Available online at www.sciencedirect.com
ScienceDirect
Acta Psychologica 124 (2007) 274–295
acta psychologica
www.elsevier.com/locate/actpsy

Age-of-acquisition differences in young and older adults affect latencies in lexical decision and semantic categorization

Simon De Deyne *, Gert Storms

Research Group of Higher Cognition and Individual Differences, Psychology Department, Tiensestraat 102, B-3000 Leuven, Belgium

Received 14 March 2005; received in revised form 3 March 2006; accepted 30 March 2006
Available online 14 June 2006

Abstract

An ongoing discussion about the role of age of acquisition (AoA) in word processing concerns the confound with word frequency. This study removed possible frequency confounds by comparing AoA and word familiarity differences in young (18–23 years) and older (52–56 years) adults. A first study investigated the differences in AoA and word familiarity ratings. The norms of AoA and familiarity were significantly different for young and older adults whereas these were previously considered equivalent [Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology*, *14*, 435–459]. In the second study, AoA and familiarity effects were significantly different for the older and young adults in a lexical decision task. The third study replicated these findings in a semantic artifact/naturally occurring categorization experiment, thus providing further evidence for AoA-effects when word processing requires semantic mediation. Results from both studies were in line with the hypothesis that AoA effects on word processing cannot be accounted for by word frequency or other possible confounds.

PsycINFO classification: 2343

* Corresponding author. Tel.: +32 16 326088; fax: +32 16 325916.
E-mail addresses: simon.dedeyne@psy.kuleuven.be (S.D. Deyne), gert.storms@psy.kuleuven.be (G. Storms).

0001-6918/$ - see front matter
doi:10.1016/j.actpsy.2006.03.007

Keywords: Psycholinguistics; Age differences; Lexical decision; Classification

1. Introduction

Consider the words *apple* and *mango*. These words are similar in that they both refer to a type of fruit. They are different, though, in the moment in life when they are learned by most people living in Western Europe (i.e., their age of acquisition; AoA) and in how frequently they occur in their natural language. When a person reads a particular word such as *apple* or *mango*, the latter word will take more time to process. The processing advantage for words like *apple*, as compared to words like *mango*, has been found in a number of tasks such as word and picture processing (e.g., Barry, Morrison, & Ellis, 1997; Brysbaert, Lange, & Van Wijnendaele, 2000; Gerhand & Barry, 1999; Morrison & Ellis, 1995, 2000).

A central and persisting issue is the reason for this processing advantage, with some authors accounting for it in terms of AoA, whereas others account for it in terms of word frequency. The key difficulty to settle this debate is that words that are learned early in life are often those words that occur most frequent in language. In other words, both variables correlate highly. Although this debate has been going on for more than 30 years (Carroll & White, 1973) and many studies have addressed the issue, the debate is still not settled.

One of the concerns that has recently been addressed in this debate is the use of the word frequency norms in studies where AoA effects were reported. Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004), for instance, investigated word naming and lexical decision performance for 2428 words. They reported considerable variability in the amount of variance accounted for by the different word-frequency estimates from Kučera and Francis (1967), CELEX (Baayen, Piepenbrock, & van Rijn, 1993), the Educator's Word Frequency Guide (Zeno, Ivens, Millard, & Duvvuri, 1995), HAL (Lund & Burgess, 1996) and MetaMetrics (MetaMetrics Inc., 2003) word frequency norms. Their findings showed that more recent and expanded sets of word frequency norms explained an additional 10% of the variance in a lexical decision task compared with older norms such as the widely used Kučera and Francis (1967) word counts. Similar findings reported by Zevin and Seidenberg (2002) cast doubt on AoA effects reported in a series of recent studies where words were matched for the Kučera and Francis frequencies. Zevin and Seidenberg showed that the matched lists significantly differed when other frequency measures such as the Educator's Word Frequency Guide (Zeno et al., 1995) and CELEX (Baayen et al., 1993) were used.

In this paper we report AoA effects in two studies that avoid the problems associated with word frequency (such as the size of the corpus, the sample of the texts used in generating the corpus, and the fact that word frequency measures underestimate exposure to words at a young age). Our studies succeed in avoiding these difficulties by using words that were introduced at different moments in the life span to otherwise comparable groups of participants. More specifically, we compared word processing latencies of participant groups with a different age and that therefore learned relatively recently introduced words at different moments in their lives. Returning to the example, since many exotic foods were not commonly available in Western Europe until recently, the AoA of a word like *mango* is different for older and younger persons, while variables like word frequency, word length, imageability, and number of neighbors do not differ. To understand the implications of this proposed procedure, a closer investigation of recent criticism on previously reported effects of AoA is required.

Most of the current views regarding the role and importance of AoA in numerous tasks depend on how word frequency is measured and interpreted. One reason why the AoA and frequency confound became a point of interest is the availability of expanded frequency norms like the ones mentioned earlier. However, recent arguments against previously reported effects of AoA in tasks such as reading aloud rely on distinguishing two notions: cumulative frequency and frequency trajectories (e.g., Zevin & Seidenberg, 2004).

Cumulative frequency reflects the total exposure to a word at a given moment in the life span. The number of exposures to words increases with age and reflects a more dynamic version of the traditional frequency measures (which refer to the frequency of word use at the moment when the norms were gathered). Related to cumulative frequency is the frequency trajectory of a word, a measure that reflects the distribution of word exposures over time. For instance, common words, like *apple*, have a relatively stable frequency distribution over the complete life span, whereas words like *gnome* are relatively more used in childhood than later in life.

Some researchers have argued that AoA is an outcome of the frequency trajectory (e.g., Bonin, Barry, Méot, & Chalard, 2004; Zevin & Seidenberg, 2002, 2004). In this view, the AoA of a word can be explained by the moment in life in which its exposure reaches a sufficient level, resulting in its acquisition. Other researchers attribute a more causal role to AoA and claim that solely the order of acquisition of representations affects processing irrespective of the factors that cause one word to be learned sooner than another (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, in press). According to this view, the variable AoA cannot be reduced to the frequency trajectory resulting in the word's acquisition, but captures the different structure of the lexicon when the word is acquired. The manipulation performed in our studies allows a contrast of both views. This can be done by comparing two groups of persons where the frequency trajectories of words can be considered identical, but the temporal order of acquiring these trajectories differs.

Crucial is the use of relatively recently introduced, but very common words. These are words that are acquired relatively late in young adults, but very late in older adults. Because these words are used very commonly now, current and cumulative exposure to these new words can be considered identical in both groups, and thus cumulative frequency differences as well as differences in frequency trajectory are of small concern. The exposure and type of the words is identical in different age groups, but the moment in life when participants are exposed to them differs strongly in these groups. According to the frequency trajectory view there should be no AoA effect when frequency trajectories are similar. According to the learning order view, AoA effects do occur, because the order in which words are learned differs irrespective of their frequency trajectory.

In a first study we investigated how young and older adults' ratings of AoA differ. Words were selected that presumably differed considerably in AoA estimates between young and older adults. Furthermore, word familiarity ratings (Gernsbacher, 1984) were collected to investigate whether all words were sufficiently known for both groups. The obtained group-specific ratings were used in two further studies in which the lexical decision processing and semantic categorization of younger and older adults were compared.

Although AoA effects in different age-groups have been compared to investigate effects of AoA before (e.g., Morrison, Hirsh, Chappell, & Ellis, 2002; Morrison, Hirsh, & Duggan, 2003), the present study was different in three important respects. First, our study differs from these earlier studies in that we compared results of relatively recently introduced words in two age groups, allowing us to isolate a possible AoA effect, independent of the possible nuisance variables described above. Second, these previous studies assumed that AoA of a given word would be comparable for different age groups. We investigated this claim in a norming study. Third, unlike these earlier studies that concentrated on naming tasks, we investigated whether AoA influences lexical decision and a semantic categorization, two tasks that arguably require deeper processing.

