# FEATURAL DENSITY IN PICTURE NAMING AMONG COLLEGE AGE AND OLDER ADULTS

A Master's Thesis
Presented to
The Academic Faculty

by

Robert T. Ashwill

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
August 2020

# FEATURAL DENSITY IN PICTURE NAMING AMONG COLLEGE AGE AND OLDER ADULTS

Approved by:

Dr. Daniel H. Spieler, Advisor
School of Psychology
*Georgia Institute of Technology*

Dr. Christopher Hertzog
School of Psychology
*Georgia Institute of Technology*

Dr. Thackery Brown
School of Psychology
*Georgia Institute of Technology*

Date Approved: February 14, 2020

## ACKNOWLEDGEMENTS

I would like to thank Don and Cheryl Ashwill for their unconditional support which has allowed me to achieve what I have. Additionally, I thank Emily Madden who, on multiple occasions, provided the support and push necessary to reach this milestone. I thank Dr. Audrey Duarte, Taylor James, Dr. Chris Hertzog, and Emily Lustig who’s assistance with recruitment was crucial in completing this project. I would also like to thank Dr. Daniel H. Spieler who’s mentoring has led directly to completion of this document.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# SUMMARY

When naming pictures, speakers are slower to name pictures with multiple appropriate labels (e.g., couch/sofa) compared to pictures consistently given a single label. This increased naming time is generally seen as a reflection of the time needed to resolve competition between the competing labels. Older speakers show a greater influence of name agreement that could reflect a specific age-related increase in sensitivity to lexical competition when speaking. The present study examines speakers' sensitivity to a more pervasive form of lexical competition. Using normative data in which individuals report features associated with object concepts, it is possible to measure the extent to which concepts share features with other concepts. Pictures matched with concepts with high featural overlap with other concepts should show greater competition during naming than those matched to concepts with lower levels of featural overlap. Initial evidence in younger speakers is consistent with this prediction. Here, we conducted a set of experiments to replicate this result in younger speakers and test the prediction that older speakers will be more sensitive to variations in featural overlap than younger speakers. We observed a marginal negative relationship between featural overlap and response times if participants were not pre-exposed to stimuli. With pre-exposure we saw a significant negative effect of feature overlap and response times in both young and older adults. There was no clear differential effect of featural overlap on semantic competition for young and older adults.

# CHAPTER 1. INTRODUCTION

Language through adulthood is often thought of as static during normal aging. There are often minimal changes to aspects of language processing observed in non-pathologic aging (e.g. Burke & Mackay, 1997; Schaie, 2013) compared to changes observed in other cognitive processes such as memory (e.g. Ronnlund & Nilsson, 2008; Salthouse, 2016; Schaie, 2013; Schroeder & Salthouse, 2004) or speed of processing (e.g. Salthouse, 2016; Schaie, 2013). There is even some evidence older adults perform better than their younger counterparts such as increased scores on tests of vocabulary (e.g.Salthouse, 2016; Schroeder & Salthouse, 2004; Verhaeghen, 2003). However, as in many other domains, older adults show differences from younger adults on tasks related to language processing, such as increased picture naming times (Au et al., 1995; Nicholas, Barth, Obler, Au, & Albert, 1997; Thomas, Fozard, & Waugh, 1977), increased picture naming error rates (Au et al., 1995; Nicholas et al., 1997), increased interference during picture naming from highly related concepts (Taylor & Burke, 2002), and increased probability of entering a tip of the tongue state (Burke, MacKay, Worthley, & Wade, 1991).

Picture naming is commonly used in language processing research, particularly in building models of word production (e.g. Dell, 1986; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Levelt, Roelofs, & Meyer, 1999; Oppenheim, Dell, & Schwartz, 2010; Schriefers, Meyer, & Levelt, 1990). Studies using picture naming take advantage of the variation in the labels speakers provide to investigate aspects of word production, such as word frequency (Alario et al., 2004; Barry, Morrison, & Ellis, 1997; Jescheniak & Levelt, 1994; Snodgrass & Yuditsky, 1996), number of syllables (Ferrand, Segui, & Grainger, 1996; Santiago, MacKay, Palma, & Rho, 2000), or name agreement (e.g. Alario et al., 2004; Barry et al., 1997, Lagrone & Spieler, 2006).

The work discussed here is specifically concerned with the influence of age on the process of selecting and producing a specific word label for a pictured object. Thus, before turning to the specific aspects of the experiments, I will briefly review the consensus view of the processes involved in word production.

## 1.1 Word Selection in Picture Naming

In a language production context, our interest in picture naming typically starts after an individual's recognition of a picture when the process of selecting and producing an object label begins. Several levels of processing and representation intervene between the activation of semantic features during object recognition and the articulation of an object label (e.g. Abdel Rahman & Melinger, 2009; Belke & Stielow, 2013; Berg & Levelt, 1990; Dell, 1986; Howard et al., 2006; Levelt et al., 1999; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Oppenheim et al., 2010; Roelofs, 1992). At the first level, speakers have some representation of what they wish to convey, referred to as the 'message' (Levelt et al., 1999). This representation is made up of semantic and conceptual features that capture what a speaker wishes to communicate. The semantic and conceptual content of the message is not directly tied to the speaker's language and accessing this information does not appear to involve processes for language production. For example, individuals can judge the typical size of pictured objects and the judgments are not influenced by properties of the object label such as the frequency of occurrence (Kroll & Potter, 1984). In the context of a production task such as picture naming, the semantic and conceptual information about the pictured object must be mapped to a lexical representation corresponding to the label the speaker will apply to the object. At this second level of representation, the lexical (or *lemma*) representation acts as a point of convergence for the collection of semantic features accessed during picture recognition. The task at this level is to select the lemma most consistent with the activated semantic features. Finally, the phonological form for the selected lemma is assembled and readied for articulation.

During lemma selection, multiple lemmas may be partially consistent with the activated set of semantic features specified as part of the message. For example, if an individual is told to name a picture of a *bed*, recognition of the picture should involve activation of semantic and conceptual representations such as *"used to sleep in", "found in bedrooms", "comfortable",* etc.. The active semantic/conceptual representations feed activation to various connected lemmas such as *bed*, *couch*, or *night stand*[1]. The word *bed* will be highly active because all the active semantic features feed activation to it while *couch* and *night stand* will be less active because they share some but not all the activated features. Selection should favor *bed* as it is the most highly active, and best fitting, lemma representation.

If semantic features sufficiently activate multiple lemmas, then the process of lemma selection is slowed by the need to select one amongst a competing set of activated lemmas (Dell, 1986; Levelt et al., 1999; Roelofs, 1992). In the example of an individual naming a picture for *bed,* the system will have multiple other active lemma representations such as

[1] In linguistics and psycholinguistics a word is considered to be the smallest unit maintaining meaning. Hence, 'ax handle' may be defined as a single word because deleting either 'ax' or 'handle' substantially changes the meaning.

