Category norms with a cross-sectional sample of adults in the United States: Consideration of cohort, age, and historical effects on semantic categories

Nichol Castro^{1,2} · Taylor Curley¹ · Christopher Hertzog¹

Published online: 8 September 2020
© The Psychonomic Society, Inc. 2020

Abstract

This paper describes normative data for newly collected exemplar responses to 70 semantic categories described in previous norming studies (Battig & Montague, *Journal of Experimental Psychology*, 80(3, pt.2): 1–46, 1969; Van Overschelde, Rawson, & Dunlosky, *Journal of Memory and Language*, 50(3): 289–335, 2004; Yoon et al., *Psychology and Aging*, 19(3), 379–393, 2004). These categories were presented to 246 young (18–39 years), middle (40–59 years), and older (60 years and older) English-speaking adults living in the United States who were asked to generate as many category exemplars as possible for each of the 70 categories. In order to understand differences in normative responses, we analyzed these responses a) between age groups within the current sample and b) in comparison to three previously published sets of norms. Experimental studies using such norms typically assume invariance of normative likelihoods across age and historical time. We replicate previous findings such that exemplar frequency correlations suggest moderate stability in generated category members between age groups and cohorts for many, but not all, categories. Further, analyses of rank-order correlations highlight that the traditional measure of typicality may not capture all aspects of typicality, namely that for some categories there is high consistency in the frequency of exemplars across age groups and/or norms, but the ordering of those exemplars differs significantly. We include a cluster analysis to aid in grouping categories based on relative stability across time, cohort, and age groups. These results emphasize the importance of maintaining and updating age-differentiated category norms.

Keywords Verbal fluency · Exemplar typicality · Aging · Language · Semantic memory

Common nouns (e.g., apple and pear) can be classified as members (exemplars) of a superordinate taxonomic category (e.g., fruit; Rosch & Mervis, 1975). Historically, categorizable stimuli have played and continue to play an important role in research on memory, cognition, and language (e.g., Brainerd, Gomes, & Nakamura, 2015; Hertzog, Cooper, & Fisk, 1996; Hunt, Davidson, & Lansman, 1981; Hunt, Smith, & Toth, 2016; Lewellen, Goldinger, Pisoni, & Greene, 1993;

Maxfield & Zelinsky, 2012; McCloskey & Glucksberg, 1979; Smith, Shoben, & Rips, 1974). Scientific use of verbal stimuli from noun categories requires normative data on the exemplars and their typicality as representatives of a given category. For instance, research has shown that exemplar differences in category typicality affect response times when people are asked to verify category membership or whether pairs of stimuli are drawn from the same or different categories, although other variables like category dominance also play a role (Casey, 1992; Hunt et al., 1981).

Over 60 years have passed since the first category norm dataset was published by Cohen, Bousfield, and Whitmarsh (1957). Given their importance for experimental psychology, these norms have since been updated and expanded (Battig & Montague, 1969; Van Overschelde et al., 2004; Yoon et al., 2004). The common method for collecting normative data is to use a generation task, in which individuals are asked to name exemplars belonging to a provided category label (e.g., animals, vegetables, fruits). Exemplar typicality is then scaled to capture the aggregate frequency of exemplar

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-020-01454-9>) contains supplementary material, which is available to authorized users.

✉ Nichol Castro
ncastro1@uw.edu

¹ School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

² Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd St., Seattle, WA 98105, USA

productions in a sample of respondents. Additionally, in all of the three published norming studies (Battig & Montague, 1969; Van Overschelde et al., 2004; Yoon et al., 2004), university student participants provided the responses, with known limitations in the chosen sample (e.g., geographical influences on category exemplar generation).

From early on, there has been recognition that category representations can change with historical time given the dynamic nature of popular language usage and cultural change. Battig and Montague (1969) reported that their update was driven by Cohen's insistence for an update given that 8 years had passed since the Cohen et al. (1957) publication. The Van Overschelde et al. (2004) norms greatly expanded the number of categories relative to earlier efforts, including taxonomic labels such as "a football penalty" or "a type of car", to capture changes in normative taxonomic membership over time and cohorts (e.g., "In 1965, undergraduates wore *rubbers* on their feet, they did not listen to *rap*, and they did not drive *SUVs*"; p. 290). In addition, other researchers have extended category norms into other languages, cultures, and age groups (Brown & Davies, 1976; Hasselhorn, Jaspers, & Hernando, 1990; Howard, 1980; Marful, Díez, & Fernandez, 2015; Posnansky, 1978; Yoon et al., 2004). Thus, continuing to update the category norms is essential for maintaining a current basis of exemplars, and their typicality, for use in studying human memory and language.

One of the opportunities that has arisen due to the continued updating and expansion of the category norms is the ability to better understand the effect of aging on semantic category representation. Indeed, there remains a debate as to what effect, if any, aging has on semantic representations, and whether any observed age-related changes are actually due to aging processes or due to cohort or historical effects. On the one hand, researchers argue that semantic representations remain consistent across age groups (Bowles, Williams, & Poon, 1983; Burke & Peters, 1986; Howard, 1980; Yoon et al., 2004). In particular, Howard (1980) replicated Battig and Montague's (1969) procedure with the same stimuli in a sample of older adults, finding similar proffered exemplars and similar typicality rankings, a finding subsequently replicated by Yoon et al. (2004). However, this consistency might be most likely when examining basic categories of language concepts (e.g., animals, fruits, vegetables) that are less likely to change over historical time—and hence more likely to be invariant over birth cohorts. Invariance might be less likely with other categories, such as weapons or toys. For example, it is relatively unlikely that a 'phaser' or 'laser gun' would have been a generated exemplar prior to 1960. Furthermore, category fluency tests that present a taxonomic category label and request production of category exemplars (e.g., a verbal fluency task) typically find age differences and age changes in the number of valid exemplars produced (e.g., Bolla, Gray, Resnick, Galante, & Kawas, 1998; Brickman et al., 2005;

Gordon, Young, & Garcia, 2018; Troyer, Moscovitch, & Winocur, 2000), even for categories like 'animals'. This method is isomorphic with the typical exemplar-generation method for norming categorizable stimuli. It is possible that older adults' lower levels of exemplar production, which could be due in part to perseveration of exemplars (Henry & Phillips, 2006), could affect estimated typicality of less-common taxonomic exemplars in norming studies. Alzheimer's disease also appears to affect category fluency above and beyond declines seen under normal cognitive aging (Aschenbrenner et al., 2015; Brandt & Manning, 2009; Venneri et al., 2019). Furthermore, additional evidence suggests that older adults may not access or use category exemplars as efficiently, or in the same manner, as younger adults (Kintz & Wright, 2017; White, Storms, Malt, & Verheyen, 2018). For all these reasons, further evaluation of age differences in category norms seems important.

We provide another update of the category norms using adults in the United States. The last major updates based on young adult responses was published in 2004, as was the most recent study with older adults (Van Overschelde et al., 2004; Yoon et al., 2004). However, the current study goes beyond mere norm updating by engaging in an analysis of how category norms have changed, if any, with the passing of historical time. By capturing category norms across decades and across cohorts, researchers can better assess how semantic representations are influenced by cohort, age, and historical effects. Figure 1 provides a representation of the data available for comparisons across age groups and across cohorts at different measurements. Specifically, with the use of the Battig & Montague (1969) norms (henceforth called the "1969 norms"), the Van Overschelde, et al. (2004) norms (henceforth called the "v2004 norms"), the Yoon et al. (2004) norms (henceforth called the "y2004 norms") and the present norms (henceforth called the "2018 norms"), we can assess longitudinal, cross-sectional, and time-lag data. In particular, we focus on those comparisons where at least one parameter (i.e., time of measurement, cohort, or age at test) is maintained between groups. From Fig. 1, we can see four possible comparisons across groups: comparisons where participants are the same age at test (1969-YA to v2004-YA; 1969-YA to y2004-YA; 1969-YA to 2018-YA; y2004-OA to 2018-OA), are from the same cohort (1969-OA to 2018-OA), are the same age at test and from the same cohort (v2004-YA to y2004-YA; v2004-YA to 2018-YA; y2004-YA to 2018-YA), and are tested at the same time point (v2004-YA to y2004-OA; y2004-YA to y2004-OA; 2018-YA to 2018-OA).

Knowing which categories remain consistent and which do not across time, cohort, and age will be particularly useful for stimulus selection and research inquiry. For example, if a study aims to assess a process theory of cognitive aging (e.g., working memory deficits and/or inhibition deficits with linguistic stimuli), having a category representation that is

Fig. 1 A representation of the comparisons available using the Battig & Montague (1969) norms, the Van Overschelde et al. (2004) norms, the Yoon et al. (2004) norms, and the present 2018 norms in which one parameter (i.e., time of measurement, cohort, or age at test) is maintained

consistent across time and age is important so as not to build in a history-graded confound in stimulus properties into the cognitive task, creating a rival explanation for age differences in inhibitory function. On the other hand, if the research aims to assess how historical and developmental changes in semantic representation influence process theories of cognitive aging, then it may be particularly useful to select categories that are known to differ in normative properties across time and/or age. Thus, the present study provides an update of the category norms from a cross-sectional sample of adults in the United States, compares the new norms to previous datasets, and assesses the influence of cohort, age, and historical effects on categories.

Method

To keep as much consistency in methodology as possible, we modeled the current 2018 norms on the v2004 norms, which in turn modeled their norms on the 1969 norms. Table 1 provides a snapshot of the similarities and differences in methods between each of the norms. We elaborate here on the methodology of the current 2018 norms and refer the readers to the previous norms for more details on those. All of the data presented in this manuscript are publicly available here: categorynorms.psychology.gatech.edu.