2. Norming of AoA and word familiarity for young and older adults

As mentioned in the introduction above, in Morrison et al.'s (2002) study with younger and older adults the same AoA norms were used for both groups. They defended the use of uniform norms on the basis of a study by Hodgson (1999) who found a high correlation between AoA ratings of young and old adults. However, Morrison et al. did not use words that are recently added to the common language, such as *mango* or *modem*. In contrast to most common words, such recently introduced words display large differences in AoA among young and older adults. To explicitly document these differences, norms for AoA and word familiarity were gathered for young and older adults. Such age-specific norms were necessary to manipulate AoA and to control for familiarity in two experiments described later in this paper.

An important concern when comparing results of two age groups is that both groups may have different exposures to the language material used. In order to minimize the exposure differences between a young and an older population, the older participants for which the norms were gathered came from a teachers population. Their students were teenagers that differed only a couple of years from the participants in the young group (first year university students) and are therefore exposed to similar patterns of language use. Furthermore, care was taken to restrict the age in both samples to be of an equal range.

2.1. Method

2.1.1. Participants

For the AoA ratings, 49 older adults with an average age of 54 years (range 51–56) and 52 young adults with an average age of 18 years (range 17–20) participated. For the word familiarity ratings, a different group was used. In this group, 48 older adults with an average age of 54 years (range 51–56) and 56 young participants with an average age of 18 years (range 17–23) participated. All participants were native Dutch speakers. The older adults were mainly teachers from secondary schools in Flanders (the Dutch speaking part of Belgium) who volunteered to participate in this study. All the young adults were students from the University of Leuven who participated for course credits.

2.1.2. Stimulus materials and rating procedure

We selected 309 Dutch nouns to allow for a maximum range of AoA.[1] The materials were composed to allow for differences in AoA both within and between age-groups and were selected from previous norms (e.g., Ghyselinck, De Moor, & Brysbaert, 2000), Dutch

[1] All materials are available upon request from the first author.

dictionaries and interviews with older persons. Words were either learned early (low AoA) or late (high AoA) for both the young and the older adults. Moreover, certain words such as *mango* are newly introduced over the last decades and are therefore acquired only recently by the older adults. There were four different randomized permutations. For each permutation the stimulus set was randomly divided in two subsets: one of 154 and one of 155 stimuli. Each participant completed only one list of 154 or 155 words.

As in Ghyselinck et al. (2000), participants were asked to indicate for each word at which age they estimated they first learned the word. If they did not know the word, they were asked to underline it. Every participant was handed a booklet that consisted of an instruction sheet and a sheet with one of the possible permutations.

The same stimulus materials and permutations described above were used to gather subjective word familiarity ratings. As in Gernsbacher (1984), participants were asked to indicate on a five-point scale how often they thought they encountered or used each word. Again, every participant was handed a booklet that consisted of an instruction sheet and a sheet with one of the possible permutations. On the sheet with the instructions, the five-point scale and its labels were explained. For every number on the scale the following labels were provided: *The word has been encountered or used: 1 – never*, *2 – almost never*, *3 – sometimes*, *4 – often*, *5 – very often*. This scale was also printed above the list with words on the second sheet.

2.2. Results and discussion

Two of the young participants provided ratings that deviated strongly from the others (a correlation below 0.50 with the mean ratings of the group). They were excluded from further analysis. Table 1 shows the number of participants, means and standard deviations, mean percentage known (based on the underlined words) and the Spearman–Brown split-half reliability measure for the AoA judgments of both age groups. As can be seen from Table 1, the reliability of these ratings for both groups was high.

For the familiarity ratings, two participants from the older group and four participants from the young group were removed because their ratings deviated strongly from the other results. The fourth column in Table 1 shows the Spearman–Brown split half reliability measure for the familiarity ratings of both age-groups. The reliability was again high for both groups.

Table 1
Average number of participants, mean, standard deviation and reliability for age of acquisition (AoA), familiarity (Fam) and percentage known (% known) for young and older adults

Age groups	*N*	Mean	SD	Reliability
Young				
AoA	50	8.2	2.7	0.97
Fam	52	2.9	0.5	0.94
% Known	50	98.3	6.3	0.91
Older				
AoA	49	14.8	10.0	0.99
Fam	46	3.1	0.5	0.93
% Known	49	97.2	17.7	0.85

Table 2
Correlations between AoA ($n = 309$), familiarity ($n = 309$) and CELEX word frequency (log(Freq), $n = 207$) for young (Young) and older (Older) adults

	AoA		Fam		log(Freq)
	Young	Older	Young	Older	
AoA					
Young		0.75**	−0.34**	−0.44**	−0.33**
Older			−0.10	−0.35**	−0.39**
Fam					
Young				0.74**	0.27**
Older					0.42**

Note. Significant correlations (two tailed).
* $p < 0.05$.
** $p < 0.01$.

The correlation matrix of the AoA and familiarity ratings of the young and older adults is provided in Table 2. Additionally, the correlation with log(Frequency) is given for 270 words for which CELEX (Baayen et al., 1993) word lemma frequency measures were available. The word lemma frequency counts in the CELEX are based on a corpus of 42,380,000 written words coming from a large variety of sources.[2]

Ghyselinck et al. (2000) collected AoA ratings for 2816 Dutch four- and five-letter nouns and found a correlation between AoA and CELEX log transformed frequency counts of $r = -0.59$. As can be seen in Table 2, in the present studies the correlations with the CELEX lemma values were lower ($r = -0.33$ for the young and $r = -0.39$ for the older adults). To further investigate the validity of the norms described in the present studies we compared them with the student ratings collected by Ghyselinck et al. in 2000 and a more recent study that collected AoA student ratings of words belonging to 49 semantic categories (Ghyselinck, Custers, & Brysbaert, 2003). There was an overlap of 136 words between our stimulus set and the set reported by Ghyselinck et al. (2000). The correlation between the AoA ratings of the young adults and the ratings by Ghyselinck et al. (2000) was very high, $r = 0.95$, but this correlation was lower for the older adults, $r = 0.74$. These results are not surprising since the participants from Ghyselinck et al. (2000) were approximately of the same age as our young adults but considerably younger than our older adults. There was an overlap of 172 words with the Ghyselinck et al. study from 2003. Similarly, there was a high correlation between their ratings and the ratings from the young adults, $r = 0.95$, whereas the correlation for the older adults was lower, $r = 0.66$. Note that both studies reported by Ghyselinck and her collaborators did not include all recently acquired words. For example, the rating study from 2000 only included words that were in the CELEX frequency count from 1993.

Next, we tested how AoA ratings for common words that were not recently added to the language correlated in the two age groups. To avoid the intrusion of new words, 202 words were selected that were learned before the age of 17 in both groups and where the rated

[2] Recent studies have used word form counts rather than lemma counts (e.g., Balota et al., 2004). Using word form counts instead of lemma counts, we obtained correlations very similar to those reported in Table 2, AoA and log(WordformFreq): $r = -0.33$ for the young, $r = -0.40$ for the older adults, familiarity and log(WordformFreq): $r = 0.28$ and $r = 0.41$.

