Abstract

When developing theories of how the brain processes words, it helps to create categories of words and see how their processing differs. Two natural categories are Age of Acquisition and frequency. In this experiment, based on work by Gerhand and Barry[1], we tested response times in a Lexical Decision Task and their relationship between both AoA (Age of Acquisition) and frequency. We were able to successfully replicate the findings of that previous work; however, modern analysis of the categorization of words by their frequency and AoA necessitates discussion of these results uses.

Background and Literature Review

Before looking at Age of Acquisition and frequency's effects on lexical analysis, let us look at general theories around lexical comprehension. The most common theory, and the one that many more specific theories are based on, is the Dual Route Model. In this model, the brain has 2 separate pathways to understand a word. The first is the lexical, or direct, route. When using the lexical route, the brain directly associates the shapes and combinations of letters with a specific meaning. This works in a similar manner to other forms of object detection; in the same way one can look at an apple and know what it is, so too can they look at the word "apple". The other route is know as the graphonological, or indirect, route. In this route, the reader sounds out the word (either silently or aloud), and then connects the meaning to that sound.

While this serves as a starting ground for understanding how humans process words, there are a number of theories that build on this base model. For instance, Grainger and Ziegler [2] propose a pair of coarse and fine "chunking" methods to break down words into letter pairs and small sections. These components can then be analyzed graphically, be sounded out, or a mixture of the two through the more traditional dual route model.

The effect of frequency and AoA (Age of Acquisition) of words on their comprehension is an active area of research, with a range of competing theories and claims about the extent of their affects. Some researchers (see Zevin & Seidenberg, 2002[3]) dispute that AoA effects are anything more than side effects of other factors. One of the reasons that some believe that AoA does not have an effect on experiments such as Lexical Decision Tasks is that AoA is thought to be an indirect measurement of word frequency. Words that are more commonly used are more likely to be learnt earlier in life, and they argue that when accounting for this correlation AoA effects are not seen.

This argument, however, has been heavily called into question. In experiments which preemptively account for both AoA and frequency, AoA effects are still seen[4], and in fact in many cases seen more strongly than frequency effects.

One of the other reasons that AoA is seen as a secondary factor in lexical processing is that it's effect differs drastically between tasks. In general, picture naming tasks tend to have the highest AoA effects, whereas word naming tasks see little to no effect.[4] Lexical decision tasks, like the one performed here, seem to have an AoA effect, but not as strongly as in picture naming tasks.

In a review by Barbara J. Juhasz[4], she identified 6 distinct theories explaining the effect of AoA on lexical processing. Some of these, like the Bilateral Representation Hypothesis, have been discounted in recent years, and others, like the Phonological Completeness and Lexical-Semantic Competition Hypotheses, focus on explaining picture naming tasks rather than Lexical Decision Tasks. The theories that best explain the effects seen in Lexical Decision Tasks are the Semantic Locus Hypothesis and the Network Plasticity Hypothesis.

The Semantic Locus Hypothesis is based on previous research and theories on how humans learn the meaning of words throughout their life. When young, a speaker learns to directly connect words to their meaning, as they have a limited vocabulary to use to describe that word. For instance, one may learn the word "ball" and directly associate it with the concept of a ball. However, words learnt later in life will be remembered through their connections to already known words (that same speaker may learn the word "sphere", and connect it back to ball rather than the base concept of a sphere). Recalling words learnt later in life, then, is slower, as understanding the semantics of a word is more difficult when navigating through connections to other words instead of directly to a concept. This theory suggests that tasks requiring the semantics of a word are more likely to show an AoA effect. This does hold true in research, as picture naming tasks require more semantic understanding than Lexical Decision Tasks, and as such show a strong AoA effect. This also implies that a Lexical Decision Task requires semantic comprehension, and this implication is supported by Perea and Rosa's[5] research showing that semantic priming has a significant effect on Lexical Decision Tasks.

