed among the three specificity classes. In fact, none of the pictures fell into the low class for either of the American groups; in interpreting this, two facts should be considered: (1) that the latent class solution arose from the conjoined data of the four groups, so that

Table 4. Latent class picture naming specificity: agreement (and disagreement) proportions for Americans and Chinese, by age group

Group	% Agreement			% Disagreement
	latent class picture naming specificity			
	high	moderate	low	
Comparison across all groups	5.4	66.9	0.0	27.7
Comparison within group				
American younger vs. older	6.9	90.0	0.0	3.1*
Chinese younger vs. older	17.7	68.1	4.6	9.6*
Younger American vs. Chinese	5.4	71.9	0.0	22.7
Older American vs. Chinese	8.5	71.2	0.0	20.4

* $p < 0.0001$ for differences in percentages between the two rows based on calculations from exact sampling distribution using Fisher's Exact test.

the dispersion of the Americans is intrinsically relative to that of the Chinese, and (2) that the latent solution fit no worse than one estimated for each of the four groups separately. Using the proportions in table 3 as a 'base', it is possible to test whether further constraints, of the form 'the proportions for group X must match those for group Y', are supported. By so doing, we can rigorously address the three issues at the outset, which we do in turn.

Differences in Response Patterns across Groups

Table 3 suggests, and nonparametric tests support, that there are no naming specificity differences between American-older and American-younger (n.s.), nor between Chinese-older and Chinese-younger (n.s.); and further that there are differences between American-older and Chinese-older ($p < 0.01$), as well as between American-younger and Chinese-younger ($p < 0.01$). That is, we see consistent differences across culture (within age), but not across age (within culture).

Relative Specificity across Groups

It is difficult to claim that any group uniformly demonstrates greater naming specificity than another. As per table 3, although neither of the American groups contained any low specificity pictures, the proportion of High specificity pictures – 7.3% for the younger and 9.6% for the older – were far smaller than the analogous figures for the Chinese (20.8 and 21.5%, respectively). Kolmogorov-Smirnov tests demonstrate that American and Chinese distributions are in fact strongly distinct ($p < 0.01$), though neither stochastically dominates the other (and, moreover, that no claims whatever can be made about age differences). We must conclude that neither culture nor age leads to directional naming specificity differences. However, naming specificity dispersion is far greater

among the Chinese, suggesting that they perceive many more pictures as lying outside their item-specific 'basic' level than do age-matched American counterparts. We know of no precedent for this finding in the picture norms literature. Whether this is an artifact of the Western genesis of Snodgrass and Vanderwart's [12] stimuli remains, of course, an open question, one which we hope future research will take up.

Latent Class Picture Agreement across Groups

The suitability of a particular picture can be determined by comparing which of the three classes it fell into across any groups of interest. Table 4 lists such proportions for 'all groups', 'all Americans', etc. For example, we find that, across all four culture-by-age groups (i.e. 'all groups'), 66.9% (or 174 in total) of the pictures always fell into the moderate specificity class, and another 5.4% (14) into the high; none fell into the low across all four groups. Because the moderate specificity class consists of those items perceived closest to their (picture-specific) 'basic' level, such picture items would be among the most broadly suitable stimuli choices. Further, we find that, for 27.7% (72) of the pictures, there is disagreement in terms of specificity (see 'Appendix' or project site); and moreover that many of these could not be anticipated based on prior studies using measures like latency and familiarity [15]. While we must stop short of suggesting these stimuli not be used in research on age or cross-cultural differences, any results stemming from them should be cautiously interpreted when subjects' responses are even partly verbal.

Table 4 presents similar figures for latent class agreement across the culture or age dimension. Of four possible comparisons, there is by far the greatest agreement across age for the American subjects; older Americans

and younger Americans disagree on the classification on only 3.1% (8) of the pictures. The analogous comparison for the older Chinese versus younger Chinese adults is 9.6% (25). Fisher's exact test indicates these proportions differ strongly ($p < 0.0001$). Disagreement values between Americans and Chinese for the younger as well as older adults are far higher: 22.7% (59 pictures) for the younger, and 20.4% (53 pictures) for the older adults; even though these proportions are not significantly different from one another (n.s.). In sum, the Snodgrass and Vanderwart [12] stimuli are in broad within-class agreement among Americans (younger vs. older: 96.9%), less so among Chinese (younger vs. older: 90.4%), and much less so among the younger (American vs. Chinese: 77.3%) and older (American vs. Chinese: 79.6%) groups.