AoA-difference between the groups was less than 5 years. The correlation was $r = 0.78$. The difference between the young ($M = 7.22$, SD = 2.14) and the older adults ($M = 9.38$, SD = 2.82) was significant, $t(201) = -23.34$, $p < 0.001$. Although Morrison et al. (2002) did not use new words, the current study thus indicates that even for common words, the AoA for young and older adults are not necessarily equivalent. One possibility is that the AoA ratings from the older adults are biased by the use of an anchor heuristic. In this case older adults would use their current age as a reference point in judging the AoA of certain words.

A similar comparison for the familiarity ratings yielded a significant difference (see Table 1), $t(308) = -6.02$, $p < 0.001$. These results are different from earlier studies where familiarity differences in young and older adults were investigated. For instance, Balota, Pilotti, and Cortese (2001), compared familiarity in young (mean age: 19.5 years) and older adults (mean age: 73.4 years) for monosyllabic English words. In their study, the familiarity norms for both groups were strongly correlated, $r = 0.92$, and there were no significant differences between both groups. As can be inferred from Table 1, the means for familiarity in the current study differed by only 0.2. This suggests that care should be taken to actually compare familiarity scores for two groups of participants, even when the correlation is high and the difference in means is small. Whereas previously reported studies deliberately avoided the use of recently acquired words because they might differ in AoA or familiarity, we specifically wanted to investigate differences between recent, early, and late acquired words. To allow such a detailed analysis, two subsets were selected from the stimuli set. A first set consisted of 202 common words that were not recently added to the language. The second set consisted of 67 words that were learned by the older adults after the age of 20. The criteria to remove the new words were identical as above. In the first subset the correlation between both groups was $r = 0.74$, whereas there was a significant difference for familiarity between the young ($M = 2.93$, SD = 0.49) and the older adults ($M = 3.16$, $S = 0.42$), $t(201) = -10.41$, $p < 0.001$. In the second set of new words the correlation for familiarity for both groups was $r = 0.78$. The familiarity ratings were significantly different, $t(66) = 2.09$, $p < 0.05$ for the young ($M = 2.92$, SD = 0.63) compared to the older adults ($M = 2.81$, SD = 0.64). For the set of common words, older adults judged words to be more familiar compared to the young adults, while the opposite pattern was observed for the new words. Note that these differences imply that (subjective) frequency norms for both populations should be based on age-specific corpora and cannot be assumed to be equal (contrary to common practice, as for instance in the study of Morrison et al., 2003).

An argument in line with this reasoning was also made by Balota et al. (2004), when they stated that there might be cohort differences in the judgment of subjective word frequency. This has been a recurring issue in most AoA-research, because word frequency measures are usually based on adult language found in newspapers that were often more than 10 years old when an experiment was performed.

3. Experiments

In the above described norming study, the norms for the young and older adults were different for AoA and word familiarity regardless of whether these norms included recently acquired or only common words. In the following two studies we proceeded with the same two age groups, but we used different participants to test whether these differences in AoA and familiarity can actually explain differences in processing time. Processing time was measured as decision latencies in two commonly used tasks, a lexical decision task and a semantic categorization task. The procedure of comparing two age-groups allowed a strong test of the effect of AoA in the absence of possible confounding word-frequency effects.

Since the norms for familiarity were different for young and older adults, we decided to use a correlation design instead of a full factorial design. Correlational analysis has been employed in numerous studies that investigated the effects of AoA and frequency on the speed of a variety of tasks (e.g., Barry et al., 1997; Brown & Watson, 1987; Gilhooly & Logie, 1981; Morrison & Ellis, 2000; Morrison et al., 2003). Recently, a number of researchers have defended the use of regression techniques instead of full factorial designs (e.g., Balota et al., 2004; Lewis, 1999) to avoid (amongst others) problems with bias in selecting items based on intuition (Rosenthal, 1995), and categorizing continuous variables such as word frequency in high and low frequency words which can lead to a decrease in statistical power and reliability (see e.g., Cohen, 1983; Humphreys, 1978; Maxwell & Delaney, 1993).

3.1. Lexical decision task

The first experiment was a lexical decision task. If the AoA effect is genuine, then there should be an effect caused by the different ages at which young and older adults acquired the words. Additionally, there should also be effects of AoA when using the age-specific norms within each group separately.

Since the norms showed that for certain words there are differences in word familiarity, this word familiarity difference between the two groups is taken into consideration. Furthermore, by using familiarity as an additional predictor a more stringent test of AoA can be obtained. Presumably, subjective word ratings in both measures are influenced by factors related to conceptual accessibility (Johnson, Paivio, & Clark, 1996). Adding both predictors in the analyses provides a correction for the subjective aspect of the AoA ratings that is also present in the familiarity ratings.

3.1.1. Method

3.1.1.1. Participants. Two groups of participants were used. The first group consisted of 22 young participants. They were all students at the University of Leuven who participated for course credits. The mean age of the participants was 18 years (range 18–20). The second group consisted of 20 older participants who volunteered and received a film ticket for their effort. Their mean age was 54 years (range 52–56). All participants spoke Dutch as their first language and had normal or corrected to normal vision. All the older adults received higher education after finishing secondary school. Many of them were teachers at secondary schools in Flanders or were employed at the University of Leuven.

3.1.1.2. Stimuli and procedure. We selected 108 words from our norms with the constraints that each word had a minimum word length of four and maximum word length of six letters. These words are listed in the third column of the Appendix. The words were selected to span the entire range of AoA and word familiarity as taken from the norms. The nonwords were four- to six-letter strings in accord with the orthographic rules of Dutch, created by replacing a vowel or a consonant of an existing word by another vowel or consonant (e.g., *fuir*, *zals*).

Students were tested in a soundproof room at the University of Leuven. The older adults were tested in quiet rooms at their homes and at various schools in Flanders. Participants were instructed to answer as quickly as possible, without making too many errors, by pressing one of the keys of a response box. The stimulus response mappings were counterbalanced over participants. There was a practice phase with 12 stimuli, 6 words and 6 nonwords. The practice and test phase were identical, except for a sound signal that was given when the response was incorrect during practice. The stimuli were presented using a 24pt lowercase Helvetica font on a 21″ Dell screen positioned at 60 cm from the participants. Each trial lasted for 4300 ms. A stimulus fixation point was presented for 300 ms and replaced by the stimulus word, which remained on the screen until the participant responded or until it timed out after 3000 ms. For the remainder of the trial, a blank screen was presented.

3.1.2. Results

Response latencies smaller than 200 ms and larger than 1400 ms were removed from the analysis. This resulted in a data loss of three responses or 0.07% of the data. The correct mean response time was significantly faster for the young participants ($M = 934$ ms, $SD = 80$) compared to the older adults ($M = 968$ ms, $SD = 100$), $t(101) = -4.97$, $p < 0.001$. The data of one young adult were removed because the percentage of errors was higher than 15%. The mean number of errors for the remaining participants was 5.7%. Five words (*klak*, *kabas*, *darts*, *ruif* and *slab*) were removed because more than 40% of young participants indicated these as nonwords. Although these are correct words, words like *klak, kabas* and *ruif* have an archaic use for some people and can be considered dialect words which participants might evaluate as nonwords. For the older adults, the only word that was wrongly indicated as a nonword in more than 40% of the cases was *gaia*. The word was therefore also removed from analyses. Following Ratcliff (1993), harmonic means were calculated for the decision latencies to further reduce the effect of outliers. Only the correct responses on the experimental trials were analyzed. As can be seen from the upper panel of Table 3, the correlations of the decision latencies, AoA, familiarity and word frequency measures were all significant.