The Network Plasticity Hypothesis is similar to the Semantic Locus Hypothesis, in that they are both based on changes to how the brain learns over time. This theory suggests that when words are learnt at an early age, the brain can create strong, new, and unique patterns to represent them. When words are learnt at a later age, however, the new patterns are weaker, and often rely on existing patterns of other words. While this is similar to the Semantic Locus Hypothesis, in the Network Plasticity Hypothesis the difficulty of accessing a word is detached from the semantic understanding of those words. Additionally, this hypothesis is largely based on simulated studies of artificial neural networks, rather than human trials.

Due to the related nature of AoA and frequency, some of these modern theories are moving towards the idea that frequency may be a side effect of AoA, in contrast with older theories that claim that AoA effects are a byproduct of frequency. There are also theories that both correlated and independent effects are present, but this continues to be an area of active research. Systematically evaluating these differences has proved difficult, due to the ambiguity in the scale of frequency and AoA. There is no easy way to map "early" vs "late" Age of Acquisition to "low" and "high" frequency words. This makes it difficult to effectively quantify AoA and frequency independent of one another. When researching AoA effects, researchers did propose a statistical technique using the z score between the two categories of both.[8] This method, however, is imperfect, and minimal research has used it due to its recent publication.

Research Questions

With this understanding of the theories surrounding AoA and frequency's effect on lexical tasks, and the background of previous studies, let us move on to the experiment discussed in this paper. This experiment aims to add to the corpus of data generated through previous experiments, and to corroborate their conclusions.

Specifically, this experiment aims to replicate and verify the results of Gerhand & Barry[1], by analyzing the correlations between the Age of Acquisition (AoA) and frequency of words in English and the time taken to identify them as existing words in a Lexical Decision Task.

As per that original study, we expected that frequency would be the primary contributor to response times, and that AoA would play a smaller role, appearing primarily as a determiner in low frequency words.

Experiment

To study the effects of word frequency and average Age of Acquisition (AoA) on lexical access times, we designed and published a website for participants to self test. These tests were taken over the course of a month, and were designed to largely replicate the tests performed in Gerhand & Barry.[1]

Participants

In total, over 50 people participated in the trials, but only 43 completed the tests. Any response over 10 seconds was discounted as the result of a distraction. Of those who completed the test, only 1 result was discounted for excess time. The most common L1 (first language) of participants was English (40/43 participants), with 1 L1 speaker each of: Chinese (Mandarin), Spanish, and Telugu. All participants were, however, fluent in English.

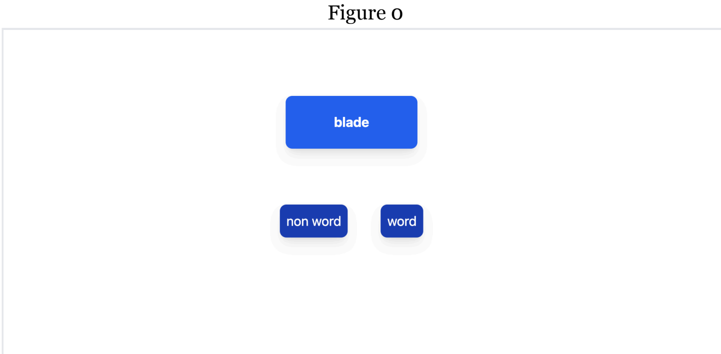
Participants were mostly college-age students, with a median age of 20. There were, however, four participants over the age of 30, and one participant under 18. Participants were both left and right handed, and were not asked their handedness before beginning. Participants were not asked about their vision. All participation in the trials was volunteer, and no incentives were given.

Methods

Before beginning the testing, all participants were asked for their name, age, and first language. Names were recorded to avoid to avoid duplicate test takers, and are not included in any publicly available data.