Discussion

Owing to their relative neutrality compared with linguistic analogs, pictorial stimuli will continue to play a large role in studies of cross-cultural cognition. As such, it is crucial to identify ways in which standard stimuli systematically differ across various population groups, both to help verify general theories of cognition and toward the pragmatic end of selecting appropriate stimuli from the outset. Whereas prior research based stimuli comparisons on measures like name agreement, concept agreement, latency and familiarity, here we considered naming specificity, finding a distinct pattern across culture-by-age (younger-older, American-Chinese) groups.

Our analysis offers several advantages over prior approaches, and was made possible by the application of latent class techniques, which we believe deserve wider currency in cross-cultural research and cognitive studies in general. Combining latent class methods with a parsimonious description of the (ordered categorical) naming specificity measure allows the corpus of naming data itself to determine several key conclusions: how many naming specificity classes are supported; which pictorial stimuli fall into each; and whether there are indeed cohort differences. On this last issue, our findings are novel: while there are no differences across age groups (within culture), Chinese participants showed much greater dispersion in naming, with many more low specificity and high specificity items than their age-matched American counterparts.

In terms of agreement across all four culture-by-age groups, most items (67%) were uniformly in the moderate

specificity class, 5.7% were high, and none low. Thus, approximately 27% of the picture items differed across some pair of groups, calling into question their suitability for use in studies investigating age and/or cross-cultural differences. To the extent that the specificity with which objects are named is not explicitly accounted for, it could lead researchers to draw incorrect inferences about the presence or absence of differences in cognition. These findings highlight the importance of identifying subsets of pictorial stimuli that are judged suitable not only for particular research goals, but for the specific cultural or age groups being studied.

Acknowledgements

This study was supported by Grant R01 AG015047 from the National Institute on Aging awarded to A.H.G. We wish to acknowledge the assistance of Denise Park, Dick Nisbett, Katherine Burson, Phaythoune Chothmounethinh, Suzanne Hambright, Ping Hu, Alice Huang, Ting Luo, Serena Salloum, Cat Woods, and Vickie Yeung.

Appendix 1

Naming specificity classes for each culture-by-age group (1 = low, 2 = moderate, 3 = high)

Picture	Item	American		Chinese	
		younger	older	younger	older
1	accordion	2	2	2	2
2	airplane	2	2	2	2
3	alligator	2	2	2	2
4	anchor	2	2	2	2
5	ant	2	2	2	2
6	apple	2	2	2	2
7	arm	2	2	2	2
8	arrow	2	2	2	2
9	artichoke	2	2	3	3
10	ashtray	2	2	2	2
11	asparagus	2	2	3	3
12	axe	2	2	2	2
13	baby carriage	2	3	3	3
14	ball	2	2	3	2
15	balloon	2	2	2	2
16	banana	2	2	2	2
17	barn	2	2	2	2
18	barrel	2	2	3	3
19	baseball bat	2	3	3	3
20	basket	2	2	2	3