Table 3
Correlations between response latencies (RT) of the Lexical Decision and Semantic Categorization task, AoA ($n = 309$), familiarity (Fam, $n = 309$) and CELEX word frequency (log(Freq), $n = 207$) for young and older adults

	RT		Fam		log(Freq)	
	Young	Older	Young	Older	Young	Older
Lexical decision ($n = 102$, n(log(Freq)) = 94)						
AoA	0.62**	0.74**	−0.40**	−0.28**	−0.36**	−0.52**
Fam	−0.45**	−0.36**	–	–	0.17**	0.30**
log(Freq)	−0.58**	−0.65**	–	–	–	–
Semantic categorization ($n = 146$, n(log(Freq)) = 130)						
AoA	0.36**	0.35**	−0.31**	−0.19**	−0.44**	−0.46**
Fam	−0.09	−0.18*	–	–	0.06	0.24**
log(Freq)	−0.20*	−0.17	–	–	–	–

Note. Significant correlations (two tailed).
* $p < 0.05$.
** $p < 0.01$.

Two classes of separate group regression analyses were used. First, simultaneous regression analyses within the young and older age group separately were performed on the harmonic means of the latencies averaged for each word over participants. These analyses are commonly reported and allow easier comparison with previously reported data. Second, for the subject-level analyses, the Lorch Myers (Lorch & Myers, 1990) procedure was used. It consists of first calculating the regression weights for each individual separately, and then running within every group, a group *t*-test to see whether the mean group values differed significantly from zero. In both types of regression, the contribution of AoA, word familiarity and word frequency for the young and older group were investigated. Finally a common analysis that used data fom both groups was carried out. In this analysis, difference scores based on the average decision latencies for young and older adults were predicted using AoA and familiarity.

Separate group analyses: Table 4 shows the regression coefficients for the first set of regression analyses. Table 5 contains the standardized regression coefficients from the Lorch Myers analysis and the corresponding *t*-values. Only effects significant in *both types* of regressions will be indicated as significant. In all regressions, the values of the appropriate collinearity diagnostics were investigated. Both tolerance and variance inflation factors indicated that multicollinearity was not a concern.

In the first regression model shown in Table 4, the decision latencies are predicted by regressing AoA and word familiarity on the mean decision latencies averaged over subjects. For both groups, AoA and word familiarity were significant. Next, the CELEX (Baayen et al., 1993) lemma log(Frequency) values were added to the equation. The CELEX

Table 4
Summary of regression analyses for variables predicting latencies in the lexical decision experiment

Regression analysis predictors	Young				Older			
	B	SE	*β*	*t*	*B*	SE	*β*	*t*
AoA and word familiarity ($n = 102$)								
Intercept	913	47.26		19.33***	959	47.04		20.40***
AoA	13.79	2.19	0.52	6.29***	6.48	0.63	0.70	10.28***
Fam	−35.93	12.20	−0.24	−2.95**	−34.44	13.95	−0.17	−2.47*
	$R^2 = 0.43$				$R^2 = 0.58$			
AoA, word familiarity (Fam) and CELEX log(Freq) ($n = 94$)								
Intercept	1020	47.63		21.42***	1059	48.40		21.88***
AoA	11.07	2.27	0.38	4.88***	3.87	0.94	0.35	4.13***
Fam	−34.55	11.06	−0.23	−3.12**	−24.61	13.38	−0.14	−1.84†
log(Freq)	−45.90	8.25	−0.41	−5.56***	−50.15	9.94	−0.43	−5.04***
	$R^2 = 0.58$				$R^2 = 0.54$			
	Older–Young							
Difference in AoA and Word familiarity ($n = 102$)								
Intercept	3.96	6.03		0.66				
AoA	4.92	0.61	0.63	8.13***				
Fam	−35.12	22.03	−0.12	−1.59				
	$R^2 = 0.43$							

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.
† $p < 0.1$.

Table 5
Summary of the mean standardized reaction time regression coefficients for the subjects analyses in the lexical decision experiment using the Lorch Myers procedure

Regression analysis predictors	Young		Older	
	β	*t*(20)	*β*	*t*(19)
AoA and word familiarity ($n = 102$)				
AoA	0.21	9.75***	0.37	13.97***
Fam	0.10	−4.81***	−0.09	−3.62***
AoA, word familiarity (Fam) and CELEX log(Freq) ($n = 94$)				
AoA	0.16	7.36***	0.15	4.81***
Fam	−0.11	−5.05***	−0.06	−2.25*
Freq	−0.20	−7.69***	−0.24	−9.83***

Significance in *t*-test based on two sided test.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.
† $p < 0.1$.

database provides frequency counts for only 94 of the 102 words. Therefore, only those 94 words where frequency counts were available were used. All predictors were significant, except word familiarity, which was only marginally significant for the older adults (Table 4). The analyses per participant showed similar effects as those per item: significant effects of AoA and familiarity in the first regression model and significant effects of AoA, familiarity and lemma frequency (Table 5). Together, these results imply that both AoA and word frequency contribute within each group to the explanation of the decision latencies.

Difference score analyses: Next, the differences between AoA, word familiarity and decision latencies of the young and older adults were calculated for every stimulus word. Specifically, for each word the difference scores for AoA were calculated by subtracting the mean AoA estimate of the young adults from the old. The same procedure was used for calculating the familiarity difference scores and the difference in decision latency. As can be seen in Table 4, there was a significant effect of AoA, but not of familiarity, in the prediction of the difference in response times.[3]

3.1.3. Discussion

The results of our experiments showed that, when controlling for word frequency, the difference in reaction time latencies between the young and older adults is determined only by the word's AoA, but not by the word's familiarity.

The individual regressions showed significant effects of AoA, word familiarity and word frequency for both young and older adults. Apart from the main finding of AoA differences for both groups, word frequency still plays a role in lexical decision. Contrary to the

[3] It is conceivable that the reaction time differences do not only reflect a difference in AoA but also age group differences in the speed of processing. For that reason, *z*-score transformations were used as recommended by Faust, Balota, Spieler, and Ferraro (1999) For each item, average trial-level *z*-scores were calculated by using each person's overall mean and standard deviation to convert the trials into a *z*-score. Next, the difference between the mean *z*-scores of the young and older adults was calculated. The results using these *z*-score transformations did not change the observed pattern of results: only AoA differences were significant for the prediction of the difference in decision latencies between the older and young adults.

word and object experiment by Morrison et al. (2002), who failed to find a frequency effect for the older adults, we observed effects of word frequency in the analyses for both groups separately.

To assess the validity of the use of difference scores in this study, the R^2 from the separate subject groups can be compared with the R^2 in the difference score regression model. The R^2 of these three analyses were similar, indicating that the effect of measurement errors in the difference scores did not substantially bias the findings.

3.2. Semantic categorization task

Morrison, Ellis, and Quinlan (1992) conducted an influential study that investigated whether the activation of semantic information could be the source of AoA effects. In an artifact vs. naturally-occurring decision task with pictures, they failed to find an effect of AoA whereas naming the same stimuli produced a strong AoA effect. Based on these findings, the authors concluded that AoA only affects semantic mediated tasks that require the retrieval or execution of object names. However, this study has subsequently been criticized because Morrison et al. used aggregate RT's over the two semantic classes and because stimuli with only a limited range of AoA were used (Brysbaert, Van Wijnendaele, & De Deyne, 2000).

Another problem with all studies that use semantic categorization with pictoral stimuli is that participants can perform the task using only surface characteristics of the stimuli. Recognizing a picture of a tiger as an animal can be based on the identification of eyes or ears. Likewise, detection of surfaces and edges may be enough to categorize something as an artifact.[4]

Using words instead of pictures in a semantic artifact vs. natural task, Ghyselinck (2002) found a significant effect of AoA for artifacts. Early acquired words were processed 49 ms faster than late acquired words. She also reported a significant frequency effect (91 ms) in this task. In the next experiment, we attempted to investigate the effect of AoA in a semantic task. Like in the first experiment, a comparison of young and older adults was carried out to control frequency and other confounds.