Participants

We aimed to collect data from a heterogeneous sample of 250 adults, divided into three approximately equivalent age groups: young adults (18–39 years old), middle-aged adults (40–59 years old), and older adults (60 years and older). Participants were recruited using Amazon Mechanical Turk during the spring semester of 2018, with the criteria that participants were native English speakers and located in the United States. Data from four participants were excluded

due to failure to follow task instructions ($n = 3$) and an error in response recording ($n = 1$). The final sample consisted of 83 young adults ($M = 31$, $SD = 4.8$), 79 middle-aged adults ($M = 49$, $SD = 5.3$), and 84 older adults ($M = 67$, $SD = 4.6$), varying in gender, education, and location within the United States. Specifically, the young adult group was comprised of 48 females (35 males), ranged in education from elementary school to a doctorate degree ($M = 15.1$ years, $SD = 2.1$), and represented 29 different states. The middle-aged adult group was comprised of 57 females (22 males), ranged in education from high school to a master's degree ($M = 15.2$ years, $SD = 2.2$), and represented 34 different states. Lastly, the older adult group was comprised of 50 females (34 males), ranging in education from high school to a doctorate degree ($M = 16.0$ years, $SD = 2.9$), and represented 30 different states.

Procedure

The procedure was similar to that used in the v2004 norms but used Qualtrics and Amazon Mechanical Turk as the collection vehicle. The Qualtrics survey consisted of informed consent, demographic questions, and the norming task. All 70 categories from the v2004 norms, which overlaps in 56 categories with the 1969 norms and 47 categories with the y2004 norms, were shown to each participant in a random order one at a time. During the norming task, a category label was presented on the screen for 2 s and remained on the screen with a text box for input. Participants had 30 s to provide as many responses as possible by typing into a text box, and a timer was displayed for convenience. Moving forward before the 30 s expired was not allowed.

Analyses

Responses from participants were checked for spelling and corrected when possible (e.g., *sargent* corrected to *sergeant*). Any remaining unintelligible responses were removed (less

Table 1 Snapshot of methodology employed for each category norm dataset

	1969 norms	y2004 norms	v2004 norms	2018 norms
Participants	$n = 442$	$n = 216$ (113 young adults, 103 older adults)	$n = 633\text{--}710$ (varies across categories)	$n = 246$ ($n = 83$ young adults, 79 middle-aged adults, 84 older adults)
Age groups	1 group: young adults (no years of age provided, undergraduate students)	2 groups: young adults (17–23 years); older adults (60–75 years)	1 group: young adults (no years of age provided, undergraduate students)	3 groups: young adults ($M = 31$ years, $SD = 4.8$), middle-aged adults ($M = 49$ years, $SD = 5.3$), older adults ($M = 67$ years, $SD = 4.6$)
Sample recruitment	University students	Young adults = university students; Older adults = local community	University students	Amazon MTurk
Categories	56 categories	47 of the 56 categories from 1969 norms	56/56 categories from 1969 norms + 14 new categories	70/70 categories from the v2004 norms
Task	Provide as many exemplars as possible for each category in 30 s (written)	Provide 5 exemplars for each category (written; 20 min allotted for young adults and 30 min for older adults)	Provide as many exemplars as possible for each category in 30 s (typed)	Provide as many exemplars as possible for each category in 30 s (typed)

than 1% of the data). In addition, similar responses were combined under one superordinate response (e.g., responses of *boa*, *boa constrictor*, and *boa constrictors* were combined to create the reported superordinate response of *boa constrictor*), as done in the previous norms.

We conducted similar analyses as previous norming studies investigating age group differences in our sample. First, we assessed the potency, or average number of responses by participants, of each category by age group following the 1969 and v2004 norms. Then we conducted assessments of between-subject variability following Howard (1980), specifically by assessing the type-to-token ratio and the idiosyncratic type-to-total type ratio for each category by age group.

Additionally, we analyzed differences in our dataset from previous norms to consider cohort and time period effects, in addition to age-related effects. Recall that we will focus on those comparisons across norm datasets where at least one of the following parameters is maintained: time of measurement, cohort, or age at test (see Fig. 1). Note that we could only conduct correlations for categories tested by both norms; thus, not all comparisons have the same number of categories included in their correlation calculation. For example, the category “a drug” was only tested in the v2004 and 2018 norms. Thus, there are only two possible correlations calculated (i.e., v2004-YA to 2018-YA; 2018-YA to 2018-OA). There were 600 (out of 770) correlations possible for analysis. Specifically, we considered the frequency of exemplars in each category across norm datasets following the v2004 norms and extended that analysis to also consider the rank order of exemplars in each category across norm datasets.

Finally, for each set of correlations (i.e., frequency and rank order), we also conducted a cluster analysis on the 47 categories where all norm comparisons could be made to provide initial data-driven insight into how groups of categories may

begin to emerge based on patterns of frequency or rank-order correlations. We used a hierarchical clustering approach using the Ward method and the Mclust() function in R (Scrucca, Fop, Murphy, & Raftery, 2016) to determine the number of clusters (i.e., groups of categories) for each set of correlation analyses. We used standard parameters, did not set any priors, and report the results based on the best fitting model in an initial attempt to begin identifying groups of similar categories.

Results

Category norm dataset

The Appendix contains response data for each category as a separate table. In each table, we provide data for the overall sample (i.e., all age groups combined), along with the data broken apart by age group. The first column (*Total*) provides the proportion of participants that provided each response. We only list those responses that were given by at least 5% of the corresponding sample (overall and by specific age group), following the v2004 norms. The second column (*First*) provides the proportion of participants that produced the exemplar as their first response. The third column (*Rank*) provides the average output position and standard deviation of each exemplar. For example, in the category “a bird”, the response of *blue jay* was given as a response by 57% of participants overall (60% of young adults, 62% of middle-aged adults, and 51% of older adults), was given as the first response by 8% of participants overall (10% of young adults, 9% of middle-aged adults, and 5% of older adults), and had an average rank of 3.82 overall (3.96 for young adults, 3.73 for middle-aged adults, and 3.76 for older adults).

Category potency

Potency was calculated by dividing the number of responses given for a category by the total number of participants. Table 2 provides the mean and standard deviation of each category's potency for the overall data, as well as for each age group. For example, the category "a bird" had an overall mean potency of 7.15 (SD = 2.69), with a mean potency of 7.08 (SD = 7.08) for young adults, 7.63 (SD = 7.63) for middle-aged adults, and 6.75 (SD = 2.30) for older adults. This indicates that participants, on average, responded to the category cue "bird" by generating, on average, 7.15 exemplars treated as category-appropriate responses, with young adults providing 7.08 responses, middle-aged adults providing 7.63 responses, and older adults providing 6.75 responses. Overall, category potency ranged from 2.67 to 10.91, with similar ranges for each age group (young adults: 2.66–11.52, middle-aged adults: 2.53–11.14, and older adults: 2.81–10.10).

We planned a priori to examine the interaction of age group and category, and to conduct pairwise comparisons of age groups (young adults–middle-aged adults; young adults–older adults; middle-aged adults–older adults) for each category. To do so, we conducted a repeated measures ANOVA (age group by category) to assess age differences in the mean potency. We found no significant main effect of age group, $F(2, 243) = 1.062$, $p = .347$, but a main effect of category, $F(69, 16551) = 197.478$, $p < .0001$. However, these main effects should be considered in the context of a significant age group by category interaction, $F(138, 16551) = 2.66$, $p < .0001$. Using a Tukey's HSD corrected post hoc analyses, we found that 14 categories had at least one significant pairwise comparison between age groups (see Table 2). For example, there was a significant pairwise comparison between the middle-aged and older adult groups for the category "a bird", such that the middle-aged adult group had a higher mean potency than the older adult group, $p = .03$. For those categories with at least one significant pairwise comparison between age groups, the numerical difference in mean potency was small, ranging from 0.81 to 1.58.

Between-subject variability

To examine other aspects of between-subject variability, we considered the type-to-token and idiosyncratic type-to-total type ratios (Table 3). "Type" refers to a response provided by at least one person (e.g., "robin") and "token" refers to the number of times that type was provided (e.g., 31 tokens would indicate that 31 participants provided the type "robin"). In contrast, "idiosyncratic type" refers to a response provided by only one person. We considered both ratios for each age group by each category to determine how similar participant responses were within a given age group.

The type-to-token ratio reflects how common a particular type is within an age group. The more individuals that produce the same type for a given category, the lower the ratio value (i.e., the less between-subject variability). For example, using the category "a bird," young adults provided 71 different types with a total of 588 tokens (a value of 0.12), middle-aged adults provided 85 different types with a total of 603 tokens (a value of 0.14), and older adults provided 77 different types with a total of 567 tokens (a value of 0.14). Thus, for this specific category, the young-adult group had numerically lower between-subject variability than middle-aged and older adult groups.

The idiosyncratic type-to-total type ratio reflects the proportion of unique responses generated in each age group. Lower ratio values indicate fewer unique responses for a given category (i.e., less between-subject variability). For example, using the same category "a bird," young adults provided 17 idiosyncratic types with a total of 71 types (a value of 0.24), middle-aged adults provided 30 idiosyncratic types with a total of 85 types (a value of 0.35), and older adults provided 27 idiosyncratic types with a total of 77 types (a value of 0.35). Again, this specific category reflects numerically lower between-subject variability in the young adult group than the middle-aged and older adult groups.

To statistically assess between-subject variability in age groups across all categories, a Friedman's test was conducted for each of the ratios, as done by Howard (1980). First, a Friedman's test indicated that there was a significant effect of age group on the type-to-token ratio, $\chi^2(2) = 12.72$, $p = .002$. Using a symmetry test with corrected p values, the post hoc analysis indicated that the older adult group had a lower type-to-token ratio than in the middle-aged adult, $p = .03$, and younger adult, $p = .002$, groups, with no difference between the middle-aged and younger-adult groups, $p = .69$. Second, a Friedman's test indicated no significant effect of age group on the idiosyncratic type-to-total type ratio, $\chi^2(2) = 4.43$, $p = .108$. Thus, when considering all categories together, these ratio tests provide evidence of limited between-subject variability within age groups, namely that older adults had, on average, lower type-to-token ratios (i.e., more consistent responses) than the middle-aged and younger-adult groups, with no differences in the rate of idiosyncratic responses across age groups.