Picture	Item	American		Chinese	
		younger	older	younger	older
21	bear	2	2	2	2
22	bed	2	2	2	2
23	bee	2	2	3	2
24	beetle	2	2	2	2
25	bell	2	2	2	2
26	belt	2	2	3	3
27	bicycle	2	2	2	2
28	bird	2	2	3	3
29	blouse	2	3	1	1
30	book	2	2	2	2
31	boot	2	2	2	2
32	bottle	2	2	3	2
33	bow	2	2	2	3
34	bowl	2	2	2	2
35	box	2	2	2	3
36	bread	2	2	2	2
37	broom	2	2	2	2
38	brush	2	2	2	2
39	bus	2	2	2	2
40	butterfly	2	2	2	2
41	button	2	2	2	2
42	cake	2	2	2	2
43	camel	2	2	2	2
44	candle	2	2	3	3
45	cannon	2	2	3	3
46	cap	2	2	1	1
47	car	2	2	2	2
48	carrot	2	2	1	1
49	cat	2	2	2	2
50	caterpillar	2	2	1	1
51	celery	2	2	2	1
52	chain	2	2	3	2
53	chair	2	2	2	2
54	cherry	2	2	2	2
55	chicken	2	2	3	3
56	chisel	2	2	2	2
57	church	2	2	1	1
58	cigar	2	2	2	2
59	cigarette	2	2	2	2
60	clock	2	2	2	2
61	clothespin	2	2	1	1
62	cloud	2	2	2	2
63	clown	2	2	2	2
64	coat	2	2	2	2
65	comb	2	2	2	2
66	corn	2	2	2	2
67	couch	2	2	2	2
68	cow	2	2	3	3
69	crown	2	2	2	2
70	cup	2	2	2	2
71	deer	2	2	2	2
72	desk	2	2	3	3
73	dog	2	2	2	2
74	doll	2	2	3	3
75	donkey	2	2	2	2
76	door	2	2	2	2
77	doorknob	3	3	3	3

Picture	Item	American		Chinese	
		younger	older	younger	older
78	dress	2	2	1	2
79	dresser	2	2	3	3
80	drum	2	2	2	2
81	duck	2	2	2	2
82	eagle	2	2	2	2
83	ear	2	2	2	2
84	elephant	2	2	2	2
85	envelope	2	2	2	2
86	eye	2	2	2	2
87	fence	2	2	2	2
88	finger	2	2	3	3
89	fish	2	2	2	2
90	flag	2	2	2	2
91	flower	2	2	2	2
92	flute	2	2	2	2
93	fly	2	2	2	2
94	foot	2	2	2	2
95	football	3	3	3	3
96	football helmet	3	2	2	2
97	fork	2	2	1	1
98	fox	2	2	2	2
99	french horn	2	2	2	1
100	frog	2	2	2	2
101	frying pan	2	3	3	3
102	garbage can	3	3	3	3
103	giraffe	2	2	2	2
104	glass	2	2	1	2
105	glasses	2	2	3	3
106	glove	2	2	2	2
107	goat	2	2	2	1
108	gorilla	2	2	2	2
109	grapes	2	2	2	2
110	grasshopper	2	2	2	2
111	guitar	2	2	2	2
112	gun	2	3	3	3
113	hair	2	2	2	2
114	hammer	2	2	2	2
115	hand	2	2	2	2
116	hanger	2	2	3	3
117	harp	2	2	2	1
118	hat	2	2	2	2
119	heart	2	2	3	2
120	helicopter	2	2	2	2
121	horse	2	2	2	2
122	house	2	2	1	1
123	iron	2	2	2	3
124	ironing board	3	3	3	3
125	jacket	2	2	2	2
126	kangaroo	2	2	2	2
127	kettle	3	3	3	3
128	key	2	2	2	2
129	kite	2	2	2	2
130	knife	2	2	2	3
131	ladder	2	2	2	2
132	lamp	2	2	3	3
133	leaf	2	2	2	2
134	leg	2	2	2	2