3.2.1. Method

Participants: The 21 young adults who participated in this experiment were students at the University of Leuven and had a mean age of 18 years (range 18–20). The 21 older adults had a mean age of 54 years (range 52–56) and a similar demographic background as in the lexical decision task. The incentives and demands for participation were identical to those in the lexical decision experiment.

Stimuli and procedure: The materials differed from the lexical decision experiment because only words that can be unambiguously categorized as either 'artifact' or 'natural' were used. Additionally, the number of words in the set was expanded to allow separate analyses within each category and to obtain a balanced set of natural kind and artifact items. The average ($M = 5.63$) and the range of the number of letters per word (3–10) differed from the stimulus set used in the first experiment ($M = 4.63$, range = 4–6). The 160 selected stimuli are presented in the first column of the Appendix. Except for the instructions, the procedure was completely identical to the lexical decision experiment.

[4] We want to thank a reviewer for pointing this out.

3.2.2. Results

All response latencies longer than 2800 ms were removed from the analysis. The correct mean response time was significantly faster for the young participants ($M = 1005$ ms, SD = 94) compared to the older adults ($M = 1169$ ms, SD = 132), $t(145) = -22.92$, $p < 0.001$. For 14 of the 160 stimuli, the percentage of errors was larger than 40% for one of the groups. As in the lexical decision experiment, these words were removed from further analysis. Harmonic means were used and incorrect responses were removed (3.7% of responses made by the younger group, and 3.5% of responses made by the older group). The presentation of the results below will be similar to the results in the lexical decision task. As can be seen from Table 3, all correlations were significant except for familiarity and word frequency.

Separate group analyses: The results of the first series of separate group regressions are presented in Table 6. The standardized regression coefficients obtained from the Lorch Myers procedure are listed in Table 7.

In the separate group regression model (Table 6), with AoA and word familiarity as predictors, AoA was significant for both the young and the older adults but familiarity reached significance only for the older adults. Next, the log(Freq) CELEX values were added to the equation. Since CELEX did not list frequency values for 16 words, 130 instead of 146 words were analyzed. As can be seen in Tables 6 and 7, there was an effect of familiarity in the older adults, but only in the Lorch Myers analysis. For both groups there was a significant effect of AoA, but no effect of frequency. Tables 6 and 7 show that

Table 6
Summary of regression analyses for variables predicting categorization latencies in the semantic categorization experiment

Regression analysis predictors	Young				Older			
	B	SE	*β*	*t*	*B*	SE	*β*	*t*
AoA and word familiarity ($n = 146$)								
Intercept	871	54.26		16.05***	1194	79.20		15.08***
AoA	11.68	2.64	0.36	4.43***	4.09	1.01	0.32	4.06***
Fam	3.94	14.73	0.02	0.27	−39.97	24.32	−0.13	−1.64*
	$R^2 = 0.13$				$R^2 = 0.14$			
AoA, word familiarity (Fam) and CELEX log(Freq) ($n = 130$)								
Intercept	884	73.79		11.98***	1177	95.10		12.38***
AoA	13.23	3.44	0.37	3.84***	4.60	1.63	0.27	2.83**
Fam	−0.20	16.30	−0.00	−0.01	−33.05	27.93	−0.10	−1.18
log(Freq)	−6.79	14.63	−0.04	−0.46	−5.41	22.99	−0.02	−0.24
	$R^2 = 0.15$				$R^2 = 0.10$			
	Older–Young							
Difference in AoA and word familiarity ($n = 102$)								
Intercept	134	9.01		14.87***				
AoA	2.35	0.60	0.30	3.90***				
Fam	−48.16	17.22	−0.22	−2.80**				
	$R^2 = 0.17$							

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.
† $p < 0.1$.

Table 7
Summary of the mean standardized reaction time regression coefficients for the subjects analyses in the semantic categorization experiment using the Lorch Myers procedure

Regression analysis predictors	Young		Older	
	β	$t(20)$	β	$t(20)$
AoA and word familiarity ($n = 102$)				
AoA	0.17	6.34***	0.17	6.45***
Fam	0.01	0.48	−0.06	−3.02**
AoA, word familiarity (Fam) and CELEX log(Freq) ($n = 94$)				
AoA	0.17	6.43***	0.15	6.13***
Fam	0.00	0.17	−0.05	−2.35*
Freq	−0.02	−1.02	−0.01	−0.59

Note. Significance in t-test based on two sided test.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.
† $p < 0.1$.

this result was obtained for both the separate regression analysis and the Lorch Myers analysis.

Difference score analyses: Next, the difference between AoA and word familiarity of the young and older adults was calculated. The reaction times of the young were subtracted from those of the older adults. As indicated in Table 6, there was a significant effect of both AoA and word familiarity on the difference scores. As in the lexical decision experiment, for each item, the mean and standard deviation were calculated over all participants. Next, the observed decision latencies for each subject were normalized and the difference between the mean z-scores of the young and older adults was calculated. The results using the z-score transformations indicated that the difference in AoA was a significant predictor of the difference in decision latencies between the older and young adults. The word familiarity differences failed to yield significance.

3.2.3. Discussion

The results of the semantic categorization task regarding the effects of AoA were similar to the findings in the lexical decision task. In all analyses, an AoA effect was found. However, word frequency was never significant and word familiarity was only significant for the older adults. Analysis of the difference regressions (i.e., Lorch Myers analysis) showed significant effects for AoA and word familiarity. The significant intercept in the difference scores and the elimination of the familiarity effect after z-score transformation indicate that the initial significant familiarity effect when regressing on the raw RT differences could be due to slower general processing by the older adults. This can be explained by the increased complexity of the task, compared to the lexical decision task, where no significant difference for the intercept was found.

Although the pattern of results in the semantic categorization replicated the results of the lexical decision task, the R^2 of the categorization task were often somewhat lower than those of the lexical decision task. Presumably, measures of category centrality such as typicality also played an important role in the semantic categorization task and co-determined the reaction times to a large extent (Storms, De Boeck, & Ruts, 2000).

Additional analyses (which we will not further describe here for reasons of conciseness) revealed that the effects of AoA were significantly larger for the natural items. Related to this, Ghyselinck (2002) used a similar task but analyzed only the artifact items. She found, contrary to our study, both frequency and AoA effects. In a recent brain imaging study (Fiebach, Friederici, Müller, Cramon, & Hernandez, 2003) AoA was found to modulate activity in brain areas not affected by word frequency. Likewise, Fiebach et al. (2003) found that AoA modulates activity in areas devoted to semantic retrieval as well, but these findings were limited to inanimate objects. Our results suggest that the effects of AoA would be even stronger in studies (like those of Ghyselinck & Fiebach et al.) that include both artifact and natural kinds.

4. General discussion

The main finding of this article is that robust effects of AoA were observed in two experiments using difference scores of AoA and word familiarity estimates. Using this method we have avoided some pitfalls associated with matched designs and cross sectional studies that did not include age-specific AoA and familiarity ratings. The importance of these separate ratings per age group was shown through the comparison of age-specific norms gathered in Study 1. For both common and recently introduced words there was a significant difference between the AoA and word familiarity estimates for the young and the older adults.

One should note that the findings are based on the assumption that old and young people presently encounter words with equal frequency. However, one might question how strong the assumption of equal word frequency for both populations is. Although we acknowledge that ideally, word frequency should also be based on age-specific corpora, just as our subjective norms, it is unlikely that such an analysis would eradicate the AoA effects we find. After all, the test of AoA effects is conservative in the difference score regressions and in the regressions per group. Most probably, word familiarity ratings incorporate both AoA and word frequency to some extent. In this sense, adding familiarity to the equations overcorrects the effect of AoA in the regressions.