Exemplar frequency correlations

Van Overschelde et al. (2004) reported an analysis of exemplar frequency by conducting Pearson correlations between their v2004 norms and the 1969 norms, that they labeled a "generational correlation." They found significant variability in how stable exemplar frequencies were across generations/cohorts. In other words, there were several categories where the exemplar frequencies remained highly consistent between

Table 2 Potency of each category

Category	Mean potency (SD)				Significant pairwise comparisons
	Total	YA	MA	OA	
A bird	7.20 (2.62)	7.26 (2.67)	7.63 (2.85)	6.75 (2.30)	MA – OA: 0.88, $p = .03$
A building for religious services	3.67 (1.27)	3.56 (1.29)	3.68 (1.35)	3.76 (1.17)	-
A carpenter's tool	5.21 (1.91)	5.16 (2.23)	5.27 (1.72)	5.20 (1.78)	-
A chemical element	4.49 (2.41)	4.44 (2.54)	4.38 (2.31)	4.63 (2.40)	-
A city	7.11 (2.57)	7.39 (2.88)	7.29 (2.44)	6.66 (2.33)	-
A college or university	5.48 (2.22)	5.68 (2.41)	5.49 (2.29)	5.27 (1.95)	-
A color	9.37 (2.57)	9.77 (2.87)	9.46 (2.70)	8.89 (2.03)	YA – OA: 0.88, $p = .03$
A country	8.18 (3.07)	8.44 (3.45)	8.47 (3.11)	7.64 (2.57)	-
A crime	5.41 (1.86)	5.65 (2.08)	5.44 (1.79)	5.15 (1.67)	-
A disease	4.66 (1.83)	4.36 (1.81)	4.87 (2.05)	4.74 (1.59)	-
A drug	4.74 (1.90)	5.40 (2.28)	4.82 (1.59)	4.04 (1.48)	YA – OA: 1.36, $p = .001$
A female first name	8.86 (3.04)	9.41 (3.49)	9.01 (2.74)	8.18 (2.71)	YA – OA: 1.23, $p = .002$
A fish	6.14 (2.26)	5.78 (2.52)	6.53 (2.26)	6.13 (1.92)	-
A flower	5.19 (1.94)	4.93 (1.88)	5.31 (2.15)	5.35 (1.77)	-
A football penalty	2.97 (1.68)	2.95 (1.98)	2.82 (1.61)	3.15 (1.41)	-
A football position	4.84 (2.15)	4.94 (2.21)	4.49 (1.85)	5.08 (2.34)	-
A football team name	5.53 (2.69)	6.02 (3.09)	5.41 (2.35)	5.16 (2.54)	-
A four-footed animal	8.54 (2.87)	8.35 (3.36)	8.92 (2.88)	8.36 (2.27)	-
A fruit	7.98 (2.50)	8.18 (2.86)	8.42 (2.50)	7.37 (1.96)	YA – OA: 0.81, $p = .048$ MA – OA: 1.05, $p = .01$
A gardener's tool	5.04 (1.91)	4.65 (1.88)	5.20 (1.91)	5.27 (1.91)	-
A kind of money	6.14 (2.18)	5.93 (2.36)	6.24 (2.19)	6.26 (1.99)	-
A kitchen utensil	5.78 (1.89)	5.81 (2.04)	5.87 (1.64)	5.67 (1.97)	-
A liquid	5.39 (2.07)	5.24 (2.09)	5.55 (2.16)	5.39 (1.98)	-
A male first name	9.33 (2.89)	9.65 (3.18)	9.70 (2.77)	8.68 (2.61)	YA – OA: 0.97, $p = .01$ MA – OA: 1.02, $p = .01$
A member of the clergy	4.63 (1.76)	4.09 (1.75)	4.82 (1.75)	4.94 (1.68)	YA – OA: -0.85, $p = .03$
A metal	5.34 (1.88)	5.21 (2.06)	5.41 (1.77)	5.40 (1.82)	-
A military title	4.78 (1.78)	4.28 (1.71)	4.96 (1.58)	5.11 (1.92)	YA – OA: -0.83, $p = .047$
A musical instrument	7.51 (2.70)	7.51 (3.07)	7.77 (2.54)	7.26 (2.45)	-
A natural earth formation	5.30 (2.29)	5.44 (2.71)	5.37 (1.95)	5.10 (2.16)	-
A non-alcoholic beverage	6.07 (1.86)	6.01 (1.97)	6.49 (1.90)	5.74 (1.65)	-
A part of a building	6.71 (2.30)	6.60 (2.72)	6.70 (2.14)	6.84 (1.99)	-
A part of speech	4.83 (1.90)	4.42 (1.96)	5.01 (1.97)	5.05 (1.72)	-
A part of the human body	10.96 (3.47)	11.52 (3.93)	11.14 (3.26)	10.22 (3.05)	YA – OA: 1.30, $p = .001$ MA – OA: 0.92, $p = .02$
A precious stone	4.69 (1.70)	4.59 (1.76)	4.67 (1.71)	4.81 (1.65)	-
A relative	9.37 (2.58)	9.64 (2.91)	9.72 (2.62)	8.76 (2.06)	YA – OA: 0.88, $p = .02$ MA – OA: 0.96, $p = .01$
A science	4.19 (1.67)	4.34 (1.86)	4.17 (1.54)	4.06 (1.62)	-
A snake	4.83 (1.75)	4.60 (1.76)	5.00 (1.61)	4.89 (1.86)	-
A sport	7.31 (2.33)	7.47 (2.63)	7.72 (2.24)	6.77 (2.01)	MA – OA: 0.95, $p = .02$
A state	8.52 (3.29)	8.71 (3.54)	8.82 (3.02)	8.05 (3.27)	-
A substance for flavoring food	5.78 (2.20)	5.59 (2.37)	6.22 (2.31)	5.55 (1.87)	-
A thing made of wood	5.87 (2.15)	5.89 (2.37)	6.01 (2.19)	5.71 (1.88)	-
A thing taken from a burning home	5.43 (2.13)	5.18 (2.27)	5.95 (2.21)	5.19 (1.82)	-
A thing that flies	5.79 (1.94)	5.54 (1.98)	6.09 (2.20)	5.76 (1.60)	-
A thing that is green	5.44 (2.20)	5.54 (2.60)	5.66 (2.06)	5.14 (1.84)	-

Table 2 (continued)

Category	Mean potency (SD)				Significant pairwise comparisons
	Total	YA	MA	OA	
A thing that makes noise	6.24 (2.44)	6.27 (2.63)	6.41 (2.50)	6.06 (2.20)	-
A thing women wear	7.72 (2.85)	7.41 (3.03)	7.95 (2.68)	7.80 (2.83)	-
A toy	5.41 (2.08)	5.34 (2.17)	5.57 (2.00)	5.35 (2.09)	-
A tree	6.15 (2.34)	5.69 (2.22)	6.68 (2.29)	6.08 (2.43)	YA – MA: -0.99, $p = .005$
A type of car	5.80 (2.12)	5.83 (2.18)	5.71 (2.17)	5.86 (2.03)	-
A type of dance	5.10 (2.11)	5.07 (2.24)	5.00 (2.18)	5.22 (1.91)	-
A type of fabric	5.07 (1.82)	4.73 (1.84)	5.34 (1.84)	5.16 (1.74)	-
A type of footwear	5.86 (1.88)	6.01 (1.98)	6.03 (2.01)	5.54 (1.63)	-
A type of fuel	4.36 (1.60)	4.00 (1.46)	4.53 (1.73)	4.56 (1.58)	-
A type of human dwelling	5.84 (2.06)	5.76 (2.36)	6.10 (2.10)	5.66 (1.68)	-
A type of music	7.03 (2.37)	7.64 (2.68)	7.42 (2.29)	6.06 (1.76)	YA – OA: 1.58, $p < .0001$ MA – OA: 1.36, $p = .0003$
A type of reading material	5.16 (1.88)	5.23 (1.98)	5.22 (1.92)	5.05 (1.74)	-
A type of ship or boat	5.07 (1.84)	4.71 (1.87)	5.20 (2.02)	5.30 (1.58)	-
A type of vehicle	6.18 (2.04)	6.22 (2.31)	6.35 (2.01)	5.99 (1.77)	-
A unit of distance	5.60 (1.99)	5.48 (2.04)	5.84 (1.86)	5.49 (2.05)	-
A unit of time	7.63 (2.53)	7.49 (2.73)	7.87 (2.59)	7.54 (2.28)	-
A vegetable	6.52 (2.32)	6.45 (2.69)	7.01 (2.28)	6.12 (1.86)	MA – OA: 0.89, $p = .03$
A weapon	5.71 (2.07)	5.48 (2.13)	6.11 (2.30)	5.54 (1.70)	-
A weather phenomenon	5.91 (2.17)	5.82 (2.27)	5.99 (2.26)	5.92 (1.99)	-
An alcoholic beverage	6.15 (1.99)	6.18 (2.16)	6.34 (1.97)	5.93 (1.83)	-
An article of clothing	8.73 (2.95)	8.72 (3.16)	8.94 (3.27)	8.54 (2.40)	-
An article of furniture	6.51 (2.11)	6.41 (2.44)	6.65 (2.01)	6.48 (1.88)	-
An elective office	4.98 (1.77)	4.54 (1.97)	5.16 (1.68)	5.22 (1.58)	-
An herb	4.58 (2.07)	4.35 (2.17)	4.71 (2.33)	4.69 (1.70)	-
An insect	6.11 (1.98)	6.24 (2.28)	6.28 (1.89)	5.83 (1.71)	-
An occupation of profession	6.88 (2.26)	6.98 (2.36)	7.10 (2.28)	6.58 (2.14)	-

the 1969 and v2004 norms (e.g., “a color”), as well as several categories with significant changes in exemplar frequencies across norms (e.g., “a type of dance”; Van Overschelde et al., 2004). Table 4 provides a similar analysis of exemplar frequency correlations across norms, allowing for consideration of generational, historical, and age effects.