*couch* or *night stand*. In this case the system must select a lemma representation from the active set of representations taking more time than if activation had converged only on *bed*.

### *1.1.1 Competition for Selection*

Given the relatively weak context in which picture naming occurs, some pictures will have multiple appropriate labels. Across speakers, such pictures will have lower agreement among responses compared to other pictures. This response agreement can be quantified as the proportion of speakers who give the modal response for a given picture, referred to as *name agreement*. The typical finding is that pictures with high name agreement are named faster than those with lower name agreement (Alario et al., 2004; Barry et al., 1997; Bonin, Chalard, Méot, & Fayol, 2002), even when controlling for other variables such as the visual complexity of the picture, word frequency, and length of the picture label (Alario et al., 2004; Barry et al., 1997; Bonin et al., 2002). It is worth noting that while low name agreement could occur for psycholinguistically uninteresting reasons such as visual ambiguity, the stimuli typically used in these studies include items with the proportion of modal responses ranging from over 80% for high name agreement pictures to 40 to 70% for medium name agreement pictures.  For properly selected items, the variation in name agreement generally reflects the availability of multiple appropriate labels (e.g., *couch*/*sofa*/*love seat*). Thus, the relationship between name agreement and naming time appears to reflect variation in competition amongst lemmas for selection and the time needed to resolve this competition.

While older adults are expected to show increased picture naming times relative to younger adults (Au et al., 1995; Nicholas et al., 1997; Thomas et al., 1977) older speakers also show a greater influence of name agreement on picture naming times (Britt, Ferrara, & Mirman, 2016; LaGrone & Spieler, 2006). This result suggests that older speakers are more sensitive to lemma competition and take longer to resolve this competition compared to younger speakers.

## 1.2 Naming and Inter-correlational Density

While naming pictures with medium name agreement represents a situation with near maximal overlap of semantic to lemma mapping, activated semantic features should commonly result in the activation of more than one lemma.  For example, naming a picture of a *cat* is unlikely to elicit alternative responses and thus would have high name agreement but the activated features are likely shared amongst other lemmas, (e.g. *dog, rabbit, mouse,* etc.). The multiple active lemma should result in competition during lexical selection similar to that observed when naming medium name agreement items.

To measure the connections a concept has to other concepts one can estimate its semantic density. Semantic density is based on the number and overlap of a concept's semantic connections to other similar concepts. One way to estimate the semantic density of a concept is to consider the features of a concept to be key parts of its semantic meaning. Using features produced for a concept we can calculate the featural overlap, and implicitly semantic overlap, a concept shares with other concepts. Inter-correlational density (McRae, Cree, Seidenberg, & McNorgan, 2005) is an example of one such semantic overlap metric. McRae et al. (2005) collected feature norms for a set of 541 English object word concepts by presenting individuals with the different concept words and asking them to write properties describing each word concept. Using those produced features as the pieces of a concept vector McRae et al. (2005) calculated inter-correlational density from pairwise correlations between each concept word[2]. Inter-correlational density is a measure of a concepts featural overlap with other concepts and a way to estimate semantic overlap (McRae et al., 2005; Mirman & Magnuson, 2008). Concept words with high inter-correlational density have a large amount of featural overlap, thus when producing names for those concepts semantic connections to multiple other lemmas would become active. While inter-correlational density is specifically calculated from concept words we can extend its implications to pictures representing those concepts.

Individuals are slower to name pictures of words with higher semantic overlap than those with lower semantic overlap (Mirman & Magnuson, 2008; Rabovsky et al., 2016). When speakers named pictures matched to concepts from McRae et al. (2005), naming times showed a positive linear relationship with inter-correlational density (Rabovsky et al., 2016). That is, the more shared features a concept has with other concepts, the slower speakers are to select and produce the corresponding picture label, depicted in Figure 1. This mirrors the finding that naming pictures with medium name agreement is slower than naming high name agreement pictures (Alario et al., 2004; Barry et al., 1997; Britt et al., 2016; LaGrone & Spieler, 2006).

[2] Inter-correlational density is calculated by creating a matrix of concept rows and feature columns with cell contents being the produced frequency for each feature on each concept. Features which were produced on fewer than three concepts are excluded. Using concept row vectors, individual pairwise correlations are conducted for each concept. Then the shared variance (for correlations accounting for 6.5% or more of the variance) of a concept’s correlations are summed for each concept.

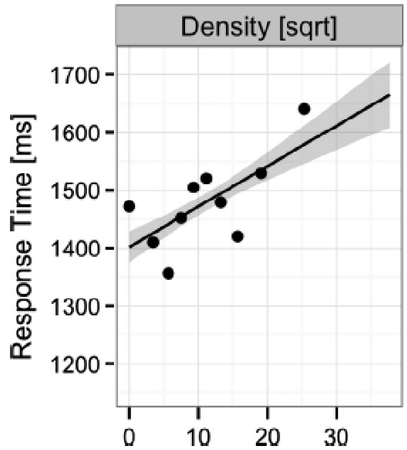


**Figure 1 – Response time by inter-correlational density**

**Response time modeled as a function of square-root inter-correlational density with 95% confidence intervals. Dots represent averages for 10 quantiles. Reproduced from (Rabovsky, Schad, & Abdel Rahman, 2016)**

If inter-correlational density influences the amount of competition during lemma selection, and older adults are more sensitive to competition, then we would predict that older speakers will show a greater influence of inter-correlational density (Britt et al., 2016; LaGrone & Spieler, 2006). Alternatively, if inter-correlational density estimates a separate type of competition from that observed when naming medium name agreement pictures, we might not observe age differences across the range of inter-correlational density.

The following experiments examine the relationship between inter-correlational density and naming times for pictures. We conducted two experiments using the same stimuli and similar methods as previous work (Rabovsky et al., 2016). Our first experiment was an attempt to replicate the observations of Rabovsky et al. (2016). In our second experiment we aimed to address the question of whether older adults are more impacted by the semantic competition predicted by inter-correlational density in a picture naming paradigm.

# CHAPTER 2. EXPERIMENT 1

Prior to recruitment of an older adult group it was important to confirm similar observations to those in Rabovsky et al., (2016). An initial experiment was conducted with college age individuals at the Georgia Institute of Technology.

## 2.1 Methods

### *2.1.1 Participants*

Thirty-five undergraduate students were recruited from the Georgia Institute of Technology School of Psychology subject pool and were awarded course credit for their participation. All subjects reported corrected to normal vision. Fourteen participants in total had to be excluded for various reasons: six for indicating English was not their first language, five were excluded for having too many experimental equipment errors, one was excluded because changes were made to the experimental protocol after they participated, one was excluded because they made excessive extraneous noise such as tapping on the table or fidgeting with experimental equipment meaning accurate voice onset times could not be determined, and one was excluded because their audio file was corrupted making it impossible to code their responses. After excluding these individuals, a total of twenty-one participants were included in the analysis. Participant characteristics can be seen in Table 1a.