In most of the analyses, significant effects were found for the familiarity predictor. However, interpreting these familiarity effects is not straightforward. Previous studies have shown that word familiarity ratings are not only influenced by the objective frequency of words but also by their meaningfulness (Balota et al., 2001). This makes it difficult to attribute these effects to an unique underlying mechanism that is often assumed for word frequency effects.

For word frequency, our findings deviate from the results reported in Morrison et al. (2003). They found no frequency effects in a word naming task with young and old adults while these effects were present in our lexical decision task. Our use of age-specific norms might explain this difference. To further illustrate the importance of age-specific norms, we ran an additional regression analysis on the lexical decision experiment (see Table 4, first analysis). The norms of AoA for the young were switched with those of the older adults and vice versa. In both regressions, the R^2 dropped with 0.11 for the young and 0.15 for the older adults and the AoA standardized betas were reduced from 0.52 to 0.36 for the young and from 0.63 to 0.58 for the older adults. These findings clearly show that age-specific norms should be used when AoA is investigated.

The results reported here provide further experimental evidence about the conditions under which the AoA-effect occurs. Recently, Brysbaert and Ghyselinck (in press) proposed an integrated account of AoA and word frequency by proposing that all AoA effects might be partly frequency-related. They compared frequency effects in different languages and multiple tasks and concluded that in most tasks such as word naming and lexical decision, frequency can account for about the same amount of explained variance as AoA, though the values of the latter variable are more restricted in range. Still, in other studies such as object naming and word associate generation, there is a frequency-independent AoA effect (Brysbaert & Ghyselinck, in press). A similar pattern was observed in our studies. Our lexical decision experiment showed consistent frequency effects for both young and older adults, whereas for the semantic categorization task these frequency effects were completely absent. Although these results fit well with previous findings of an entire array of studies with different stimuli in different languages, it can only be seen as weak evidence of a dissociation of AoA and word frequency since this evidence is based upon null results. The interpretation of the meta-analysis reported by Brysbaert and Ghyselinck concurs with the results from a number of simulation studies aimed at understanding the conditions under which AoA-effects show up. Central in these simulations is the finding that when there is extensive overlap between early and late acquired patterns, which is the case in mapping from orthography to phonology of English regular words, the AoA effect is strongly reduced or absent (see for example, Lambon Ralph & Ehsan, in press; Zevin & Seidenberg, 2002). However, in tasks that require semantic access, AoA-effects might occur because the relationship between orthography or phonology and meaning is much less systematic. More specifically, words that overlap in spelling tend to overlap in sound, but not in meaning (Zevin & Seidenberg, 2002). It would therefore be interesting to see whether the current results can be replicated in tasks such as reading that involve quasi-consistent mapping.

Although some of the effects of AoA in the separate group analyses of the lexical decision task might be related to word frequency, the finding of an AoA effect in the difference scores is critical, since there is no difference between the mappings for both groups. Furthermore, the involvement of semantics in the lexical decision task cannot be excluded, and reducing the locus of the AoA effects in this and other tasks to one source might be premature.

In conclusion, the lexical decision and semantic categorization task demonstrated robust AoA effects when word frequency was controlled by comparing two age-groups. This study is the first attempt to investigate AoA effects by comparing two groups that acquired words at different ages and allows a test where an array of confounding factors such as word frequency are removed. These results strongly indicate that AoA-effects in these tasks cannot be attributed to other word-specific factors beside the order in which words are acquired.

Acknowledgements

The work described in this paper was supported by Grant G.0266.02 from the Belgian National Science Foundation (Fundamental Human Sciences), and Grant OT/01/15 of the Leuven University Research Council to G. Storms. We thank the schools and the numerous volunteers for their participation in the norm gathering and experiments. We also thank Mandy Ghyselinck and Marc Brysbaert for helpful comments on an earlier version of this article.

Appendix

Age of acquisition (AoA) and familiarity (Fam) norms for Experiments 1 and 2: lexical decision (LD) and semantic categorization (SEM)

Word	Translation	LD/SEM	AoA		Fam		% Known	
			Young	Old	Young	Old	Young	Old
aids	AIDS	LD	11.61	37.47	3.11	3.19	100	100
drama	tragedy	LD	9.78	14.00	3.81	3.50	100	100
gaia	gaia	LD	11.63	40.68	2.62	2.83	100	96
hoest	cough	LD	4.95	5.89	4.08	3.79	100	100
holebi	lesbigay	LD	12.67	43.53	2.96	2.70	100	100
kreng	beast	LD	8.78	13.80	3.15	2.83	100	100
maand	month	LD	5.73	6.76	4.30	4.33	100	100
match	match	LD	8.11	9.81	4.00	3.90	100	100
pond	pound	LD	9.91	11.90	2.64	3.09	100	100
roes	fuddle	LD	11.00	15.81	2.81	3.14	100	100
salto	somersault	LD	7.55	12.14	2.73	3.00	100	100
samba	samba	LD	11.61	26.56	2.73	2.57	100	100
scampi	scampi	LD	9.25	22.80	3.69	3.50	100	100
schelp	shell	LD	5.24	6.94	3.15	3.42	96	100
smurf	smurf	LD	4.32	24.50	2.85	2.73	100	100
trend	trend	LD	11.09	19.28	3.27	3.46	100	100
abdij	abbey	LD-SEM	8.54	10.19	2.84	2.95	100	100
adder	viper	LD-SEM	8.00	11.29	2.68	2.77	100	100
airbag	airbag	LD-SEM	11.09	37.56	2.92	3.13	100	100
airco	air-condition	LD-SEM	10.89	31.41	3.65	3.71	100	100
ananas	pineapple	LD-SEM	5.91	9.27	3.37	3.55	100	100
ballet	ballet	LD-SEM	6.75	12.56	3.04	3.14	100	100
bamboe	bamboo	LD-SEM	8.46	13.14	2.76	2.95	100	100
barbie	barbie	LD-SEM	4.50	24.23	2.96	2.68	100	100
bink	tough guy	LD-SEM	10.05	21.35	2.58	2.36	100	92
byte	byte	LD-SEM	13.00	33.94	2.89	2.83	95	95
cactus	cactus	LD-SEM	6.11	7.94	2.92	3.09	100	100
cello	cello	LD-SEM	8.93	15.10	2.84	2.68	100	100
choco	chocolate spread	LD-SEM	3.59	7.27	4.27	4.04	100	100
clown	clown	LD-SEM	4.63	5.61	3.08	3.09	100	100
cowboy	cowboy	LD-SEM	5.70	8.25	2.85	3.13	100	100
darts	darts	LD-SEM	11.44	25.81	2.52	2.29	100	92
dynamo	dynamo	LD-SEM	8.35	11.60	3.26	3.23	100	100
egel	hegdehog	LD-SEM	5.52	6.57	3.12	2.92	100	100
elpee	LP	LD-SEM	10.96	15.40	2.67	2.90	100	100
engel	angel	LD-SEM	5.04	6.13	3.48	3.41	100	100
farao	pharaoh	LD-SEM	8.79	11.27	2.56	2.76	100	100
fluit	flute	LD-SEM	6.26	6.71	3.08	3.17	100	100

Appendix (*continued*)