After entering their information, participants were prompted with a screen explaining how to take the test. They were told to rest their fingers on the "f" and "j" keys of the keyboard (the left and right centering keys on the standard QWERTY keyboard). They were told that they would see a series of words on the screen, and to press the "f" or "j" key if a word was a real word in English, and the other if it was not. Whether the "word" option was on "f" or "j" was randomized between participants. Word vs. non-word cues were included throughout testing above the keys on the screen, as shown in figure 0.

Figure 0



Participants were then shown a series of potential words and asked to identify whether or not they were true words in English. In total, participants were shown 64 potential words, in addition to 10 "warm up" potential words. These first 10 were the same for all participants, but the 64 test words were picked randomly from the potential word pairs, which will be discussed further in Stimuli. Each participant saw 16 words from each of the following categories, totaling 64 tests per person:

- High Frequency, Early Age of Acquisition
- High Frequency, Early Age of Acquisition
- Low Frequency, Late Age of Acquisition
- Low Frequency, Late Age of Acquisition

Of these, 8 of each were true words, and 8 were non-words designed to emulate the phonotactics of their corresponding real word. All of these test words were randomized per participant. After pressing either "f" or "j", the time taken to respond was recorded, along with which word it was and whether or not they chose correctly.

Stimuli (Word Choice)

Each potential word shown to participants was part of a word non-word pair. The nonword in the pair was created by modifying one or more letters of the true word, while still obeying English orthographic rules and phonotactics. In total there were 100 test pairs.

Of these 100 test pairs, 64 were from the dataset used in Gerhand & Barry[1], and the other 36 were written by Eleanor Olson. Of those not from the Gerhand & Barry, words were segregated into frequencies and average AoA with analysis by Kuperman, Stadthagen-Gonzales, and Brysbaert.[6] Words with an average AoA over 10 (years old) were said to be learnt late, and those with an AoA under 5 were said to be early. Words with a frequency over 150 occurrences per million words were said to be high frequency, and those with less than 2 occurrences per million were classified as low frequency. In total, there were 22 high frequency late AoA, 26 low frequency late AoA, 26 high frequency early AoA, and 26 low frequency early AoA word pairs. In this dataset, there are fewer high frequency and late AoA pairs, as words that are frequent are more likely to be learnt at a younger age. Since, however, each participant was always shown the same number of each category, this imbalance does not skew results.

Results

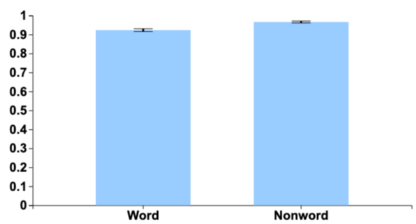
Word Identification Accuracy

Overall, participants were highly accurate at distinguishing word from non words. This aligns with previous studies on LDTs. Participants were slightly more likely to misidentify a word as a non word, with a mean word identification accuracy of 92.3% and non word identification accuracy of 96.6%, with respective standard errors of 0.7% and 0.5%, as seen in figure 1.

The relationship between word status (whether or not a word was a real English word) and accuracy was analyzed with a logistic regression model. There was a significant effect of word status on response accuracy ($z = 4.747$, $p < 0.001$).

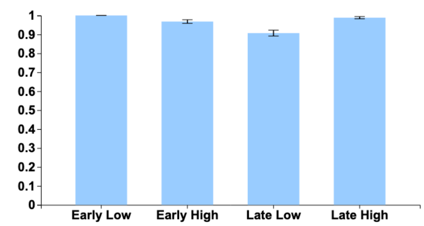
When broken down by frequency and AoA, as in figure 2, we see that there are larger differences in accuracy. When measuring response accuracy based on frequency and AoA, corresponding non words were ignored, and only the true words were used. The highest response accuracy occurred in early AoA low frequency words, with 100% accuracy, and the lowest accuracy in late AoA low frequency words, with a 90.7% accuracy and a 1.6% standard error.

Figure 1



The average accuracy of users at identifying words and nonwords.

Figure 2



The average accuracy of users at identifying words depending on their AoA and Frequency. Non words are ignored.