**Table 1 – Participant characteristics**

| | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| Table 1a. Younger adult participant characteristics, no familiarization | | | | | | | |
| | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
| Age | 21 | 19.476 | 1.632 | 18 | 18 | 20 | 24 |
| Digit Symbol | 21 | 47.476 | 6.29 | 38 | 43 | 53 | 61 |
| Vocabulary | 21 | 49.952 | 6.553 | 36 | 47 | 54 | 59 |
| Table 1b. Younger adult participant characteristics, familiarized | | | | | | | |
| | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
| Age | 26 | 19.731 | 2.475 | 18 | 19 | 20 | 31 |
| Digit Symbol | 26 | 52.346 | 12.525 | 34 | 42.5 | 55 | 88 |
| Vocabulary | 26 | 44.192 | 5.671 | 33 | 40.2 | 48 | 55 |
| Table 1c. Older adult participant characteristics | | | | | | | |
| | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
| Age | 28 | 70.464 | 3.717 | 65 | 68.8 | 73.2 | 80 |
| Digit Symbol | 28 | 42.571 | 8.875 | 26 | 36.5 | 46.8 | 62 |
| Vocabulary | 28 | 39.893 | 11.58 | 15 | 32.5 | 50.2 | 54 |

Experiment 1, young adults sample with no familiarization; Experiment 2a, young adult sample with familiarization session; Experiment 2b, older adult sample with familiarization.

### *2.1.2 Materials*

Stimuli ($n = 541$) were pictures previously used in Rabovsky et al. (2016) and are grayscale photos of real world objects set on a light blue background. Pictures were selected to match the word concepts used in the feature norms by McRae et al. (2005). Pictures were scaled to be 3.5° x 3.5° of visual field at a viewing distance of 42 inches.

### *2.1.3 Measures*

**Word frequency** was estimated using the Zipf scale from SUBTLEX-US (Brysbaert & New, 2009; Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The Zipf scale is a logarithmic scale based on word counts from subtitles of American English films. Zipf scale values fall in the range of slightly below 1 and slightly above 7 with greater values indicating greater word frequency.

Participants' familiarity with a word concept was estimated using mean **word familiarity** from McRae et al. (2005). Familiarity ratings were provided for each word concept on a scale of 1-9 (9 being extremely familiar) and averaged across individuals to estimate an average word familiarity.

Each pictures' objective **visual complexity** was estimated using its compressed file size in kilobytes (Donderi, 2006).

**Word length** was estimated using the number of phonemes for the response provided by the participant.

The **number of features** for a word concept was estimated using number of features excluding taxonomic features produced in McRae et al. (2005). Only features which were produced by more than five individuals were included. Taxonomic features were excluded as they represent categorization of concepts which does not reflect the same type of semantic information as other feature types (e.g. part or function).

Square-root **inter-correlational density** was used to account for non-linearity of the measure and to better reflect the analyses conducted in Rabovsky et al. (2016).

The **WAIS III vocabulary** (Wechsler, 1997) test has participants provide definitions for a set of 30 words which are then scored on a 0-2 scale by the experimenter. Bonus items were not included.

The **WAIS III digit symbol** (Wechsler, 1997) test involves participants matching and copying as many symbols to a set of numbers in 90 seconds.

Summary statistics for the item level predictors are shown in Table 2 and correlations between predictors can be found in Table 3.

### *2.1.4 Procedure*

Each participants’ session began with a demographics questionnaire and the WAIS III Digit Symbol test. Individuals were then taken to an individual testing room to complete the experimental task. Upon completion of the experimental task a researcher administered the WAIS III Vocabulary test.

| **Table 2 – Measure Summary Statistics** | | |
|---|---|---|
| Predictor | Mean | SD |
| Word Frequency | 3.59 | 0.81 |
| Word Familiarity | 5.62 | 2.00 |
| Word Length[1] | 4.83 | 1.73 |
| Number of Features | 13 | 3 |
| Visual Complexity[2] | 59.74 | 23.50 |
| Inter-Correlational Density[3] | 10.97 | 7.40 |

1. Number of phonemes
2. Visual Complexity on kilobyte scale
3. Square-root inter-correlational density

**Table 3 – Predictor Correlations**

| | Word Frequency | Word Familiarity | Visual Complexity | Word Length | NOF | ICD |
|---|---|---|---|---|---|---|
| Word Frequency | 1 | | | | | |
| Word Familiarity | 0.351 | 1 | | | | |
| Visual Complexity | 0.086 | 0.045 | 1 | | | |
| Word Length | -0.232 | -0.142 | 0.024 | 1 | | |
| NOF | 0.259 | 0.263 | -0.034 | -0.042 | 1 | |
| ICD | 0.100 | -0.071 | 0.012 | 0.032 | 0.405 | 1 |

NOF = Number of Features, ICD = Inter-correlational Density

The stimuli were presented using EPrime 2.0 (Psychology Software Tools, Pittsburgh, PA). Voice onsets were detected using a microphone connected to a Psychology Software Tools serial response box. During the experiment, participants were first given a verbal description of the task by the experimenter. Participants then read written instructions to themselves before starting the task. Participants were instructed to name the pictures shown as quickly and accurately as possible when they appeared on the screen. The experimenter remained in the room with the participant as they read the instructions and for the first 60 trials to answer any questions and verify proper equipment set up. Participants saw 8 blocks with 60 pictures and a final block with 61. There were 7 breaks spaced evenly throughout the experiment and a check in with the experimenter at the halfway point. Pictures were shown once in a different random order for each participant.

Each trial sequence started with a fixation cross presented in the center of the screen for 1000 ms followed by a brief blank screen for 200 ms.  A picture was then presented for either 4000 ms or until a voice onset was registered after which it remained on the screen for an additional 500 ms. A blank screen was displayed as the inter-trial interval and lasted for 1000 ms. Response times were recorded as the time between presentation of a picture and registration of a voice onset.

Responses were transcribed offline. All responses were coded into 8 different categories described in Table 4.

**Table 4 – Trial coding descriptions**

| Response Coding | Code Description |
|---|---|
| Experimental Error | Instances of outside influence: Hardware didn't detect response, clear interruption during trial. |
| Wrong Response | Response was clearly incorrect as determined by coder |
| Near Response | Response was an alternative name, superordinate name, or sufficiently semantically close as determined by coder (e.g. Lion as cat) |
| Matched Response | Response matched the concept in McRae et al. (2005) |
| Filler sounds | Respondent uttered a clear filler sound prior to their response such as "uh" or "um" |
| Lip smack | Respondent produced a loud smacking noise prior to their response. Used to locate possible quick responses |
| Correction | Respondent either changed their response mid utterance or changed their response after finishing their first utterance. Used to exclude following trial as subject may still be recovering from previous trial. |
| No response | Respondent didn't produce a response. |

## 2.2 Results

For the participants included there were 11,361 observations coded and transcribed. Of those observations 47 (0.4%) were coded as experimental errors, 443 (3.9%) as incorrect responses, 6500 (57.2%) matched the concept from McRae et al (2005), 3999 (35.2%) were deemed semantically close to the matched concept, 111 (0.98%) coded as filler sounds (e.g. "uh" or "um"), 19 (0.17%) registered lip smacks, 90 (0.79%) were response corrections, and 152 (1.34%) had no response. A total of 255 (2.24%) observations did not register a voice onset even though a response was provided and were also excluded from analyses. A total of 9 items had more responses coded as incorrect than near and matched responses so were excluded to avoid including confusing items ($n$ = 14 matched responses).