Word	Translation	LD/SEM	AoA		Fam		% Known	
			Young	Old	Young	Old	Young	Old
gadget	gadget	LD-SEM	11.37	27.32	2.92	3.08	100	96
gamba	gamba	LD-SEM	12.14	20.35	2.76	2.96	92	96
geit	goat	LD-SEM	5.26	6.43	3.32	3.43	100	100
grot	cave	LD-SEM	5.54	7.05	3.16	3.27	100	100
hagel	hail	LD-SEM	5.77	7.09	3.12	3.36	100	100
harp	harp	LD-SEM	7.59	12.59	2.65	2.75	100	100
hert	deer	LD-SEM	5.22	7.91	3.08	3.13	100	100
hostie	host	LD-SEM	5.87	6.25	2.92	2.96	100	100
hyena	hyena	LD-SEM	8.82	11.59	2.56	2.79	100	100
jojo	yo-yo	LD-SEM	5.09	8.93	2.67	2.55	100	100
kabas	bag	LD-SEM	7.26	9.18	2.59	2.42	86	85
kalf	calf	LD-SEM	5.44	6.23	2.96	3.33	100	100
kano	canoe	LD-SEM	7.43	11.27	2.93	2.90	100	100
kanon	gun	LD-SEM	6.19	8.95	2.73	2.92	100	100
keet	hut	LD-SEM	10.86	15.47	2.23	2.32	96	95
kers	cherry	LD-SEM	5.05	6.05	3.23	3.48	100	100
ketel	kettle	LD-SEM	6.48	7.25	3.23	3.55	100	100
kikker	frog	LD-SEM	4.57	6.85	3.15	3.32	100	100
kilt	kilt	LD-SEM	11.00	14.47	2.41	2.45	100	100
klak	cap	LD-SEM	6.90	8.94	3.08	3.00	92	100
koek	cookie	LD-SEM	3.70	4.52	4.28	4.09	100	100
komeet	comet	LD-SEM	9.00	14.00	2.80	3.09	100	100
lama	lama	LD-SEM	7.67	10.83	2.44	2.68	96	100
lasso	lasso	LD-SEM	7.58	8.33	2.42	2.48	96	100
lego	Lego	LD-SEM	4.54	15.43	3.16	2.82	100	100
lens	lens	LD-SEM	9.13	13.05	3.58	3.52	100	100
lolly	lolly	LD-SEM	4.82	12.13	3.28	2.68	100	100
lont	fuse	LD-SEM	8.19	10.56	2.92	2.65	96	100
modem	modem	LD-SEM	13.52	41.13	3.56	3.55	100	100
motel	motel	LD-SEM	10.46	21.22	2.54	2.65	96	100
oester	oyster	LD-SEM	8.04	13.00	2.89	3.22	100	100
olijf	olive	LD-SEM	7.96	15.33	3.62	3.48	100	100
paasei	Easter egg	LD-SEM	3.74	5.77	3.00	3.17	100	100
paella	paella	LD-SEM	8.82	21.44	3.15	2.96	100	100
panda	panda	LD-SEM	5.91	17.05	2.88	2.87	100	100
panty	panty	LD-SEM	9.83	17.00	3.12	3.18	100	100
papaja	papaya	LD-SEM	12.00	26.29	2.52	2.42	91	96
patat	potato	LD-SEM	4.48	4.11	3.88	3.91	100	100
piama	pajamas	LD-SEM	4.04	5.43	4.40	4.18	100	100
pilaar	pillar	LD-SEM	9.48	10.59	2.74	2.96	100	100
prozac	prozac	LD-SEM	13.89	45.35	2.27	2.68	96	77
pull	pullover	LD-SEM	5.87	9.27	4.44	4.05	100	100

(*continued on next page*)

Appendix (*continued*)

Word	Translation	LD/SEM	AoA		Fam		% Known	
			Young	Old	Young	Old	Young	Old
puree	puree	LD-SEM	4.89	5.70	4.00	3.71	100	100
radar	radar	LD-SEM	10.82	15.57	2.72	2.95	100	100
rasp	grater	LD-SEM	7.75	11.25	3.08	3.27	100	100
rodeo	rodeo	LD-SEM	11.11	12.07	2.37	2.62	96	100
ruif	rack	LD-SEM	10.06	11.57	2.00	2.09	67	63
sauna	sauna	LD-SEM	9.78	19.95	3.31	3.29	100	100
slab	bib	LD-SEM	5.08	10.39	2.54	2.74	96	100
slee	sled	LD-SEM	4.79	6.60	2.89	3.19	100	100
slurf	trunk	LD-SEM	4.78	6.55	2.88	2.92	100	100
speer	spear	LD-SEM	6.89	8.96	2.64	2.75	100	100
spook	ghost	LD-SEM	4.67	5.67	2.96	2.91	100	100
steeg	alley	LD-SEM	8.22	11.17	3.04	3.08	100	100
stereo	stereo	LD-SEM	8.63	19.95	4.04	3.79	100	100
tandem	tandem	LD-SEM	7.91	11.50	2.73	2.83	100	100
tijger	tiger	LD-SEM	5.04	6.78	3.12	2.78	100	100
tulp	tulip	LD-SEM	5.89	6.59	2.96	3.29	100	100
urne	urn	LD-SEM	9.56	15.12	2.63	2.70	100	100
valk	falcon	LD-SEM	7.19	9.35	2.68	2.96	100	100
velg	rim	LD-SEM	10.54	12.56	2.92	2.96	96	100
viagra	viagra	LD-SEM	14.70	48.13	2.62	2.68	100	100
vinyl	vinyl	LD-SEM	11.52	17.59	2.54	2.83	100	100
wafel	waffle	LD-SEM	4.73	6.24	3.67	3.71	100	100
wesp	wasp	LD-SEM	4.93	7.35	3.32	3.46	100	100
wolf	wolf	LD-SEM	5.18	6.13	3.08	2.86	100	100
zalm	salmon	LD-SEM	6.91	10.44	3.58	3.75	100	100
zebra	zebra	LD-SEM	6.00	9.13	2.89	3.05	100	100
zombie	zombie	LD-SEM	9.00	22.68	2.96	2.83	100	96
zwaard	sword	LD-SEM	5.29	7.85	2.85	2.91	100	100
adelaar	eagle	SEM	8.73	11.09	2.62	2.92	100	100
artisjok	artichoke	SEM	10.33	19.83	2.68	2.63	100	100
avocado	avocado	SEM	10.09	17.24	2.81	2.92	100	100
boon	bean	SEM	5.04	6.87	3.44	3.68	100	100
braambes	blackberry	SEM	6.48	7.27	2.73	3.17	100	100
buggy	buggy	SEM	6.23	23.74	3.08	3.32	100	100
cipres	cypress	SEM	11.45	16.40	1.65	2.82	87	100
disco	disco	SEM	9.50	24.63	3.73	3.05	100	100
diskette	disk	SEM	11.59	36.36	3.88	3.88	100	96
dolfijn	dolphin	SEM	5.21	10.00	2.96	3.09	100	100
donder	thunder	SEM	5.17	7.14	3.32	3.70	100	100
eekhoorn	squirrel	SEM	5.54	6.56	3.12	3.05	100	100
fazant	pheasant	SEM	6.52	8.89	2.77	3.04	100	100
fitness	fitness	SEM	10.36	27.61	3.65	3.58	100	100

Appendix (*continued*)