To compute the Pearson correlations, we obtained the frequency of each exemplar in each category for each norm dataset. We included all exemplars provided by at least 5% of the sample (rather than using a raw frequency count cutoff as done in the v2004 norms) so that consideration of low typicality responses would be included, while also reducing the influence of idiosyncratic responses and controlling for the issue in varying sample sizes across norms. Of the 600 correlations performed, over half ($n = 321$) had a Pearson's $r > .9$ and over 80% ($n = 498$) had a Pearson's $r > .75$, indicating high similarity in exemplar frequency across norm datasets (i.e., high-frequency exemplars tend to be consistently high

frequency, low-frequency exemplars tend to be consistently low frequency). For example, the exemplar frequency correlations across all group comparisons for the category “a precious stone” were extremely high (r range, 0.93–0.99). However, it is important to point out that the Pearson r values ranged from 0.01 to 0.99, which replicates the similarly substantial range in the generational correlation reported in the v2004 norms. For example, young adults' response frequencies for “a type of dance” from the current 2018 norms are highly correlated with those from young adults in the v2004 norms ($r = 0.91$); however, the same response frequency correlation between young adults from the current 2018 norms and those from the 1969 norms is practically zero ($r = 0.03$). This finding continues to harken the need to use up-to-date norms in psychological research, particularly for those categories where there may be significant influence of social and historical factors (e.g., “a disease”, “a toy”, “a weapon”, “a type of fuel”).

Table 3 Between-subject variability within age groups measured by type-to-token and idiosyncratic type-to-total type ratios

	Total Types / Total Tokens				Idiosyncratic Types / Total Types			
	Overall	YA	MA	OA	Overall	YA	MA	OA
A bird	0.06	0.12	0.14	0.14	0.19	0.24	0.35	0.35
A building for religious services	0.06	0.10	0.11	0.10	0.37	0.52	0.61	0.52
A carpenter's tool	0.09	0.16	0.18	0.15	0.34	0.53	0.55	0.50
A chemical element	0.10	0.18	0.21	0.19	0.28	0.40	0.40	0.38
A city	0.21	0.31	0.34	0.31	0.49	0.65	0.64	0.61
A college or university	0.21	0.34	0.35	0.34	0.41	0.63	0.60	0.58
A color	0.03	0.06	0.07	0.06	0.25	0.37	0.39	0.27
A country	0.06	0.15	0.12	0.14	0.19	0.34	0.34	0.34
A crime	0.08	0.17	0.15	0.15	0.23	0.39	0.41	0.45
A disease	0.13	0.23	0.24	0.22	0.34	0.55	0.53	0.46
A drug	0.13	0.15	0.20	0.29	0.38	0.43	0.53	0.67
A female first name	0.22	0.41	0.40	0.37	0.33	0.55	0.53	0.51
A fish	0.10	0.18	0.19	0.17	0.27	0.45	0.45	0.44
A flower	0.09	0.16	0.14	0.17	0.32	0.47	0.35	0.42
A football penalty	0.12	0.24	0.26	0.21	0.27	0.43	0.48	0.40
A football position	0.07	0.14	0.12	0.13	0.29	0.45	0.31	0.44
A football team name	0.06	0.11	0.14	0.14	0.25	0.29	0.41	0.35
A four-footed animal	0.05	0.11	0.11	0.11	0.15	0.30	0.32	0.29
A fruit	0.03	0.08	0.07	0.07	0.13	0.28	0.26	0.27
A gardener's tool	0.07	0.14	0.16	0.14	0.21	0.36	0.46	0.43
A kind of money	0.06	0.13	0.12	0.13	0.25	0.39	0.38	0.46
A kitchen utensil	0.07	0.14	0.15	0.13	0.23	0.49	0.43	0.41
A liquid	0.09	0.17	0.18	0.16	0.33	0.53	0.53	0.56
A male first name	0.18	0.34	0.32	0.28	0.35	0.55	0.53	0.51
A member of the clergy	0.07	0.14	0.13	0.10	0.35	0.52	0.55	0.42
A metal	0.05	0.08	0.07	0.09	0.32	0.35	0.35	0.51
A military title	0.05	0.11	0.12	0.07	0.24	0.46	0.56	0.30
A musical instrument	0.04	0.09	0.07	0.08	0.22	0.36	0.27	0.36
A natural earth formation	0.10	0.19	0.20	0.19	0.30	0.40	0.50	0.57
A non-alcoholic beverage	0.06	0.12	0.11	0.10	0.24	0.38	0.41	0.28
A part of a building	0.10	0.18	0.17	0.18	0.29	0.47	0.41	0.46
A part of speech	0.12	0.26	0.20	0.16	0.41	0.61	0.55	0.57
A part of the human body	0.04	0.08	0.08	0.09	0.19	0.33	0.26	0.31
A precious stone	0.04	0.10	0.08	0.08	0.24	0.47	0.21	0.26
A relative	0.02	0.05	0.06	0.05	0.07	0.08	0.21	0.26
A science	0.09	0.16	0.18	0.17	0.25	0.32	0.40	0.46
A snake	0.04	0.10	0.09	0.09	0.20	0.28	0.26	0.26
A sport	0.05	0.09	0.11	0.10	0.20	0.29	0.41	0.30
A state	0.03	0.07	0.08	0.08	0.03	0.06	0.02	0.02
A substance for flavoring food	0.09	0.18	0.16	0.18	0.30	0.46	0.42	0.45
A thing made of wood	0.13	0.25	0.22	0.23	0.33	0.50	0.51	0.54
A thing taken from a burning home	0.09	0.19	0.16	0.16	0.31	0.54	0.43	0.38
A thing that flies	0.08	0.14	0.17	0.12	0.34	0.43	0.59	0.36
A thing that is green	0.15	0.25	0.23	0.27	0.43	0.48	0.48	0.61
A thing that makes noise	0.14	0.26	0.25	0.26	0.37	0.47	0.48	0.53
A thing women wear	0.07	0.15	0.14	0.11	0.29	0.39	0.38	0.39
A toy	0.13	0.24	0.23	0.23	0.35	0.50	0.49	0.52

Table 3 (continued)

	Total Types / Total Tokens				Idiosyncratic Types / Total Types			
	Overall	YA	MA	OA	Overall	YA	MA	OA
A tree	0.08	0.16	0.15	0.14	0.28	0.45	0.43	0.33
A type of car	0.13	0.24	0.22	0.21	0.35	0.41	0.46	0.47
A type of dance	0.12	0.22	0.20	0.18	0.38	0.52	0.48	0.45
A type of fabric	0.08	0.12	0.14	0.17	0.36	0.30	0.53	0.53
A type of footwear	0.06	0.14	0.14	0.12	0.28	0.44	0.45	0.43
A type of fuel	0.07	0.15	0.15	0.13	0.29	0.38	0.47	0.33
A type of human dwelling	0.08	0.16	0.13	0.13	0.37	0.53	0.46	0.50
A type of music	0.08	0.14	0.13	0.15	0.38	0.52	0.40	0.47
A type of reading material	0.11	0.20	0.20	0.22	0.40	0.58	0.52	0.49
A type of ship or boat	0.10	0.16	0.19	0.18	0.35	0.35	0.41	0.43
A type of vehicle	0.10	0.19	0.18	0.14	0.39	0.52	0.49	0.46
A unit of distance	0.03	0.05	0.07	0.06	0.30	0.29	0.50	0.41
A unit of time	0.03	0.06	0.06	0.04	0.31	0.41	0.44	0.28
A vegetable	0.05	0.11	0.11	0.10	0.27	0.35	0.33	0.31
A weapon	0.10	0.19	0.18	0.18	0.34	0.44	0.44	0.52
A weather phenomenon	0.05	0.11	0.10	0.09	0.29	0.39	0.37	0.38
An alcoholic beverage	0.07	0.14	0.13	0.09	0.41	0.55	0.52	0.44
An article of clothing	0.05	0.12	0.10	0.09	0.29	0.48	0.39	0.38
An article of furniture	0.05	0.10	0.10	0.10	0.28	0.31	0.37	0.34
An elective office	0.06	0.11	0.11	0.09	0.39	0.45	0.57	0.54
An herb	0.07	0.15	0.13	0.13	0.32	0.47	0.32	0.37
An insect	0.05	0.09	0.13	0.10	0.27	0.30	0.49	0.32
An occupation or profession	0.15	0.25	0.24	0.25	0.40	0.54	0.51	0.50

A cluster analysis was also conducted to identify if some categories were more similar in their correlation patterns across norm comparisons than other categories. Using standard parameters and no priors, the Mclust() function (Scrucca et al., 2016) identified two clusters, or groups of categories (see Fig. 2), best fitted by the VEE model. Cluster 1 (Fig. 2, pink) consisted of ten categories (e.g., a type of dance, a disease, a toy, a type of music). Cluster 2 (Fig. 2, teal) consisted of the remaining 37 categories (e.g., a bird, a science, a vegetable, a relative, a color). We assessed the average frequency correlations for each norm comparison within clusters as an initial attempt at describing the clusters (see Table 5). Higher r values indicate greater consistency between the frequencies of exemplars produced within a category between the two norms compared. This simple assessment indicated that Cluster 1 represents categories that were more variable across norms (r values ranging from 0.465 to 0.871), while Cluster 2 represents categories that were more stable across norms (r values ranging from 0.849 to 0.952).