Only responses coded as matching the word concepts from McRae et al (2005) were used in analyses. If a participant did not produce a label matching the normed concept it is not possible to map accurate feature predictors to that observation. Without a corresponding predictor any estimated model would have excessive missing data or would be using a potentially erroneous predictor for some responses hindering our ability to make strong inferences.

Because the experimenter was present and actively interacting with both the participant and equipment during the first 60 trials, those observations were treated as practice trials. Excessively quick responses (i.e. < 300 ms) were also excluded as these are likely spurious voice onsets due to a filler sound, lip smack, or other vocalization from the participant. Excluding these quick responses removed 24 (0.21%) observations overall and

13 (0.2%) of matched responses. Particularly slow responses were excluded by removing those which were more than 2.5 within participant standard deviations above the within participant mean response time for matched responses. In total 197 (3.03%) matched responses were removed from analyses for being slow.

**Table 5 – Summary statistics of name agreement and proportion matched**

| | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| Experiment 1 | | | | | | | |
| Name Agreement | 541 | 0.716 | 0.226 | 0.152 | 0.515 | 0.939 | 1 |
| Prop. Matched | 541 | 0.544 | 0.354 | 0 | 0.212 | 0.879 | 1 |
| Experiment 2a | | | | | | | |
| Name Agreement | 300 | 0.79 | 0.208 | 0.174 | 0.654 | 0.962 | 1 |
| Prop. Matched | 300 | 0.766 | 0.243 | 0.038 | 0.615 | 0.962 | 1 |
| Experiment 2b | | | | | | | |
| Name Agreement | 300 | 0.802 | 0.204 | 0.182 | 0.663 | 0.962 | 1 |
| Prop. Matched | 300 | 0.708 | 0.272 | 0.038 | 0.5 | 0.923 | 1 |

Experiment 1, young adults naming all pictures from Rabovsky et al. (2016) without familiarization; Experiment 2a, young adults naming subset of 300 pictures from original 541 with familiarization; Experiment 2b, older adults naming same 300 pictures from experiment 2a with familiarization.

Prior to calculating name agreement pluralized forms of the same word were summed to represent a non-plural form (e.g. *blueberries* and *blueberry* counted as the same) as were repeats of the same word (e.g. *blueberry blueberry* and *blueberry* counted as the same). Name agreement values were calculated as a proportion of the dominant response relative to all the responses for a given picture. Additionally, we calculated the proportion of responses for each picture matching the word concept from McRae et al. (2005) referred to as proportion matched. Summary statistics for name agreement and proportion matched can be found in Table 5.

Picture name agreement showed substantial variation (min =15%, max =100%) across items and a slight negative skew (M (sd) = 72% (23%); Mdn = 76%, $g_1$ = -0.41). Because the mean agreement was below our intended inclusion criteria of 80%, many matched responses would be excluded ($n = 2106, 38.28\%$ of included responses) and a large number of the pictures ($n = 213, 45.51\%$ of included items). Instead of further limiting included observations, we decided not to use a name agreement exclusion criteria. This decision was made because name agreement is thought to be capturing aspects of the semantic overlap for a concept word describing a picture. Because of this, we were concerned having a name agreement cutoff would be limiting the range of semantic overlap for responses provided.

Two mixed effects models with crossed random effects (Baayen, Davidson, & Bates, 2008) were fit with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in the R statistical environment (R Development Core Team, 2008). Statistical tests were conducted using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Because response times are highly skewed, they were transformed to response speed, 1/RT(s), to compensate for this non-normality. The initial model predicted response speed with fixed effects word frequency, word familiarity, visual complexity, word length, number of features, and inter-correlational density. Intercepts were allowed to vary over both subjects and items. Slopes for number of features and inter-correlational density varied over subjects.

Word frequency ($b = 0.07, t(463.79) = 4.31, p < 0.001$), word familiarity ($b = 0.02,\ t(444.65), = 4.46,\ p < 0.001$), and number of features ($b = 0.01,\ t(196.94) = 3.84,\ p < 0.001$) all positively predicted response speeds. Visual complexity was a significant negative predictor of response speed ($b = -0.001,\ t(461.75) = -2.19,\ p = 0.029$). While inter-correlational density was not a significant predictor, it showed a marginal negative relationship with response speed ($b = -0.002,\ t(398.84) = -1.74$, $p = 0.083$). Word length was not a significant predictor of response speed ($b = 0.006$, $t(446.31) = 1.12,\ p = 0.265$).

Because word length showed no significant relationship with response speed, a second model was fit excluding word length as a predictor. Again, word frequency ($b = 0.063,\ t(459.51) = 4.34,\ p < 0.001$), word familiarity ($b = 0.02,\ t(446.86) = 4.60$, $p < 0.001$), and number of features ($b = 0.01,\ t(196.67) = 3.89,\ p < 0.001$) were all significant positive predictors of response speed. As we observed in our first model, visual complexity was a significant negative predictor of response speed ($b = -0.001$, $t(460.87) = -2.12, p = 0.035$. Inter-correlational density and response speed also still showed a marginal negative relationship ($b = -0.002, t(399.17) = -1.78, p = 0.077$). A chi-square test of model fit between the two models indicated no improvement of fit

($\chi^2(1) = 1.27,\ p = 0.262$) with the inclusion of word length, so for subsequent analyses it was excluded. Fixed and random effects for both models can be found in Tables 6 and 7.

| **Table 6 – Model estimates Experiment 1, model 1** | | | | |
|---|---|---|---|---|
| Fixed effects | | | | |
| Coefficient | Estimate | Std. Error | df | t-value |
| Intercept | 0.478 | 0.082 | 392.567 | 5.847*** |
| Word Frequency | 0.073 | 0.017 | 463.794 | 4.308*** |
| Word Familiarity | 0.020 | 0.005 | 444.650 | 4.461*** |
| Visual Complexity | -0.001 | 0.000 | 461.754 | -2.191* |
| Word Length | 0.006 | 0.005 | 446.313 | 1.115 |
| ICD | -0.002 | 0.001 | 398.848 | -1.738 |
| NOF | 0.011 | 0.003 | 196.936 | 3.836*** |
| Random Effects | | | | |
| Groups | Name | Variance | Std. Dev. | |
| Subject | Intercept | 0.013 | 0.113 | |
| | NOF | 0.00002 | 0.004 | |
| | ICD | 0.0000005 | 0.001 | |
| Item | Intercept | 0.023 | 0.153 | |
| Residual | | 0.063 | 0.250 | |