Word	Translation	LD/SEM	AoA		Fam		% Known	
			Young	Old	Young	Old	Young	Old
floppy	floppy disk	SEM	12.00	35.76	2.62	3.13	100	95
fluo	fluorescent marker	SEM	8.79	25.71	3.76	2.86	100	96
fossiel	fossil	SEM	9.86	12.89	2.77	2.75	100	100
framboos	raspberry	SEM	6.32	8.76	3.44	3.23	100	100
frisbee	frisbee	SEM	6.39	31.50	3.19	2.62	100	95
geiser	geyser	SEM	10.60	14.95	2.35	2.76	91	96
gorilla	gorilla	SEM	5.52	8.83	2.88	2.75	100	100
goudvis	goldfish	SEM	5.45	6.67	3.38	3.25	100	100
grafitti	graffiti	SEM	9.57	26.94	3.23	2.91	100	100
havik	hawk	SEM	8.32	12.06	2.37	2.92	100	100
ijsbeer	polar bear	SEM	5.30	6.28	2.85	2.78	100	100
inktvis	cephalopod	SEM	6.30	10.73	2.93	3.09	100	100
kakkerlak	cockroach	SEM	7.39	14.50	2.92	2.62	100	100
kameleon	chameleon	SEM	7.61	11.80	2.73	2.96	100	100
karper	carp	SEM	10.09	12.38	2.50	3.00	100	100
kebab	kebab	SEM	12.70	35.44	3.77	2.35	100	100
koolmees	oxeye	SEM	8.36	11.25	2.38	3.05	100	100
krokodil	crocodile	SEM	5.44	6.33	2.92	2.87	100	100
larve	larva	SEM	8.26	9.63	2.65	3.14	100	100
loft	loft	SEM	12.12	32.90	2.36	2.58	96	92
luis	louse	SEM	5.81	8.83	2.72	3.13	100	100
mango	mango	SEM	8.82	27.00	3.30	2.92	100	100
neushoorn	rhinoceros	SEM	5.54	7.53	2.96	3.05	100	100
nijlpaard	hippopotamus	SEM	6.17	8.67	2.85	3.09	100	100
olifant	elephant	SEM	4.74	6.05	3.15	3.30	100	100
orchidee	orchid	SEM	10.27	16.43	2.96	3.48	100	100
orkaan	hurricane	SEM	8.00	12.00	2.81	3.27	100	100
papegaai	parrot	SEM	5.45	7.55	2.85	3.24	100	100
parkiet	parakeet	SEM	6.32	8.50	2.88	2.91	96	100
pesto	pesto	SEM	13.76	42.07	2.88	2.14	93	90
pinguin	pinguin	SEM	5.96	7.30	2.76	2.75	100	100
playmobil	playmobil	SEM	6.04	28.88	2.85	2.62	100	100
pluim	plume	SEM	5.36	5.31	3.04	3.57	100	100
poema	puma	SEM	8.64	11.60	2.96	2.67	100	100
ratelslang	rattlesnake	SEM	8.32	11.29	2.80	2.77	100	100
reiger	heron	SEM	8.11	8.86	2.96	3.36	100	100
rog	ray	SEM	11.43	10.00	2.46	3.13	91	100
rotonde	roundabout	SEM	10.22	26.45	3.58	4.00	100	100
rups	caterpillar	SEM	5.37	7.82	2.74	3.17	100	100
salamander	salamander	SEM	6.43	9.47	2.96	3.00	100	100
schol	plaice	SEM	9.62	11.95	2.38	2.88	95	100
schorpioen	scorpion	SEM	8.18	11.94	2.89	2.96	100	100

(*continued on next page*)

Appendix (*continued*)

Word	Translation	LD/SEM	AoA		Fam		% Known	
			Young	Old	Young	Old	Young	Old
snoek	pike	SEM	7.85	10.09	2.50	2.96	100	100
snooker	snooker	SEM	11.79	29.30	3.15	2.82	100	100
spar	spruce	SEM	7.26	7.94	3.15	3.82	100	100
spreeuw	starling	SEM	8.17	9.71	2.76	3.30	100	100
sprinkhaan	grasshopper	SEM	6.39	7.50	3.04	2.91	100	100
sweater	sweater	SEM	7.36	32.14	3.12	3.44	100	100
tonijn	tuna	SEM	8.22	12.28	3.62	3.17	100	100
triatlon	triathlon	SEM	10.50	31.27	2.96	3.10	100	100
veulen	foal	SEM	5.36	7.85	3.04	3.41	100	100
vijg	fig	SEM	8.50	11.13	2.85	3.33	100	100
vink	finch	SEM	7.48	8.13	2.56	3.32	100	100
vlinder	butterfly	SEM	5.00	5.65	3.26	3.63	100	100

References

Baayen, R. H., Piepenbrock, R., van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM)*. Linguistic Data Consortium. University of Pennsylvania, Philadelphia, PA.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283–316.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*, 639–647.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology, 50A*, 560–585.

Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language, 50*, 456–476.

Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15*, 208–216.

Brysbaert, M., Ghyselinck, M. (in press). The effect of age-of-acquisition: Partly frequency-related partly frequency-independent. *Visual Cognition*.

Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology, 12*(1), 65–85.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition of words is a significant variable in semantic tasks. *Acta Psychologica, 104*, 215–226.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25*, 85–95.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249–253.

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflects loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*, 1103–1123.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777–799.

Fiebach, C. J., Friederici, A. D., Müller, K., Cramon, D. Y., & Hernandez, A. E. (2003). Distinct brain representations for early and late learned words. *Neuroimage, 19*, 1627–1637.

Gerhand, S., & Barry, C. (1999). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition, 27*, 592–602.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*, 256–281.

Ghyselinck, M. (2002). *Age of acquisition as an organizing variable of the semantic system*. Unpublished PhD Thesis, Ghent University.

Ghyselinck, M., Custers, R., & Brysbaert, M. (2003). Age-of-acquisition ratings for 2322 Dutch words from 49 different semantic categories. *Psychologica Belgica, 43*, 181–214.

Ghyselinck, M., De Moor, W., & Brysbaert, M. (2000). Age-of-acquisition ratings for 2816 Dutch four- and five-letter words. *Psychologica Belgica, 40*, 77–98.

Gilhooly, K. J., & Logie, R. H. (1981). Word age-of-acquisition, reading latencies and auditory recognition. *Current Psychological Research, 1*, 251–262.

Hodgson, C. (1999). *Object naming in the elderly: Some experiments and a model.* Unpublished D.Phil. thesis, University of York, UK.

Humphreys, L. G. (1978). Research on individual differences requires correlational analysis, not ANOVA. *Intelligence, 2*, 1–5.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of pictures naming. *Psychological Bulletin, 120*, 113–139.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lambon Ralph, M. A., & Ehsan, S. (in press). Age of acquisition effects depend on the mapping between representations and the frequency of occurrence: Empirical and computational evidence. *Visual Cognition.*

Lewis, M. B. (1999). Age of acquisition in face categorisation: Is there an instance-based account? *Cognition, 71*, 23–39.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 149–157.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers, 28*, 203–208.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*, 181–190.

MetaMetrics, Inc. (2003). *MetaMetrics word frequency counts* [Database]. Durham, NC: Author. (Available from MetaMetrics, Inc., Developers of the Lexile Framework, attention: A. Jackson Stenner, 2327 Englert Drive, Suite 300, Durham, NC 27713).

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 116–133.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology, 91*, 167–180.

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition, 20*, 705–714.

Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology, 14*, 435–459.

Morrison, C. M., Hirsh, K. W., & Duggan, G. B. (2003). Age of acquisition, ageing, and verb production: Normative and experimental data. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*, 705–730.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.

Rosenthal, R. (1995). Critiquing Pygmalion: A 25-year perspective. *Current Directions in Psychological Science, 4*, 171–172.

Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language, 42*, 51–73.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47*, 1–29.

Zevin, J. D., & Seidenberg, M. S. (2004). Age of acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition, 32*, 31–38.