The relationship between AoA + frequency and response accuracy was also analyzed with a logistic regression model. This model showed that there was no significant effect of either AoA or frequency on accuracy, nor was there a significant effect of them combined (although their interaction trended towards significance). These relationships can be seen in figure 3.

Figure 3

	Z Value	P Value
AoA (late)	-0.249	0.803
Frequency (low)	0.655	0.513
AoA + Frequency	-1.841	0.0657

Word Identification Response Times

In addition to being slightly more accurate at identifying nonwords, participants also tended to take more time. As shown in figure 4, the mean identification time for words was 836ms, with a standard deviation of 478ms, and the mean time to identify nonwords was 1146ms, with a standard deviation of 796ms. These response times varied significantly, which was expected due to the natural variations in human reaction times. More information can be seen by hovering over figure 4.

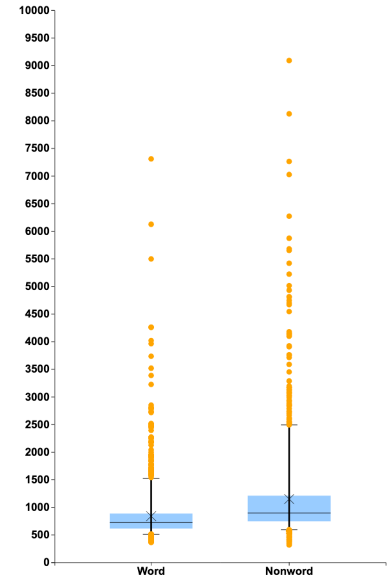
The relationship between word status and response time was analyzed with a linear regression model, and found a significant effect ($z = -12.74$, $p < 0.001$) of word status on response time.

When categorized by frequency and AoA, as in figure 5, we again see a pairing of higher response accuracy with longer average response times. The highest mean identification time occurred in infrequent late AoA words, with a mean of 981ms, and the lowest mean identification time was in frequent late AoA words, with a mean of 764ms.

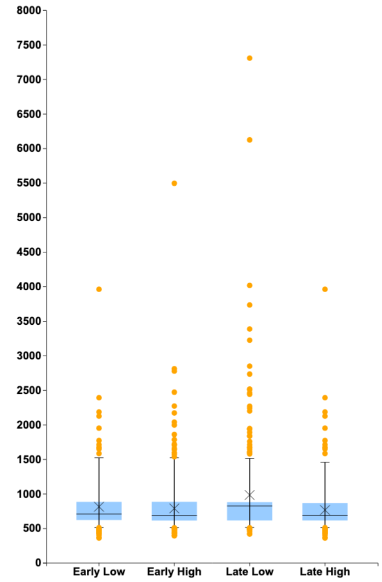
Figure 4

Figure 5

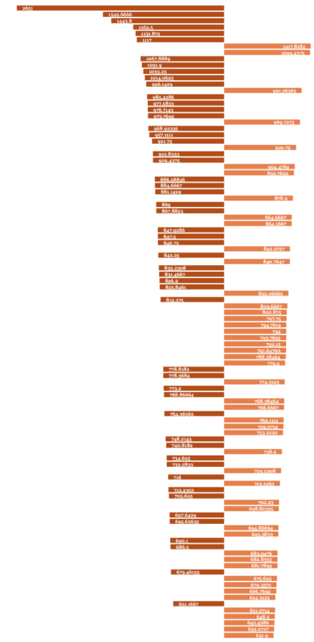
Figure 6



Blue box surrounds the center 2 quartiles. Whiskers surround the 5th and 9th percentiles. All data the falls outside of this is plotted



Blue box surrounds the center 2 quartiles. Whiskers surround the 5th and 9th percentiles. All data the falls outside of this is plotted



On the left are words with a low frequency in english, on the right are those with a higher frequency.