NOF = Number of features, ICD = Inter-correlational Density.
* p < 0.05, ** p < 0.01, *** p <0.001

**Table 7 – Model Estimates Experiment 1, model 2**

| Fixed effects | | | | |
|---|---|---|---|---|
| Coefficient | Estimate | Std. Error | df | t-value |
| Intercept | 0.538 | 0.062 | 257.521 | 8.680*** |
| Word Frequency | 0.063 | 0.015 | 459.508 | 4.341*** |
| Word Familiarity | 0.021 | 0.005 | 446.863 | 4.598*** |
| Visual Complexity | -0.001 | 0.0004 | 460.866 | -2.119* |
| NOF | 0.011 | 0.003 | 196.669 | 3.894*** |
| ICD | -0.002 | 0.001 | 399.168 | -1.775 |

| Random Effects | | | |
|---|---|---|---|
| Groups | Name | Variance | Std. Dev. |
| Subject | Intercept | 0.013 | 0.113 |
| | NOF | 0.00002 | 0.004 |
| | ICD | 0.0000005 | 0.001 |
| Item | Intercept | 0.023 | 0.153 |
| Residual | | 0.063 | 0.250 |

NOF = Number of features, ICD = Inter-correlational Density.

* p < 0.05, ** p < 0.01, *** p <0.001

### 2.3 Discussion

One key goal of this initial experiment was to replicate the observation of inter-correlational density predicting picture naming speeds (Rabovsky et al. 2016). While we did not observe this relationship, these data showed the patterns we would expect given the observations from Rabovsky et al. (2016). Naming speed was positively predicted by word familiarity, word frequency, and number of features. Additionally, while visual complexity is not usually included as a predictor in picture naming studies due to mixed evidence for its prediction of response latencies (Alario et al., 2004) and difficulty creating an objective definition (Donderi, 2006; Forsythe, 2009) with these data it reliably predicts slower picture naming response times. We would anticipate visual complexity to negatively predict response speed, as it is here. The assumption is that more visually complex images require more time to process during visual recognition and thus response times should be slower.

Failure to observe inter-correlational density as a significant predictor of picture naming speed is counter to previous observations (Rabovsky et al., 2016). One primary concern in this experiment was the number of observations removed. Given the few constraints participants had when producing names, they often provided responses that differed from the matched word concepts in McRae et al. (2005). Because both inter-correlational density and number of features are only available for specific words it is not possible to map all responses to these item predictors resulting in many responses being excluded (n = 3744, 37.3 % of valid responses). With such a large proportion of responses removed power may not have been sufficient to detect a statistical effect. It should be noted that we do not believe most unmatched responses were removed because the participant was confused on what the image was as most (86.58 %) unmatched responses were coded as semantically near responses (e.g. *bird*, *hawk*, or *eagle* for a picture of *buzzard*)

An additional concern with excluding so many responses is the possibility inter-correlational density and response variability are connected. Inter-correlational density represents the featural overlap of a concept word with other concepts from the McRae et al. (2005) corpus. This means that concepts with high inter-correlational density likely share features and are potentially describable with a similar superordinate response. As an example, 39 of the concepts in McRae et al (2005) are different types of birds and share multiple produced features. However, when naming these different birds many (51%) responses were simply *bird* and on average few responses (M = 0.29) matched the expected concept word. Due to this, it may be the case that certain groups of high inter-correlational density items were excluded systematically.

A follow up item analysis was conducted to address whether high inter-correlational density items were more likely to have unmatched responses. We fit a linear regression model predicting the proportion of matched responses for a picture. The model included the predictors word frequency, word familiarity, visual complexity, number of features, and square root inter-correlational density. The model was fit using the lm function in the R statistical environment (R Development Core Team, 2008). In this model, each of the predictors uniquely contributed to the proportion of matched responses. Word frequency ($b = 0.16, t(535) = 6.78, p < 0.001$), word familiarity ($b = 0.03, t(535) = 3.65, p < 0.001$), and number of features ($b = 0.02,\ t(535) = 5.61,\ p < 0.001$) positively predicted proportion of matched responses while visual complexity ($b = -0.001, t(535) = -2.29, p = 0.022$) and inter-correlational density ($b =$

$-0.01, t(535) = -5.97, p < 0.001$) negatively predicted the proportion of matched responses for a picture. Importantly, pictures of high inter-correlational density concepts appear to elicit a lower proportion of matched responses so are more likely to be excluded from our analyses.

# CHAPTER 3. EXPERIMENT 2A

To increase the number of matching responses, individuals were exposed to each of the pictures were accompanied by the matched concept name from McRae et al., (2005). During the familiarization phase, speakers saw each picture paired with labels. Similar procedures are common within the psycholinguistic literature (e.g. Alario et al., 2004; Llorens, Trébuchon, Riès, Liégeois-Chauvel, & Alario, 2014).

## 3.1 Methods

### *3.1.1 Participants*

A total of 26 undergraduate students from the Georgia Institute of Technology School of Psychology participant pool were recruited. All participants reported English as their first language and normal or corrected to normal vision. All participants were compensated with course credit. Participant characteristics can be found in Table 1b.

### *3.1.2 Materials*

Stimuli were a subset of 300 images from those used in Experiment 1. The 300 images were randomly selected from the original 541 to avoid any experimenter bias and to maximize comparability of predictor distributions. During the familiarization phase, pictures were shown along with the matched concept word (McRae et al., 2005). Demographic information, WAIS Digit Symbol, WAIS Vocabulary, word frequency, word familiarity, visual complexity, number of features, and inter-correlational density were all gathered in the same way as in Experiment 1.

### *3.1.3 Procedure*

When participating in this experiment, individuals started by completing a demographics questionnaire and WAIS Digit Symbol before the experimental task. The experimental task was completed on a computer in a quiet room. After completing the experimental task the experimenter administered the WAIS Vocabulary to the participant

The experiment was presented using EPrime 2.0 (Psychology Software Tools, Pittsburgh, PA). The experimental task was split into two parts: an initial familiarization session and an experimental session. Instructions were administered in the same way as in Experiment 1 with an additional refresher after the participant completed the familiarization session. Trial timing and stimuli presentation for both the familiarization and experimental sessions was identical to Experiment 1.

During familiarization, participants were shown each of the 300 picture stimuli in a random order with the associated word from McRae et al. (2005) displayed below in Arial font. Participants were instructed to look at the picture before saying the displayed word out loud. The experimenter was initially in the room with the participant to describe the task, answer questions, and verify the equipment was set up correctly before stepping out of the room after the first 60 trials.