Picture	Item	American		Chinese	
		younger	older	younger	older
135	lemon	2	2	2	2
136	leopard	2	2	2	2
137	lettuce	2	2	2	2
138	light bulb	3	3	3	3
139	light switch	3	3	2	3
140	lion	2	2	2	2
141	lips	2	2	2	2
142	lobster	2	2	1	1
143	lock	2	2	2	2
144	mitten	3	3	2	2
145	monkey	2	2	2	2
146	moon	2	2	2	2
147	motorcycle	2	2	2	2
148	mountain	2	2	2	3
149	mouse	2	2	2	2
150	mushroom	2	2	2	2
151	nail	2	2	2	2
152	nail file	3	3	2	3
153	necklace	2	2	2	2
154	needle	2	2	2	2
155	nose	2	2	2	2
156	nut	2	2	2	2
157	onion	2	2	2	2
158	orange	2	2	2	2
159	ostrich	2	2	2	2
160	owl	2	2	2	2
161	paintbrush	2	2	2	2
162	pants	2	2	2	2
163	peach	2	2	2	2
164	peacock	2	2	2	2
165	peanut	2	2	2	2
166	pear	2	2	2	2
167	pen	2	2	2	2
168	pencil	2	2	2	2
169	penguin	2	2	2	2
170	pepper	2	2	2	3
171	piano	2	2	2	2
172	pig	2	2	2	2
173	pineapple	2	2	2	2
174	pipe	2	2	2	2
175	pitcher	2	2	3	3
176	pliers	2	2	2	2
177	plug	2	2	3	2
178	pocketbook	2	2	2	2
179	pot	2	2	2	2
180	potato	2	2	3	3
181	pumpkin	2	2	2	2
182	rabbit	2	2	2	2
183	raccoon	2	2	2	2
184	record player	2	2	2	2
185	refrigerator	2	2	2	2
186	rhinoceros	2	2	2	2
187	ring	3	3	3	3
188	rocking chair	3	3	2	2
189	roller skate	2	2	1	1
190	rolling pin	2	2	2	2
191	rooster	3	3	3	3

Picture	Item	American		Chinese	
		younger	older	younger	older
192	ruler	2	2	2	2
193	sailboat	3	3	3	3
194	salt shaker	3	3	3	3
195	sandwich	2	2	2	2
196	saw	2	2	2	3
197	scissors	2	2	2	2
198	screw	2	2	2	2
199	screwdriver	2	2	2	2
200	seahorse	2	2	2	2
201	seal	2	2	2	2
202	sheep	2	2	2	1
203	shirt	2	2	2	2
204	shoe	2	2	3	3
205	skirt	2	2	1	1
206	skunk	2	2	2	2
207	sled	2	2	3	3
208	snail	2	2	2	2
209	snake	2	2	2	2
210	snowman	2	2	2	2
211	sock	2	2	2	2
212	spider	2	2	2	2
213	spinning wheel	2	2	2	2
214	spool of thread	2	3	3	3
215	spoon	2	2	2	2
216	squirrel	2	2	2	2
217	star	2	2	3	3
218	stool	2	2	2	2
219	stove	2	2	2	2
220	strawberry	2	2	2	2
221	suitcase	2	2	3	3
222	sun	2	2	2	2
223	swan	2	2	2	2
224	sweater	2	2	2	2
225	swing	2	2	2	2
226	table	2	2	2	2
227	telephone	2	2	2	2
228	television	2	2	2	2
229	tennis racket	3	3	3	3
230	thimble	2	2	3	2
231	thumb	2	2	3	3
232	tie	2	2	2	2
233	tiger	2	2	2	2
234	toaster	2	2	2	2
235	toe	2	2	2	2
236	tomato	2	2	2	2
237	toothbrush	2	2	2	2
238	top	2	2	2	2
239	traffic light	3	3	3	3
240	train	2	2	2	2
241	tree	2	2	2	2
242	truck	2	2	2	2
243	trumpet	2	2	1	1
244	turtle	2	2	2	2
245	umbrella	2	2	3	2
246	vase	2	2	3	3
247	vest	2	2	2	2
248	violin	2	2	2	2

Picture	Item	American		Chinese	
		younger	older	younger	older
249	wagon	2	2	2	2
250	watch	2	3	3	3
251	watering can	3	3	3	3
252	watermelon	2	2	2	2
253	well	2	2	3	3
254	wheel	2	2	3	3
255	whistle	2	2	2	2
256	windmill	2	2	2	2
257	window	2	2	2	2
258	wineglass	3	3	3	3
259	wrench	2	2	2	2
260	zebra	2	2	2	2