The relationship between AoA, frequency, and response time was also analyzed with a linear regression model. It did not find a significant effect between either AoA or frequency on their own and response time (AoA: $z = -0.724$, 0.469, frequency: $z = -0.152$, $p = 0.8791$); however, when combined they did have a statistically significant effect on response time ($z = 2.493$, $p = 0.0127$, $F(1,2684) = 6.213$). There was a higher effect of AoA on response time for low frequency words than higher frequency words.

While not part of the originally planned analysis, one other result found was the correlation between response accuracy and response time. When analyzed with a linear model, a significant effect was found ($z = -3.92$, $p < 0.01$). However, no significant effect was found when correlating accuracy with any other metrics.

Figure 6 shows a complete list of the tested words and their mean response times. In addition to the statistical analysis, this histogram visually shows the relationship between frequency and response time.

Discussion

These results matched the predictions outlined by Gerhand & Barry.[1] While we were not able to significantly show that frequency was a predictor of response time on its own, we did show that it was a predictor when accounting to AoA as well. Additionally, we showed that there was a more prominent effect from AoA in low frequency words.

This analysis was based on the analysis done by Gerhand & Barry, and as such it is unsurprising that it aligns with their results. If word selection for AoA and frequency was guided by more modern techniques as outlined in the various theories of AoA's effect on lexical tasks, we may have seen results that aligned with those theories.

While our analysis does align with the results in previous studies, there are a number of limitations in our data that should be considered. All participants took the tests on their own computers at their own discretion, and as such many outside variables could not be avoided. While only 1 test was discarded due to extended delay, it is not possible to tell whether there were other stimuli during testing that affected results. Additionally, participants were primarily from a similar cohort, meaning that these results may not hold true in the wider population. This is a particularly important note, as some studies[7][4] have found frequency effects present in young adult populations, but not in older adults. Since, however, our results matched those found by numerous studies, they add to the body of tests suggesting the findings presented, and help to solidify the conclusions drawn by others.

Although there is a plethora of data analyzing a number of lexical processing tasks and their relationship to both frequency and Age of Acquisition, there remain a number of competing theories on the reasoning behind their effects. Additionally, some continue to argue that AoA and frequency are simply manifestations of the same property. While more research on new data is useful, we now have a number of more modern techniques to partition words by characteristics like frequency and AoA. With these techniques, we may find more information by reevaluating existing data from the numerous studies over the year, and comparing their findings with consistent categorization and analysis.

References

1: Gerhand, S., Barry, C. Age of acquisition, word frequency, and the role of phonology in the lexical decision task. Memory & Cognition 27, 592–602 (1999). https://doi.org/10.3758/BF03211553

2: Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. Frontiers in Psychology, 2, Article 54. https://doi.org/10.3389/fpsyg.2011.00054

3: Jason D Zevin, Mark S Seidenberg, Age of Acquisition Effects in Word Reading and Other Tasks, Journal of Memory and Language, Volume 47, Issue 1, 2002, Pages 1-29, ISSN 0749-596X, https://doi.org/10.1006/jmla.2001.2834.

4: Juhasz, B. J. (2005). Age-of-Acquisition Effects in Word and Picture Identification. Psychological Bulletin, 131(5), 684–712. https://doi.org/10.1037/0033-2909.131.5.684

5: Perea, M., Rosa, E. The effects of associative and semantic priming in the lexical decision task. Psychological Research 66, 180–194 (2002). https://doi.org/10.1007/s00426-002-0086-5

6: Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. Age-of-acquisition ratings for 30,000 English words. Behav Res Methods. 2012 Dec;44(4):978-90. doi: 10.3758/s13428-012-0210-4. Erratum in: Behav Res Methods. 2013 Sep;45(3):900. PMID: 22581493.

7: 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(4), 435–459. https://doi.org/10.1080/09541440143000159

8: Lewis, M.B., Chadwick, A.J. & Ellis, H.D. Exploring a neural-network account of age-of-acquisition effects using repetition priming of faces. Mem Cogn 30, 1228–1237 (2002). https://doi.org/10.3758/BF03213405