Once the familiarization session was completed, the participant was instructed to alert the experimenter. The experimenter then described the experimental task to the participant again and answered any questions they might have before leaving the participant to continue. Participants then saw each of the 300 pictures again in a random order without the associated concept word. During this part of the experiment, participants were instructed to name the pictures as quickly and accurately as possible. In order to discourage participants from defaulting to superordinate names (e.g. all birds called *bird*), experimenters indicated to participants their memory was not being tested but that they should name the pictures with the most specific label which felt appropriate. To reduce participant exhaustion, there were 8 breaks spaced evenly throughout the experiment and one experimenter check in at the halfway point.

## 3.2 Results

Responses were coded and transcribed using the coding scheme laid out in Table 4. Responses were excluded following the same criteria as Experiment 1 with the exception that familiarity trials were treated as practice trials.

Data was analyzed using a mixed-effects model with crossed random effects (Baayen, Davidson, & Bates, 2008). Models were fit in the R statistical environment (R Development Core Team, 2008) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and statistical tests conducted using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). The model estimated was based off the final model in Experiment 1 predicting response speed with fixed effects of word frequency, word familiarity, visual complexity, number of features, and inter-correlational density. Random intercepts were estimated for both subjects and items while random slopes were estimated for subjects on inter-correlational density and number of features.

As in our previous experiment, word frequency ($b = 0.08$, $t(289.98) = 5.08$, $p < 0.001$) and number of features ($b = 0.008, t(264.56) = 2.46, p = 0.015$) were both positive predictors of response speed. However, with familiarization, neither word familiarity ($b = 0.0004,\ t(275.47) = 0.79,\ p = 0.433$) nor visual complexity ($b = -0.0005,\ t(287.04) = -1.46,\ p = 1.457$) showed a significant relationship with response speed. Inter-correlational density, however, was a significant negative predictor of response speed ($b = -0.006,\ t(259.09) = -4.537,\ p < 0.001$), predicting slower responses as inter-correlational density increased. Random and fixed effects for this model can be found in Table 8.

Inspection of correlations between random effects indicated that much of the variation across subjects for the number of features was being captured by the random slope of inter-correlational density ($r = 0.94$). A second model was fit matching the previous analysis but excluded the random slope for number of features in order to test whether estimating the additional parameters greatly improves model fit. The results from this analysis closely matched those from our previous model. Word frequency ($b = 0.08$, $t(290.00) = 5.08,\ p < 0.001$) and number of features ($b = 0.008,\ t(289.25) = 2.48,\ p = 0.014$) were both significant positive predictors of response speed. Response speed also showed a significant negative relationship with inter-correlational density ($b = -0.006,\ t(233.14) = -4.50,\ p < 0.001$). As in the previous model, visual complexity ($b = -0.0005,\ t(287.06) = -1.46,\ p = 0.145$) was a non-significant negative predictor and familiarity ($b = 0.004,\ t(275.47) = 0.79,\ p = 0.431$) was a non-significant positive predictor of response speed. A chi-square model comparison test ($\chi^2(3) = 2.69,\ p = 0.443$) indicated no significant improvement of model fit with the inclusion of the random parameters for number of features. A table of fixed and random effects for this model can be found in Table 9.

**Table 8 – Model estimates Experiment 2a, model 1**

Fixed effects

| Coefficient | Estimate | Std. Error | df | t-value |
|---|---|---|---|---|
| Intercept | 0.717 | 0.064 | 208.246 | 11.254*** |
| Word Frequency | 0.076 | 0.015 | 289.981 | 5.077*** |
| Word Familiarity | 0.004 | 0.005 | 275.473 | 0.786 |
| Visual Complexity | -0.001 | 0.0003 | 287.041 | -1.459 |
| NOF | 0.008 | 0.003 | 264.562 | 2.460* |
| ICD | -0.006 | 0.001 | 259.089 | -4.537*** |

Random Effects

| Groups | Name | Variance | Std. Dev. |
|---|---|---|---|
| Subject | Intercept | 0.022 | 0.148 |
| | NOF | 0.000007 | 0.003 |
| | ICD | 0.000002 | 0.001 |
| Item | Intercept | 0.018 | 0.132 |
| Residual | | 0.069 | 0.262 |

NOF = Number of features, ICD = Inter-correlational Density.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

| **Table 9 – Model estimates Experiment 2a, model 2** | | | | |
|---|---|---|---|---|
| Fixed effects | | | | |
| Coefficient | Estimate | Std. Error | df | t-value |
| Intercept | 0.717 | 0.062 | 275.080 | 11.531*** |
| Word Frequency | 0.076 | 0.015 | 289.999 | 5.076*** |
| Word Familiarity | 0.004 | 0.005 | 275.468 | 0.788 |
| Visual Complexity | -0.001 | 0.0003 | 287.058 | -1.460 |
| NOF | 0.008 | 0.003 | 289.251 | 2.481* |
| ICD | -0.006 | 0.001 | 233.135 | -4.496*** |
| Random Effects | | | | |
| Groups | Name | Variance | Std. Dev. | |
| Subject | Intercept | 0.017 | 0.130 | |
| | ICD | 0.000002 | 0.002 | |
| Item | Intercept | 0.017 | 0.132 | |
| Residual | | 0.069 | 0.262 | |

NOF = Number of features, ICD = Inter-correlational Density.

* p < 0.05, ** p < 0.01, *** p <0.001

### 3.3 Discussion

The addition of a familiarization session impacted our observations in meaningful ways compared to Experiment 1. As expected, given previous observations, including our own, both increasing word frequency (e.g. Alario et al., 2004) and number of features (Rabosky et al. 2016) were related to shorter response speeds. In line with our original hypothesis, higher inter-correlational density was associated with longer response speeds. Interestingly, in this experiment, we deviated from our previous observations as we did not observe slower responses with increasing visual complexity or quicker responses for pictures of concepts with higher word familiarity.

The observation of a negative relationship between response speed and inter-correlational density is in agreement with the idea that the more semantic overlap a concept has then selection and subsequent production of a word describing that concept will be slowed (Abdel Rahman & Melinger, 2009; Dell, 1986; Howard et al., 2006; Levelt et al., 1999; Roelofs, 1992). There are a few explanations for why adding familiarization led to observing an effect of inter-correlational density. The simplest explanation is either our previous experiment failed to detect a significant relationship or this experiment showed such a relationship simply due to random chance. Given the limited examples of using inter-correlational density as a semantic predictor in this type of paradigm, it is difficult to address the possibility of a chance finding. In an attempt to do so, we re-analyzed a set of data from an unfamiliarized paradigm similar to Experiment 1 (Lagrone & Spieler, 2006). After matching the responses in those data to the concept words from McRae et al. (2005), we were able to fit the same models used in Experiments 1 and 2. In that analysis, we saw a similar pattern of results as Experiment 1 without a significant relationship between inter-correlational density and response speeds. It therefore seems unlikely that chance alone explains the statistical differences observed and that familiarization impacts our observations in other ways.