Appendix 2

Agreement across All Four Culture-by-Age Groups

Moderate Specificity: (1) accordion, (2) airplane, (3) alligator, (4) anchor, (5) ant, (6) apple, (7) arm, (8) arrow, (10) ashtray, (12) axe, (15) balloon, (16) banana, (17) barn, (21) bear, (22) bed, (24) beetle, (25) bell, (27) bicycle, (30) book, (31) boot, (34) bowl, (36) bread, (37) broom, (38) brush, (39) bus, (40) butterfly, (41) button, (42) cake, (43) camel, (47) car, (49) cat, (53) chair, (54) cherry, (56) chisel, (58) cigar, (59) cigarette, (60) clock, (62) cloud, (63) clown, (64) coat, (65) comb, (66) corn, (67) couch, (69) crown, (70) cup,

(71) deer, (73) dog, (75) donkey, (76) door, (80) drum, (81) duck, (82) eagle, (83) ear, (84) elephant, (85) envelope, (86) eye, (87) fence, (89) fish, (90) flag, (91) flower, (92) flute, (93) fly, (94) foot, (98) fox, (100) frog, (103) giraffe, (106) glove, (108) gorilla, (109) grapes, (110) grasshopper, (111) guitar, (113) hair, (114) hammer, (115) hand, (118) hat, (120) helicopter, (121) horse, (125) jacket, (126) kangaroo, (128) key, (129) kite, (131) ladder, (133) leaf, (134) leg, (135) lemon, (136) leopard, (137) lettuce, (140) lion, (141) lips, (143) lock, (145) monkey, (146) moon, (147) motorcycle, (149) mouse, (150) mushroom, (151) nail, (153) necklace, (154) needle, (155) nose, (156) nut, (157) onion, (158) orange, (159) ostrich, (160) owl, (161) paintbrush, (162) pants, (163) peach, (164) peacock, (165) peanut, (166) pear, (167) pen, (168) pencil, (169) penguin, (171) piano, (172) pig, (173) pineapple, (174) pipe, (176) pliers, (178) pocketbook, (179) pot, (181) pumpkin, (182) rabbit, (183) raccoon, (184) record player, (185) refrigerator, (186) rhinoceros, (190) rolling pin, (192) ruler, (195) sandwich, (197) scissors, (198) screw, (199) screwdriver, (200) seahorse, (201) seal, (203) shirt, (206) skunk, (208) snail, (209) snake, (210) snowman, (211) sock, (212) spider, (213) spinning wheel, (215) spoon, (216) squirrel, (218) stool, (219) stove, (220) strawberry, (222) sun, (223) swan, (224) sweater, (225) swing, (226) table, (227) telephone, (228) television, (232) tie, (233) tiger, (234) toaster, (235) toe, (236) tomato, (237) toothbrush, (238) top, (240) train, (241) tree, (242) truck, (244) turtle, (247) vest, (248) violin, (249) wagon, (252) watermelon, (255) whistle, (256) windmill, (257) window, (259) wrench, (260) zebra.

High Specificity: (77) doorknob, (95) football, (102) garbage can, (124) ironing board, (127) kettle, (138) light bulb, (188) ring, (191) rooster, (193) sailboat, (194) salt shaker, (229) tennis racket, (239) traffic light, (251) watering can, (258) wineglass.