Exemplar rank-order correlations

We also evaluated Spearman's rank-order correlations (Table 6) using the average rank position of each exemplar in

each category for each norm dataset. We again included all exemplars provided by at least 5% of the sample. Of the 600 correlations performed, only 5% ($n = 32$) had a Spearman's $r > .9$, with the range of Spearman r values from -0.38 to 0.99 . These results stood in stark contrast to the previous exemplar frequency correlations. Specifically, here the rank-order correlations indicated that there was much less stability in the ordering of exemplars across norm datasets, even if the frequency of those exemplars was stable. For example, responses for the category "a four-footed animal" seem to be highly stable with regard to exemplar frequency correlations (r range, 0.90 – 0.98). However, differences in rank-order correlations were evident in key group comparisons: adults within the same cohort showed moderate stability in rank-order correlation (e.g., "2018 OA – 1969 YA": $r = 0.68$), but adults from different age groups tested at the same time point did not (e.g., "y2004 OA – y2004 YA": $r = 0.33$). In other words, the overall rates in which adults reported exemplars for "a four-footed animal" were highly similar across groups, but the order in which they reported these exemplars differed as a function of the idiosyncrasies inherent to different age and cohort groups.

Again, a cluster analysis was conducted to identify if some categories were more similar in their correlation patterns

Table 4 Exemplar frequency correlations across age groups and norms, where at least one parameter (age at test, cohort, or test point) is constant

Category	Same age at test				Same age at test and cohort				Same cohort		Same test point			
	1969 YA - 2018 YA	1969 YA - v2004 YA	1969 YA - y2004 YA	2018 OA - y2004 OA	2018 YA - v2004 YA	2018 YA - y2004 YA	2018 YA - v2004 YA	2018 YA - y2004 YA	2018 OA - 1969 YA	2018 OA - y2004 OA	2018 OA - 2018 YA	2018 OA - y2004 OA	2018 OA - y2004 YA	2018 OA - y2004 YA
A bird	0.68	0.75	0.84	0.87	0.93	0.86	0.91	0.88	0.88	0.87	0.87	0.85	0.85	0.74
A building for religious services	0.84	0.94	0.90	0.88	0.96	0.96	0.99	0.91	0.91	0.96	0.96	0.73	0.73	0.78
A carpenter's tool	0.92	0.89	0.91	0.95	0.97	0.96	0.98	0.95	0.95	0.94	0.94	0.79	0.79	0.74
A chemical element	0.89	0.94	0.80	0.70	0.96	0.83	0.85	0.88	0.88	0.91	0.91	0.89	0.89	0.69
A city	0.73	0.69	0.68	0.55	0.70	0.63	0.55	0.80	0.80	0.91	0.91	0.92	0.92	0.37
A college or university	0.48	0.19			0.25			0.44	0.44	0.96	0.96			
A color	0.98	0.98	0.92	0.91	0.99	0.90	0.92	0.99	0.99	0.98	0.98	0.98	0.98	0.89
A country	0.80	0.86	0.75	0.95	0.97	0.95	0.94	0.90	0.90	0.94	0.94	0.95	0.95	0.96
A crime	0.92	0.96	0.98	0.94	0.87	0.90	0.96	0.93	0.93	0.98	0.98	0.96	0.96	0.93
A disease	0.34	0.39	0.35	0.82	0.92	0.84	0.96	0.72	0.72	0.71	0.71	0.52	0.52	0.57
A drug					0.86					0.60	0.60			
A female first name	0.16	0.02			0.78			0.82	0.82	0.43	0.43			
A fish	0.65	0.79	0.82	0.87	0.85	0.77	0.89	0.80	0.80	0.86	0.86	0.94	0.94	0.77
A flower	0.85	0.87	0.85	0.93	0.98	0.96	0.98	0.89	0.89	0.94	0.94	0.80	0.80	0.87
A football penalty					0.89					0.84	0.84			
A football position					0.98					0.88	0.88			
A football team name					0.50					0.80	0.80			
A four-footed animal	0.93	0.95	0.90	0.95	0.98	0.94	0.95	0.97	0.97	0.96	0.96	0.91	0.91	0.92
A fruit	0.78	0.87	0.84	0.93	0.96	0.94	0.98	0.87	0.87	0.93	0.93	0.95	0.95	0.95
A gardener's tool				0.56	0.70	0.67	0.98			0.95	0.95	0.91	0.91	0.90
A kind of money	0.81	0.88			0.93			0.89	0.89	0.95	0.95			
A kitchen utensil	0.87	0.92	0.86	0.97	0.97	0.96	0.98	0.90	0.90	0.98	0.98	0.94	0.94	0.95
A liquid					0.97					0.96	0.96			
A male first name	0.66	0.52			0.86			0.87	0.87	0.79	0.79			
A member of the clergy	0.63	0.71			0.95			0.85	0.85	0.84	0.84			
A metal	0.93	0.94	0.92	0.88	0.99	0.97	0.97	0.96	0.96	0.98	0.98	0.80	0.80	0.86
A military title	0.87	0.92	0.94	0.96	0.94	0.94	0.99	0.95	0.95	0.91	0.91	0.95	0.95	0.93
A musical instrument	0.90	0.92	0.94	0.89	0.94	0.91	0.93	0.94	0.94	0.95	0.95	0.96	0.96	0.88
A natural earth formation	0.92	0.88	0.93	0.94	0.97	0.96	0.96	0.93	0.93	0.96	0.96	0.95	0.95	0.88
A non-alcoholic beverage	0.65	0.75	0.67	0.93	0.89	0.96	0.88	0.70	0.70	0.97	0.97	0.84	0.84	0.75
A part of a building	0.91	0.93	0.97	0.97	0.93	0.94	0.96	0.96	0.96	0.95	0.95	0.94	0.94	0.87
A part of speech	0.93	0.93			0.98			0.94	0.94	0.97	0.97			

Table 4 (continued)

Category	Same age at test				Same age at test and cohort				Same cohort		Same test point			
	1969 YA - 2018 YA	1969 YA - v2004 YA	1969 YA - y2004 YA	2018 OA - y2004 OA	2018 YA - v2004 YA	2018 YA - y2004 YA	2018 YA - v2004 YA	2018 YA - y2004 YA	2018 OA - 1969 YA	2018 OA - y2004 YA	2018 OA - 2018 YA	2018 OA - y2004 YA	2018 OA - y2004 YA	2018 OA - y2004 YA
A part of the human body	0.96	0.97	0.90	0.84	0.97	0.87	0.90	0.96	0.97	0.97	0.97	0.93	0.93	0.79
A precious stone	0.96	0.96	0.97	0.96	0.98	0.99	0.99	0.97	0.97	0.98	0.98	0.95	0.95	0.93
A relative	0.98	0.99	0.96	0.78	0.98	0.94	0.95	0.96	0.96	0.98	0.98	0.73	0.73	0.66
A science	0.74	0.80	0.78	0.74	0.95	0.96	0.95	0.73	0.73	0.93	0.93	0.75	0.75	0.69
A snake	0.62	0.65	0.62	0.79	0.98	0.95	0.97	0.70	0.70	0.94	0.94	0.87	0.87	0.80
A sport	0.88	0.89	0.87	0.98	0.97	0.98	0.98	0.87	0.87	0.98	0.98	0.95	0.95	0.94
A state	0.73	0.75	0.94	0.95	0.81	0.97	0.99	0.71	0.71	0.91	0.91	0.96	0.96	0.95
A substance for flavoring food	0.92	0.95	0.94	0.95	0.97	0.97	0.99	0.91	0.91	0.97	0.97	0.96	0.96	0.95
A thing made of wood					0.81					0.94	0.94			
A thing taken from a burning home					0.90					0.94	0.94			
A thing that flies					0.98					0.98	0.98			
A thing that is green					0.85					0.96	0.96			
A thing that makes noise					0.79					0.78	0.78			
A thing women wear					0.93					0.88	0.88			
A toy	0.71	0.61	0.73	0.87	0.92	0.84	0.77	0.89	0.89	0.87	0.87	0.70	0.70	0.53
A tree	0.90	0.86	0.92	0.92	0.94	0.95	0.93	0.94	0.94	0.93	0.93	0.85	0.85	0.78
A type of car					0.42					0.87	0.87			
A type of dance	0.03	0.01	0.04	0.83	0.91	0.74	0.87	0.59	0.59	0.67	0.67	0.44	0.44	0.40
A type of fabric	0.73	0.76	0.75	0.94	0.98	0.97	0.98	0.85	0.85	0.95	0.95	0.88	0.88	0.85
A type of footwear	0.39	0.40	0.47	0.87	0.92	0.85	0.90	0.58	0.58	0.93	0.93	0.88	0.88	0.77
A type of fuel	0.54	0.63	0.68	0.92	0.93	0.87	0.97	0.73	0.73	0.94	0.94	0.82	0.82	0.82
A type of human dwelling	0.83	0.88	0.84	0.92	0.93	0.93	0.97	0.88	0.88	0.95	0.95	0.90	0.90	0.87
A type of music	0.50	0.52	0.68	0.90	0.94	0.90	0.95	0.73	0.73	0.91	0.91	0.83	0.83	0.71
A type of reading material	0.69	0.69	0.67	0.97	0.98	0.96	0.98	0.68	0.68	0.98	0.98	0.99	0.99	0.98
A type of ship/boat	0.27	0.34	0.27	0.69	0.92	0.79	0.84	0.49	0.49	0.89	0.89	0.79	0.79	0.67
A type of vehicle	0.72	0.90	0.81	0.91	0.89	0.93	0.95	0.86	0.86	0.95	0.95	0.94	0.94	0.89
A unit of distance	0.84	0.85			0.98			0.85	0.85	0.98	0.98			
A unit of time	0.95	0.93	0.79	0.86	0.99	0.89	0.89	0.95	0.95	0.99	0.99	0.98	0.98	0.85
A vegetable	0.73	0.76	0.75	0.94	0.90	0.86	0.98	0.87	0.87	0.85	0.85	0.87	0.87	0.89
A weapon	0.89	0.41	0.92	0.95	0.47	0.98	0.44	0.97	0.97	0.94	0.94	0.92	0.92	0.43
A weather phenomenon	0.83	0.83	0.84	0.98	0.97	0.95	0.98	0.91	0.91	0.95	0.95	0.96	0.96	0.95
An alcoholic beverage	0.85	0.79	0.78	0.96	0.96	0.95	0.95	0.92	0.92	0.98	0.98	0.88	0.88	0.86
An article of clothing	0.78	0.75	0.75	0.91	0.97	0.94	0.96	0.91	0.91	0.93	0.93	0.78	0.78	0.75