Originally, we had assumed failure to detect an effect was due to a lack of power which familiarization would alleviate by increasing our number of valid responses. To an extent, we achieved this as there was a substantial increase in the number of matched responses in Experiment 2a ($M(sd) = 0.76(0.24)$, $Mdn = 0.85$) compared to Experiment 1 ($M(sd) = 0.54(0.35)$, $Mdn = 0.64$). This increase in proportion of matched responses led to including more observations in Experiment 2a ($n = 5{,}734$) than in Experiment 1 ($n = 5{,}502$) despite the total number of observations being greater in Experiment 1 (exp1: $n = 11{,}361$; exp2: $n = 7{,}800$).

We were also concerned that by excluding so many responses, we were systematically excluding observations of concept words with high inter-correlational density. This was in part addressed, as in Experiment 2a only one picture elicited no matching responses while in Experiment 1 sixty-six pictures had no matching responses. However, when we fit a linear regression model predicting proportion of matched responses for a picture, we found very similar results to those from our previous experiment. Word frequency ($b = 0.05$, $t(294) = 3.18$, $p = 0.002$), word familiarity ($b = 0.02$, $t(294) = 3.30$, $p = 0.001$), and number of features ($b = 0.02$, $t(294) = 4.06$, $p < 0.001$) all positively predicted the proportion of matched responses. Additionally, greater inter-correlational density ($b = -0.01$, $t(294) = -6.77$, $p < 0.001$) still predicted a lower proportion of matched responses. This indicates that even though we improved the overall proportion of matched responses, observations with high inter-correlational density were still more likely to be excluded.

With familiarization, the relationship between visual complexity and response speed is minimal, which makes some intuitive sense. Visual complexity should reduce response speeds in part because visual recognition of more complicated pictures takes more time. Once participants have been exposed to the images and some appropriate names for those pictures the process of visual recognition is likely quicker. Once familiarized to the pictures, visual recognition time may be fairly uniform across the range of pictures or at the very least reduced to the point where differences are undetectable in this context. This is further supported by our secondary analysis predicting proportion of matched responses, where more complex pictures were no more likely to have unmatched responses.

While we were expecting word familiarity to show a relationship with response speed, even with familiarization, there are reasons one might not. Some authors have argued that object familiarity is closely connected to visual recognition (Forsythe, 2009; Logothetis, & Sheinberg, 1996). To the extent word familiarity captures object familiarity, one might expect pictures described by more familiar words to be recognized easier, and thus have faster response speeds. If familiarization is impacting the process of visual recognition, then after familiarization, the benefits of a word being more familiar may be reduced sufficiently to make any relationship of word familiarity and response speed undetectable.

# CHAPTER 4. EXPERIMENT 2B

Addition of a familiarization session appears to have provided a sufficient increase in power and reduction of excluded responses to observe a negative relationship between inter-correlational density and response speed. Given this observation, we recruited a group of community-dwelling older adults to address the question of whether adult aging would have an impact on semantic interference predicted by inter-correlational density.

## 4.1 Methods

### *4.1.1 Participants*

A total of 26 community dwelling older adults were recruited from the Atlanta area. All participants reported corrected to normal vision and English as their first language. No participants reported a history of dementia, cognitive impairment, serious head injury, cerebrovascular events (e.g. stroke), or serious ongoing cardiovascular issues. Participants were compensated with around $25 depending upon the time it took to complete the experiment. Participant characteristics can be found in Table 1c.

### *4.1.2 Materials*

Materials used were identical to those used in experiment 2a, except for 20 additional pictures, which were included as practice trials. The practice stimuli were selected from the remaining 241 pictures not included in experiment 2a and were paired with the matched concept word from McRae et al. (2005).

### *4.1.3 Procedure*

Experiment 2b followed the same procedure as experiment 2a except for three primary differences. Prior to the familiarization session, individuals were shown the 20 pictures selected as practice trials in a random order. These practice trials were displayed in the same way as the familiarization session, ending with a prompt encouraging the participant to ask the experimenter any questions they might have. Additionally, the experimenter was in the room with the participant for the entirety of the experiment. This change was decided upon after data from an initial pilot participant suggested that some older adults may have trouble responding consistently at a level where voice onsets could be registered. To avoid losing excessive data, the experimenter was in the room with the participant to verify proper voice onsets were registered throughout the experimental session. The inter-trial interval was also extended to 1500 ms, allowing older adults to better recover from a response and prepare for the next trial.

## 4.2 Results

As in both previous experiments, a mixed effects model with crossed random effects (Baayen, Davidson, & Bates, 2008) was fit in the R statistical framework (R Development Core Team, 2008) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) with statistical tests conducted using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Response speeds were modeled to be predicted by word frequency, word familiarity, visual complexity, number of features and inter-correlational density as fixed effects. Intercepts were allowed to vary across random the factors, subjects and items. Because our previous experiment suggested accounting for variability in the slopes for number of features across subjects added no improvement in model fit, only the slope of inter-correlational density was allowed to vary across subjects.

**Table 10 – Model estimates Experiment 2b, older adults only**

Fixed effects

| Coefficient | Estimate | Std. Error | df | t-value |
|---|---|---|---|---|
| Intercept | 0.548 | 0.068 | 286.267 | 8.020*** |
| Word Frequency | 0.080 | 0.017 | 301.191 | 4.867*** |
| Word Familiarity | 0.019 | 0.005 | 286.248 | 3.517*** |
| Visual Complexity | -0.001 | 0.0004 | 293.727 | -1.297 |
| NOF | 0.006 | 0.004 | 292.884 | 1.789 |
| ICD | -0.005 | 0.001 | 245.315 | -3.153*** |

Random Effects

| Groups | Name | Variance | Std. Dev. |
|---|---|---|---|
| Subject | Intercept | 0.020 | 0.141 |
| | ICD | 0.000003 | 0.002 |
| Item | Intercept | 0.022 | 0.148 |
| Residual | | 0.054 | 0.233 |

NOF = Number of features, ICD = Inter-correlational Density.

* p < 0.05, ** p < 0.01, *** p <0.001

Word frequency ($b = 0.08$, $t(301.192) = 4.87$, $p < 0.001$) and word familiarity ($b = 0.02$, $t(286.25) = 3.52$, $p < 0.001$) again were significant positive predictors of response speed. As we observed with the young adults, inter-correlational density ($b = -0.005$, $t(245.32) = -3.15$, $p = 0.002$) showed a significant negative relationship with response speed. Visual complexity was also a non-significant negative predictor ($b = -0.0005$, $t(293.73) = -1.30$, $p = 0.196$) of response speed. Unlike our results from Experiment 2a, number of features ($b = 0.006$, $t(292.88) = 1.79$, $p = 0.075$) did not show a significant relationship with response speed, though the coefficient was still positive. Both random and fixed effects can be seen in Table 10.