References

- Lachman R, Lachman JL: Picture naming: retrieval and activation of long-term memory; in Poon LW, Fozard JL, Cermak LS, Greenberg D, Thompson LW (eds): *New Directions in Memory and Aging*. Hillsdale, Erlbaum, 1980, pp 313–343.
- Snodgrass JG: Concepts and their surface representations. *J Verbal Learn Verbal Behav* 1984;23:3–22.
- Kirsner K, Milech D, Stumpf V: Word and picture identification: is representational parsimony possible? *Mem Cognit* 1986;14:398–408.
- Mitchell DB, Brown AS: Persistent repetition priming in picture naming and its dissociation from recognition memory. *J Exp Psychol Learn Mem Cogn* 1988;14:213–222.
- McDermott KB, Roediger HL: Effects of imagery on perceptual implicit memory tests. *J Exp Psychol Learn Mem Cogn* 1994;20:1379–1390.
- Rajaram S: Perceptual effects of remembering: recollective processes in picture recognition memory. *J Exp Psychol Learn Mem Cogn* 1996;22:365–377.
- Nyberg L, Cabeza R, Tulving E: PET studies of encoding and retrieval: the HERA model. *Psychon Bull Rev* 1996;3:135–148.
- Mitchell KJ, Johnson MK, Raye CL, Mather M, D'Esposito M: Aging and reflective processes of working memory: binding and test load deficits. *Psychol Aging* 2000;15:527–541.
- Stark CEL, Squire LR: Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. *J Neurosci* 2000;20:7776–7781.
- Parkin AJ, Streete S: Implicit and explicit memory in young children and adults. *Br J Psychol* 1988;79:361–369.
- Park DC, Smith AD, Morell RW, Puglisi JT, Dudley WN: Effects of contextual integration on recall of pictures by older adults. *J Gerontol B Psychol Sci Soc Sci* 1990;45:P52–P57.
- Snodgrass JG, Vanderwart M: A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol [Hum Learn]* 1980;6:174–215.
- Snodgrass JG, Yuditsky T: Naming times for the Snodgrass and Vanderwart pictures. *Behav Res Methods Instrum Comput* 1996;28:516–536.
- Székely A, D'Amico S, Devescovi A, Federmeier K, Herron D, Iyer G, Jacobsen T, Bates E: Timed picture naming: extended norms and validation against previous studies. *Behav Res Methods Instrum Comput* 2003;35:621–633.
- Yoon C, Feinberg F, Luo T, Hedden T, Gutchess AH, Chen HY, Mikels J, Jiao S, Park DC: A cross-culturally standardized set of 260 pictures: American and Chinese norms for name agreement, concept agreement and familiarity. *Behav Res Methods Instrum Comput* 2004;36:639–649.

- 16 Alario FX, Ferrand L: A set of 400 pictures standardized for French: norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behav Res Methods Instrum Comput* 1999;31:531–552.
- 17 Bonin P, Peereman R, Malardier N, Méot A, Chalard M: A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behav Res Methods Instrum Comput* 2003;35:158–167.
- 18 Sanfeliu MC, Fernandez A: A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: norms for name agreement, image agreement, familiarity, and visual complexity. *Behav Res Methods Instrum Comput* 1996;28:537–555.
- 19 Cuetos F, Ellis AW, Alvarez B: Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behav Res Methods Instrum Comput* 1999;31:650–658.
- 20 Dell'acqua R, Lotto L, Job R: Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: direct comparisons with American, English, French, and Spanish published databases. *Behav Res Methods Instrum Comput* 2000;32:588–615.
- 21 Bates E, Andonova E, D'Amico S, Jacobsen T, Kohnert K, Lu CC, Szekely A, Wicha N, Federmeier K, Herron D, Iyer G, Pechmann T, Devescovi A, Orozco-Figueroa A, Gutierrez G, Hung D, Hsu J, Tzeng O, Gerdjikova G, Mehotcheva T, Pleh C: Timed picture naming in seven languages. *Psychon Bull Rev* 2003;10:344–380.
- 22 Matsukawa J: A Study of Characteristics of Pictorial Materials with Snodgrass and Vanderwart Pictures. *Memoirs of the Faculty of Law and Literature*. Shimane University, Japan, 1983.
- 23 Seo CW: 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 J Psychol* 1988;7:158–186.
- 24 Park DC, Nisbett RE, Hedden T: Aging, culture, and cognition. *J Gerontol B Psychol Sci Soc Sci* 1999;54:P75–P84.
- 25 Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P: Basic objects in natural categories. *Cognit Psychol* 1976;8:382–439.
- 26 Tanaka JW, Taylor M: Object categories and expertise: is the basic level in the eye of the beholder? *Cognit Psychol* 1991;23:457–482.
- 27 McLachlan GJ, Peel D: *Finite Mixture Models*. New York, Wiley, 2000.
- 28 Uebersax JS: *Latent Class Analysis 2004*, <http://ourworld.compuserve.com/homepages/jsuebersax/>
- 29 Rost J: A latent class model for rating data. *Psychometrika* 1985;50:37–49.
- 30 Kamakura WA, Wedel M: Life segmentation with tailored interviewing. *J Mark Res* 1995;32:308–317.
- 31 Kass R, Raftery A: Bayes factors. *J Am Stat Assoc* 1995;90:773–795.