Table 4 (continued)

Category	Same age at test			Same age at test and cohort				Same cohort		Same test point		
	1969 YA - 2018 YA	1969 YA - v2004 YA	1969 YA - y2004 YA	1969 YA - y2004 YA	2018 YA - v2004 YA	2018 YA - y2004 YA	2018 YA - y2004 YA	2018 OA - 1969 YA	2018 OA - y2004 YA	2018 OA - 2018 YA	y2004 OA - y2004 YA	y2004 OA - v2004 YA
An article of furniture	0.91	0.93	0.95	0.96	0.98	0.97	0.99	0.95	0.97	0.96	0.94	0.94
An elective office	0.76	0.95			0.86			0.77	0.96			
An herb				0.92	0.88	0.87	0.97		0.91	0.88	0.83	
An insect	0.91	0.96	0.97	0.95	0.97	0.93	0.96	0.96	0.95	0.94	0.91	
An occupation or profession	0.87	0.87	0.92	0.97	0.93	0.92	0.96	0.92	0.95	0.94	0.92	

*Blank cells indicate that one of the norms did not obtain responses for that category

across norm comparisons than other categories. Using standard parameters and no priors, the Mclust() function (Scrucca et al., 2016) identified five clusters, or groups of categories (see Fig. 3), best fitted by the EII model. Cluster 1 (Fig. 3, red) consisted of nine categories (e.g., a bird, a vegetable, a chemical element). Cluster 2 (Fig. 3, gold) consisted of 11 categories (e.g., a snake, an occupation or profession, a sport). Cluster 3 (Fig. 3, green) consisted of 15 categories (e.g., a crime, a metal, a unit of time). Cluster 4 (Fig. 3, blue) consisted of three categories (i.e., a type of dance, a disease, a science). Cluster 5 (Fig. 3, purple) consisted of nine categories (e.g., a toy, a weapon, a building for religious services). We also assessed the average frequency correlations for each norm comparison within clusters as an initial attempt at describing the clusters (see Table 7). Higher r values indicate greater consistency between the rank orders of exemplars produced within a category between the norms compared. While we initially found that there was much more variability in rank-order correlation r values across norm comparisons than exemplar frequency correlation r values, this clustering analysis provided a more fine-grained examination, allowing patterns to emerge. Clusters 1 and 3 had the most consistent average rank-order correlations across norm comparisons, with Cluster 3 having higher r values than Cluster 1. Thus, these clusters represent categories that were more stable across norms. Cluster 2 had low r values when comparisons were made including the Yoon et al. (2004) data, indicating that these norms were perhaps qualitatively different in their rank order as compared to the other norms analyzed. Cluster 4 represents categories that were highly influenced by age, cohort, and historical effects, as indicated by the low r values across the majority of norm comparisons. Finally, Cluster 5 had only very low r values for the norm comparison between younger and older adults in the present 2018 norms, perhaps indicating an age-related effect for the current period of time.

Based on the frequency and rank-order correlation clusters, we can begin to label categories on whether they are more or less stable across norms. Although we cannot be certain as to why some categories are more variable across norms (e.g., is variance due to age, cohort, or historical effects, or some combination), these labels may serve as a starting point for determining which categories need further investigation. Table 8 shows how categories might be labeled in regard to stability, from very low to high. Indeed, there are several categories which were quite stable across norms, in terms of both frequency and rank order of exemplars (e.g., “a crime”, “a fruit”, “a relative”). In contrast, there are some categories which were more variable across norms, particularly in regard to rank order. For example, while the category “a science” tended to have high consistency in exemplar frequency across norms, the ordering of the exemplars was highly variable.

Fig. 2 Cluster analysis of 47 categories across norms using frequency correlations. From *left to right*, the pink cluster (Cluster 1) represents categories with moderate frequency correlations across norms, while the teal cluster (Cluster 2) represents categories with high-frequency correlations across norms

Discussion

The present data set provides updated category norms using a heterogeneous sample of adults in the United States. There are two main differences of the present category norms in contrast to prior norms (Battig & Montague, 1969; Van Overschelde et al., 2004; Yoon et al., 2004): 1) a sampling of participants with no restriction on geographical location and 2) a sampling of participants from across adulthood who are not necessarily

tied to a specific university. Despite these differences, we provide strong evidence that the frequency of exemplars within a category tend to remain consistent, while also providing new insight regarding the variability in ranking of exemplars within categories. These updated norms will be useful for continuing investigations of language, memory, and cognition, particularly as it relates to cognitive aging.

One major impact that these data will have in cognitive aging research is the ability to have accurate exemplar

Table 5 Average frequency correlations across norms by cluster

	Norm comparison	Cluster 1 (pink) <i>n</i> = 10	Cluster 2 (teal) <i>n</i> = 37
Same age at test	1969-YA to 2018-YA	.502	.849
	1969-YA to v2004-YA	.465	.879
	1869-YA to y2004-YA	.544	.862
	2018-OA to y2004-OA	.819	.915
Same age at test and cohort	2018-YA to v2004-YA	.861	.952
	2018-YA to y2004-YA	.839	.931
	v2004-YA to y2004-YA	.822	.952
Same cohort	2018-OA to 1969-YA	.720	.900
Same test point	2018-OA to 2018-YA	.871	.948
	y2004-OA to y2004-YA	.769	.899
	Y2004-OA to v2004-YA	.607	.858

*Note color corresponds to Fig. 2

Table 6 Rank-order correlations across age groups and norms, where at least one parameter (age at test, cohort, or test point) is constant

Category	Same age at test				Same age at test and cohort				Same cohort		Same test point			
	1969 YA - 2018 YA	1969 YA - v2004 YA	1969 YA - y2004 YA	2018 OA - y2004 OA	1969 YA - y2004 YA	2018 YA - y2004 YA	2018 YA - v2004 YA	2018 OA - y2004 YA	1969 YA - 2018 OA	1969 YA - 2018 OA	2018 OA - y2004 YA	2018 OA - y2004 YA	2018 OA - y2004 YA	2018 OA - y2004 YA
A bird	.31	.51	.26	.56	.50	.42	.65	.48	.61	.31	.49	.61	.31	.49
A building for religious services	.68	.93	.82	.94	.61	.53	.85	.77	.20	.85	.92	.20	.85	.92
A carpenter's tool	.88	.83	.91	.46	.65	.61	.78	.46	.62	.38	.39	.62	.38	.39
A chemical element	.70	.57	.36	.82	.67	.44	.40	.77	.80	.44	.60	.80	.44	.60
A city	.70	.93	.50	.52	.33	.23	.31	.66	.56	.34	.40	.56	.34	.40
A college or university	.46	.32			.45			.13	.42			.42		
A color	.90	.95	.22	.35	.96	.43	.17	.88	.88	.21	.25	.88	.21	.25
A country	.52	.64	.28	.61	.70	.46	.30	.58	.56	.02	.27	.56	.02	.27
A crime	.80	.97	.96	.84	.69	.68	.79	.90	.81	.65	.76	.81	.65	.76
A disease	-.07	.40	.10	.31	.51	.62	.52	.19	.78	.27	.01	.78	.27	.01
A drug					.66				.57			.57		
A female first name	-.38	.70			.11			.38	.01			.01		
A fish	.52	.59	.26	.64	.49	.03	.43	.53	.44	-.07	.46	.44	-.07	.46
A flower	.42	.54	.91	.39	.85	.29	.39	.63	.64	.31	.45	.64	.31	.45
A football penalty					.89				.85			.85		
A football position					.48				.47			.47		
A football team name					-.12				-.10			-.10		
A four-footed animal	.59	.70	.44	.42	.63	.45	.57	.68	.65	.33	.34	.65	.33	.34
A fruit	.71	.85	.82	.61	.87	.75	.73	.86	.71	.67	.64	.71	.67	.64
A gardener's tool				.72	.94	.45	.66		.90	.25	.74	.90	.25	.74
A kind of money	.85	.96			.90			.83	.72			.72		
A kitchen utensil	.75	.89	.47	.42	.92	.84	.49	.49	.87	.50	.56	.87	.50	.56
A liquid					.68				.46			.46		
A male first name	.35	.62			-.05			.54	.14			.14		
A member of the clergy	.31	.64			.66			.86	.54			.54		
A metal	.69	.89	.83	.38	.82	.47	.58	.65	.65	.11	.63	.65	.11	.63
A military title	.31	.78	.76	.54	.62	.37	.86	.75	.32	.41	.51	.32	.41	.51
A musical instrument	.65	.85	.19	.34	.83	-.08	.45	.55	.74	.29	.46	.74	.29	.46
A natural earth formation	.46	.68	.54	.47	.66	.49	.85	.68	.28	.60	.67	.28	.60	.67
A non-alcoholic beverage	.86	.74	.53	.38	.86	.59	.70	.80	.91	.18	.31	.91	.18	.31
A part of a building	.46	.41	.01	.59	.50	.17	.72	.48	.21	.66	.41	.21	.66	.41
A part of speech	.97	.97			.95			.88	.90			.90		
A part of the human body	.58	.82	.62	.37	.74	.62	.74	.56	.71	.53	.47	.71	.53	.47

Table 6 (continued)

Category	Same age at test				Same age at test and cohort				Same cohort		Same test point			
	1969 YA - 2018 YA	1969 YA - v2004 YA	1969 YA - y2004 YA	2018 OA - y2004 OA	1969 YA - y2004 YA	2018 YA - y2004 YA	2018 YA - v2004 YA	2018 OA - y2004 YA	2018 OA - 1969 YA	2018 YA - y2004 YA	2018 OA - y2004 YA	2018 OA - y2004 YA	2018 OA - y2004 YA	2018 OA - y2004 YA
A precious stone	.94	.89	.72	.70	.89	.77	.73	.80	.84	.46	.45			
A relative	.92	.98	.37	.66	.96	.62	.84	.59	.73	.04	.63			
A science	.37	.68	.27	.30	.64	.61	.39	-.07	.69	.18	.23			
A snake	.82	.83	.45	.29	.74	.41	.49	.77	.47	.30	.32			
A sport	.67	.66	-.01	.42	.88	.17	.25	.62	.68	-.10	.56			
A state	.13	.34			.46			.28	.25					
A substance for flavoring food	.54	.81	.69	.43	.67	.42	.58	.90	.66	.42	.86			
A thing made of wood					.42				.40					
A thing taken from a burning home					.81				.56					
A thing that flies					.82				.43					
A thing that is green					.45				.70					
A thing that makes noise					.52				.69					
A thing women wear					.79				.69					
A toy	.69	.67	.59	.15	.35	.16	.52	.50	.07	.52	.33			
A tree	.65	.89	.57	.34	.70	.08	.54	.67	.58	.39	.38			
A type of car					.33				.55					
A type of dance	.24	-.04	-.07	.47	.63	.18	.32	.19	.83	.13	.23			
A type of fabric	.73	.95	.91	.56	.75	.51	.83	.72	.59	.06	.36			
A type of footwear	.25	.29	.51	.53	.51	.56	.68	.77	.46	.30	.78			
A type of fuel	.49	.44	.63	.29	.66	.59	.55	.38	.62	.57	.63			
A type of human dwelling	.59	.58	.57	.31	.68	.63	.80	.45	.45	.74	.84			
A type of music	.56	.68	.48	.06	.58	.27	.45	.86	.63	.30	.26			
A type of reading material	.94	.73	.70	.74	.79	.73	.86	.93	.76	.88	.89			
A type of ship/boat	.51	.47	.54	.42	.68	.73	.52	.46	.69	.40	.33			
A type of vehicle	.55	.84	.23	.01	.88	.01	.46	.75	.36	.08	.54			
A unit of distance	.84	.74			.83			.88	.87					
A unit of time	.99	.93	.60	.09	.89	.78	.70	.92	.92	.54	.04			
A vegetable	.41	.57	.45	.37	.58	.46	.62	.79	.48	.07	.60			
A weapon	.45	.74	.51	.74	.43	.28	.55	.78	.13	.53	.62			
A weather phenomenon	.63	.56	.69	.66	.66	.38	.61	.63	.28	.47	.56			
An alcoholic beverage	.46	.31	.34	.59	.82	.35	.63	.76	.30	.66	.54			
An article of clothing	.56	.58	.44	.38	.84	.40	.65	.73	.56	.43	.14			
An article of furniture	.85	.96	.53	.69	.88	.65	.55	.88	.85	.33	.69			

Table 6 (continued)

Category	Same age at test			Same age at test and cohort				Same cohort		Same test point		
	1969 YA - 2018 YA	1969 YA - v2004 YA	1969 YA - y2004 YA	2018 OA - y2004 OA	2018 YA - v2004 YA	2018 YA - y2004 YA	2018 YA - v2004 YA	2018 OA - 1969 YA	2018 YA - y2004 YA	2018 OA - 2018 YA	y2004 OA - y2004 YA	y2004 OA - v2004 YA
An elective office	.93	.86		.001	.93	.48	.47	.77		.80		
An herb					.68					.64	-.08	.05
An insect	.69	.84	.60	.50	.86	.67	.72	.80		.57	.52	.57
An occupation or profession	.72	.51	.42	.51	.83	.17	.42	.32		.56	.16	.28

*Blank cells indicate that one of the norms did not obtain responses from participants for that category

typicality estimates that are specific to different age groups. Recall that exemplar typicality is commonly obtained by scaling the mean number of times an exemplar is produced in a sample of participants, but previous norms usually rely only on young, college adult samples. This is a critical limitation of previous norms given that there is evidence to support change in the structure of semantic memory, or at least change in how words are retrieved from semantic memory, with increasing age. For example, Dubossarsky, De Deyne, and Hills (2017) provide evidence that a once tightly connected semantic memory structure begins to fragment into smaller clusters with increasing age (Dubossarsky et al., 2017), which is likely due in part to the degradation of exemplars on the periphery of categories (e.g., Borge-Holthoefer, Moreno, & Arenas, 2011). Although these changes in semantic structure could be accounted for by a change in higher-order cognition, like executive function and cognitive control (see reviews Siew, Wulff, Beckage, & Kenett, 2019; Wulff, De Deyne, Jones, & Mata, 2019), such that exemplars on the periphery are avoided in search tasks. Regardless of the underlying mechanism (i.e., change in structure or process), the fact that periphery exemplars are less accessible puts into question how exemplar typicality is determined, particularly for older adult samples. For example, a smaller category (i.e., fewer accessible exemplars) may impact the calculation of typicality (e.g., typicality will be larger for an exemplar if there are relatively few other exemplars to choose than if there are many other exemplars) and the rank ordering of exemplars (e.g., if there are many exemplars, variability in ordering may also increase). Using the common method of determining exemplar typicality, the current norms now provide a way to obtain exemplar typicality that is relative to younger (18–39 years), middle-aged (40–59 years), and older (60 years and older) adults (with publicly available access to the data for researchers to determine even more fine-grained age-associated typicality estimates), which may begin to address issues of age-related differences in semantic memory structure.

Although exemplar frequency correlations were moderate to high across norms, perhaps reflecting limited influence of age, cohort, and historical effects, our results on rank-order correlations indicated significant variability across norms, which points to a critical limitation of the current method in calculating exemplar typicality. That is, although an exemplar may reside in one's semantic memory (and be accessible to the extent of production during the category fluency task), the importance of that exemplar to the category may change over time and with age (Brosseau & Cohen, 1996; Verheyen, Droeshout, & Storms, 2019). Our clustering results suggest that the ordering of exemplars may be more sensitive to capturing age, cohort, and historical effects. That is, although an exemplar might be produced in equal proportion between two groups, this does not necessarily reflect that the exemplar is as equally important to both groups. For example, in the category

Fig. 3 Cluster analysis of 47 categories across norms using rank-order correlations. *From left to right*, the red cluster (Cluster 1) had moderate rank-order correlations, the gold cluster (Cluster 2) had low rank-order

correlations, the green cluster (Cluster 3) had high rank-order correlations, the blue cluster (Cluster 4) had very low rank-order correlations, and the purple cluster (Cluster 5) had moderate rank-order correlations

“a science”, the exemplar *computer science* had similar frequency proportions between young adults and older adults (5% and 6%, respectively), but differed in their average rank-order position (2.5 and 4.8, respectively). Here, we could make the suggestion that computer science has grown in recognition given the changing workforce demands and job opportunities, which are more relevant in the daily lives of younger adults who are seeking out their professions than older adults who are nearing retirement. In this case, *computer science* would have the same typicality estimate for the two

age groups given current methods, even though these two groups differ in how important this exemplar is to their category. Thus, the traditional measure of exemplar typicality may not be sensitive enough. New methods for estimating exemplar typicality can be developed to take into consideration rank ordering, as well as other potential influences, like feature saliency (Montefinese, Ambrosini, Fairfield, & Mammarella, 2014). This is beyond the scope of the present paper, but these norms could be used by other researchers to vet out alternative methods of estimating exemplar typicality.

Table 7 Average rank-order correlations across norms by cluster

	Norm comparison	Cluster 1 (pink) <i>n</i> = 9	Cluster 2 (gold) <i>n</i> = 11	Cluster 3 (green) <i>n</i> = 15	Cluster 4 (blue) <i>n</i> = 3	Cluster 5 (purple) <i>n</i> = 9
Same age at test	1969-YA to 2018-YA	.471	.660	.791	.180	.525
	1969-YA to v2004-YA	.518	.760	.872	.346	.628
	1869-YA to y2004-YA	.504	.326	.684	.100	.536
	2018-OA to y2004-OA	.464	.371	.522	.360	.554
Same age at test and cohort	2018-YA to v2004-YA	.657	.720	.816	.593	.592
	2018-YA to y2004-YA	.482	.198	.647	.470	.373
	v2004-YA to y2004-YA	.558	.388	.708	.410	.710
Same cohort	2018-OA to 1969-YA	.632	.653	.750	.103	.644
Same test point	2018-OA to 2018-YA	.612	.587	.746	.766	.248
	y2004-OA to y2004-YA	.351	.174	.418	.193	.604
	y2004-OA to v2004-YA	.484	.380	.550	.156	.600

*Note color corresponds to Fig. 3

Table 8 Stability of categories based on frequency and rank-order correlation clusters

		Frequency correlation clusters	
		Moderate (Cluster 1 - Pink)	High (Cluster 2 - Teal)
Rank-order correlation clusters	High (Cluster 3 - Green)		A carpenter's tool A crime A fruit A kitchen utensil A metal A non-alcoholic beverage A part of the human body A precious stone A relative A substance for flavoring food A type of fabric A type of reading material A unit of time An article of furniture An insect
	Moderate (Cluster 1 - Red & Cluster 5 - Purple)	A type of footwear A type of fuel A type of ship or boat A toy A weapon	A bird A building for religious services A chemical element A flower A four-footed animal A military title A natural earth formation A part of a building A type of human dwelling A vegetable A weather phenomenon An alcoholic beverage An article of clothing
	Low (Cluster 2 - Gold)	A city A snake A type of music	A color A country A fish A musical instrument A sport A tree A type of vehicle An occupation or profession
	Very Low (Cluster 4 - Blue)	A disease A type of dance	A science

Additionally, we suggest that researchers selecting categorizable stimuli for age-comparative work should attend carefully to the information provided here to identify categories with sufficient stability to warrant assumptions of measurement invariance across age and time (Baltes, Reese, & Nesselroade, 1988; Labouvie, 1980), or at least make attempts to account for potential age, historical, and cohort differences in category representation. Notably, there were several categories, based on our cluster analysis, which could be deemed as relatively stable across norms (i.e., across age, time, and cohort). For example, the category “a type of reading material” had both relatively consistent frequency (r range, 0.70–0.94) and rank-order correlations (r range, 0.68–0.99) across norm comparisons.

Thus, it is crucial that researchers continue to better disentangle which categories are more resilient to cognitive and societal factors, and which categories are not. Taken together, these findings are of crucial importance to researchers studying typicality effects on memory and language tasks, and harkens the need to better understand how our current method of estimating typicality, which does not take into rank ordering and has used norms from young adults, influences research findings, particularly in the context of cognitive aging research. Given the wide variety of research that is reliant on category norms and continuing concern regarding the replicability of psychological research (Arnett, 2016; Hanel & Vione, 2016; Henrich, Heine, & Norenzayan, 2010), it is critical that future norms continue to make attempts

at capturing the variability of individuals, including across ages and cohorts.

The findings of this study also draw attention to some of its limitations, and more broadly, to the very idea that norms can be generated with broad generalizability that remains valid across epochs of historical time. Our data are based on a relatively small sample and it would be desirable to generate large-sample norms that are not limited to sampling university students. It would also be desirable to be able to evaluate subject-level moderator variables (e.g., education, verbal ability, and region of the country), item-level moderator variables (e.g., the number of unique exemplars possible in a given category and the distribution of psycholinguistic features of exemplars, like lexical frequency, typicality, feature distinctiveness, etc.), and the interaction of subject-level and item-level moderator variables on potency and rank ordering of exemplars in a given category. Finally, the finding of historical effects and age effects suggests that norms might need to be continuously updated, as might be contemplated by regular, recurring sampling or even open-ended sampling where age and time-point of responses are recorded to enable fine-grained analysis of historical trends. Longitudinal data might be collected to differentiate within-person consistency from historical change as has been done in studies of adult intellectual development (e.g., Schaie, 2013). Finally, evaluating how exemplar generation norms differ from norms collected with other methods (such as rankings or ratings of provided exemplars) might be critical for accurate assessment of typicality for exemplars found to have limited production frequencies.

In closing, the results presented here show that the very concept of normative typicality must be qualified to consider the dynamic influences of changing language usage within and between persons across historical time. The age-graded norms we provide here are a resource than can be effectively leveraged to construct age-appropriate stimuli for studies of age differences in semantic memory access, episodic memory, and language production that rely in some manner on access to exemplars from semantic categories.

Acknowledgements This work was supported by a Ruth L. Kirschstein NRSA Institutional Research Training Grant from the National Institute on Aging of the National Institutes of Health (5T32AG000175) awarded to NC and TC.

Open Practices Statement The data generated and analyzed in the present study are available at categorynorms.psychology.gatech.edu.

References

- Amett, J. (2016). The neglected 95%: Why American psychology needs to become less American. In A. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 115–132). American Psychological Association.
- Aschenbrenner, A. J., Balota, D. A., Tse, C. S., Fagan, A. M., Holtzman, D. M., Benzinger, T. L. S., & Morris, J. C. (2015). Alzheimer disease biomarkers, attentional control, and semantic memory retrieval: Synergistic and mediational effects of biomarkers on a sensitive cognitive measure in non-demented older adults. *Neuropsychology*, 29(3), 368–381.
- Baltes, P., Reese, H., & Nesselroade, J. (1988). *Life-span developmental psychology: Introduction to research methods*. Hillsdale, NJ: Erlbaum.
- Battig, W., & Montague, W. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 80(3, pt.2), 1–46.
- Bolla, K., Gray, S., Resnick, S., Galante, R., & Kawas, C. (1998). Category and letter fluency in highly educated older adults. *The Clinical Neuropsychologist*, 12(3), 330–338.
- Borge-Holthoefer, J., Moreno, Y., & Arenas, A. (2011). Modeling abnormal priming in Alzheimer's patients with a free association network. *PLoS ONE*, 6(8), e22651.
- Bowles, N. L., Williams, D., & Poon, L. W. (1983). On the use of word association norms in aging research. *Experimental Aging Research*, 9(3), 175–177.
- Brainerd, C., Gomes, C., & Nakamura, K. (2015). Dual recollection in episodic memory. *Journal of Experimental Psychology: General*, 144(4), 816–843.
- Brandt, J., & Manning, K. (2009). Patterns of word-list generation in mild cognitive impairment and Alzheimer's disease. *The Clinical Neuropsychologist*, 23(5), 870–879.
- Brickman, A., Paul, R., Cohen, R., Williams, L., Macgregor, K., Jefferson, A., ... Gordon, E. (2005). Category and letter verbal fluency across the adult lifespan: relationship to EEG theta power. *Archives of Clinical Neuropsychology*, 20(5), 561–573.
- Brosseau, J., & Cohen, H. (1996). The representation of semantic categories in aging. *Experimental Aging Research*, 22(4), 381–391.
- Brown, W., & Davies, G. (1976). Studies in Word Listing: Testing for Group Differences in Category Norms. *The Irish Journal of Psychology*, 3(2), 87–120.
- Burke, D., & Peters, L. (1986). Word associations in old age: Evidence for consistency in semantic encoding during adulthood. *Psychology and Aging*, 1(4), 283–292.
- Casey, P. (1992). A reexamination of the roles of typicality and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(4), 823–834.
- Cohen, B., Bousfield, W., & Whitmarsh, G. (1957). Cultural norms for verbal items in 43 categories. (No. TR22). *CONNECTICUT UNIV STORRS*.
- Dubossarsky, H., De Deyne, S., & Hills, T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, 53(8), 1560–1570.
- Gordon, J., Young, M., & Garcia, C. (2018). Why do older adults have difficulty with semantic fluency? *Aging, Neuropsychology, and Cognition*, 25(6), 803–828.
- Hanel, P. H. P., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PLoS ONE*, 11(12), e0168354.
- Hasselhorn, M., Jaspers, A., & Hemando, M. (1990). Typizitätsnormen zu zehn Kategorien für Kinder von der Vorschule bis zur vierten Grundschulklasse [Category norms for verbal items in ten categories for children in kindergarten and in grades 1–4]. *Sprache & Kognition*, 9(2), 92–108.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83.
- Henry, J., & Phillips, L. (2006). Covariates of production and perseveration on tests of phonemic, semantic and alternating fluency in normal aging. *Aging, Neuropsychology, and Cognition*, 13(3–4), 529–551.

- Hertzog, C., Cooper, B., & Fisk, A. (1996). Aging and individual differences in the development of skilled memory search performance. *Psychology and Aging, 11*(3), 497–520.
- Howard, D. (1980). Category norms: A comparison of the Battig and Montague (1969) norms with the responses of adults between the ages of 20 and 80. *Journal of Gerontology, 35*(2), 225–231.
- Hunt, E., Davidson, J., & Lansman, M. (1981). Individual differences in long-term memory access. *Memory & Cognition, 9*(6), 599–608.
- Hunt, R., Smith, R., & Toth, J. (2016). Category cued recall evokes a generate-recognize retrieval process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 339–350.
- Kintz, S., & Wright, H. H. (2017). Semantic knowledge use in discourse: Influence of age. *Discourse Processes, 54*(8), 670–681.
- Labouvie, E. (1980). Identity versus equivalence of psychological measures and constructs. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological Issues* (pp. 493–502). Washington, DC: American Psychological Association.
- Lewellen, M., Goldinger, S., Pisoni, D., & Greene, B. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General, 122*(3), 316–330.
- Marful, A., Diez, E., & Fernandez, A. (2015). Normative data for the 56 categories of Battig and Montague (1969) in Spanish. *Behavior Research Methods, 47*(3), 902–910.
- Maxfield, J., & Zelinsky, G. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition, 20*(10), 1153–1163.
- McCloskey, M., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology, 11*(1), 1–37.
- Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). Semantic significance: A new measure of feature salience. *Memory and Cognition, 42*(3), 355–369.
- Posnansky, C. J. (1978). Category norms for verbal items in 25 categories for children in grades 2–6. *Behavior Research Methods & Instrumentation, 10*(6), 819–832.
- Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605.
- Schaie, K. W. (2013). *Developmental Influences on Adult Intelligence: The Seattle Longitudinal Study* (2nd ed.). Oxford University Press.
- Scrucca, L., Fop, M., Murphy, T., & Raftery, A. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal, 8*(1), 289–317.
- Siew, C., Wulff, D., Beckage, N., & Kenett, Y. (2019). Cognitive Network Science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity, 2019*, 2108423.
- Smith, E., Shoben, E., & Rips, L. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*(3), 214–241.
- Troyer, A., Moscovitch, M., & Winocur, G. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology, 22*(3), 370–378.
- Van Overschelde, J., Rawson, K., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language, 50*(3), 289–335.
- Venneri, A., Mitolo, M., Beltrachini, L., Varma, S., Della Pietà, C., Jahn-Carta, C., ... De Marco, M. (2019). Beyond episodic memory: Semantic processing as independent predictor of hippocampal/perirhinal volume in aging and mild cognitive impairment due to Alzheimer's Disease. *Neuropsychology, 33*(4), 523–533.
- Verheyen, S., Droeshout, E., & Storms, G. (2019). Age-related degree and criteria differences in semantic categorization. *Journal of Cognition, 2*(1), 17.
- White, A., Storms, G., Malt, B., & Verheyen, S. (2018). Mind the generation gap: Differences between young and old in everyday lexical categories. *Journal of Memory and Language, 98*, 12–25.
- Wulff, D., De Deyne, S., Jones, M., Mata, R., & Consortium, T. A. L. (2019). New perspectives on the aging lexicon. *Trends in Cognitive Sciences, 23*(8), 686–698.
- Yoon, C., Feinberg, F., Hu, P., Gutches, A., Hedden, T., Chen, H., ... Park, D. (2004). Category norms as a function of culture and age: comparisons of item responses to 105 categories by American and Chinese adults. *Psychology and Aging, 19*(3), 379–393.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.