## 4.3 Discussion

These data, for the most part, are consistent with our observations in younger adults. Response speed shows the expected positive relationship with word frequency, the negative relationship with inter-correlational density, and a limited relationship with visual complexity as we saw in Experiment 2a. However, in this sample, older adults did not show as strong a relationship of naming times with number of features. Additionally, even with familiarization, the older adults still seem to be quicker when naming pictures of more familiar concept words where the younger adults did not.

While number of features was not a significant predictor of response speed this does not necessarily indicate facilitation was not present in older adults. If we look at the coefficients and standard errors for number of features in both the young adult ($b = 0.008$, $se = 0.003$) and older adult group ($b = 0.006$, $se = 0.004$), we see that the magnitude and direction of the relationship predicted is very similar. Older adults tend to show greater within- and between-individual variability for response latencies (e.g. Hultsch, MacDonald, & Dixon, 2002), which can limit our ability to detect statistical effects. Failing to detect a significant relationship may be in part due to the very small effects being observed and the increase in variability and thus standard errors of the estimates for older adults.

# CHAPTER 5. EXPERIMENT 2A & 2B COMBINED ANALYSIS

## 5.1 Results

Observations from both younger adults and older adults were analyzed together to address the question of whether adult aging impacts semantic competition during picture naming predicted by inter-correlational density. A mixed effects model with crossed random effects (Baayen, Davidson, & Bates, 2008) was estimated using lme4 (Bates, Mächler, Bolker, & Walker, 2015) with statistical tests conducted using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) in the R statistical environment (R Development Core Team, 2008). Response speed was predicted by fixed effects of word frequency, word familiarity, visual complexity, age group, number of features, inter-correlational density, and the interactions of age group with number of features, inter-correlational density, and familiarity. Random intercepts were estimated for items and subjects. Random slopes for inter-correlational density were estimated for subjects. Age group was coded as “0” for young adults and “1” for older adults.

In these data, response speed had a significant positive relationship with word frequency ($b = 0.08,\ t(295.30) = 5.36,\ p < 0.001$) and number of features ($b = 0.008,\ t(340.24) = 2.56,\ p = 0.011$). Significantly slower response speeds were seen in the older adult group ($b = -0.16,\ t(82.63) = -3.59,\ p < 0.001$). Increasing inter-correlational density also showed the expected slower response speeds ($b = -0.006,\ t(311.89) = -4.632,\ p < 0.001$). When both groups were considered together, word familiarity ($b = 0.004,\ t(320.15) = 0.75,\ p = 0.452$) did not show significant prediction of quicker response speeds. Additionally, visual complexity ($b = -0.0005,\ t(290.78) = -1.47,\ p = 0.142$) showed a non-significant relationship with response speed.

The negative relationship observed in inter-correlational did not differ across age groups ($b = 0.001,\ t(70.26) = 1.30,\ p = 0.199$). Similarly, the positive relationship of response speed and number of features did not significantly vary across ages ($b = -0.001,\ t(10639.37) = -0.65,\ p = 0.517$). However, the positive influence of word familiarity on response speeds was greater in older adults ($b = 0.02,\ t(10639.52) = 6.39,\ p < 0.001$). Random and fixed effects can be found organized in Table 11.

## 5.2 Discussion

Across the two groups we see some similar patterns. Response times are shortened with increasing values of word frequency and number of features. At the same time, we see longer response latencies with increasing semantic feature overlap. With familiarization, visually complex pictures do not appear any slower to name than simpler pictures. Older adults overall were slower than younger adults to provide names for the pictures shown. This is in line with previous research (e.g. Au et al., 1995; Nicholas, Barth, Obler, Au, & Albert, 1997; Thomas, Fozard, & Waugh, 1977) but is not particularly interesting in this context as it likely reflects changes not specific to lexical processes, such as speed of processing (e.g. Salthouse, 2016; Schaie, 2013). Older adults showed more facilitation from word familiarity than younger adults. However, we did not observe differences between younger and older adults in facilitation from number of features or interference due to semantic overlap.

**Table 11 – Model estimates Experiments 2a and 2b combined**

Fixed Effects

| Coefficient | Estimate | Std. Error | df | t-value |
|---|---|---|---|---|
| Intercept | 0.710 | 0.061 | 349.139 | 11.582*** |
| Word Frequency | 0.077 | 0.014 | 295.296 | 5.363*** |
| Word Familiarity | 0.004 | 0.005 | 320.148 | 0.752 |
| Visual Complexity | -0.001 | 0.0003 | 290.783 | -1.473 |
| Age | -0.156 | 0.044 | 82.628 | -3.588*** |
| NOF | 0.008 | 0.003 | 340.237 | 2.556* |
| ICD | -0.006 | 0.001 | 311.893 | -4.632*** |
| Age*ICD | 0.001 | 0.001 | 70.255 | 1.297 |
| Age*NOF | -0.001 | 0.002 | 10639.374 | -0.648 |
| Age*Familiarity | 0.016 | 0.003 | 10639.516 | 6.385*** |

Random Effects

| Groups | Name | Variance | Std. Dev. |
|---|---|---|---|
| Subject | Intercept | 0.018 | 0.135 |
| | ICD | 0.000002 | 0.001 |
| Item | Intercept | 0.018 | 0.134 |
| Residual | | 0.064 | 0.252 |

Age coded as 0 = Young adult, 1 = Old adult; NOF = Number of features, ICD = Inter-correlational Density, Familiarity = Word Familiarity.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As we mentioned from our previous analysis, the failure to observe a significant relationship between number of features and response speed was unlikely to reflect a genuine age difference, which appears to be supported here. However, one could apply the same argument by saying that the failure to detect a significant interaction between age and number of features is due the variability in response times of older adults and the small effects we are observing. This could be the case, but there is no clear theoretical reason older adults would not show facilitation during picture naming from the increased semantic information captured by number of features.

Older adults showing greater facilitation in response times from word familiarity was not one of our primary hypotheses, but might be expected (e.g. Newman, & German, 2005). Word familiarity ratings capture many aspects of processing during picture naming (Balota, Pilotti, & Cortese, 2001; Newman, & German, 2005). As we mentioned before, to the extent that word familiarity reflects object familiarity, it may be capturing aspects recognition processing (Alario et al. 2004; Forsythe 2009). Beyond just a role in recognition, word familiarity may also impact the transmission of information following lemma selection to later stages, such as word form or phoneme selection (Newman & German, 2005). More familiar words may have stronger connections between their associated lemmas and connected representations in these later stages. Older adults might show reduced efficiency in transmission between processing stages, introducing noise and making processing slower (Burke et al., 1991). Having stronger connections between the lemmas of familiar words and later lexical representations would be protective against reductions in transmission efficiency and increase the differences of facilitation from familiarity between young and older adults.

Counter to our original hypothesis, we did not observe differences in semantic competition between age groups. In the case of naming pictures with varying name agreement, older adults appear to be more impacted by high semantic competition compared to younger adults (Lagrone, & Spieler, 2006). Here, we see no such pattern, and older adults show similar competition to younger adults across the range of inter-correlational density. This seems to suggest a difference in the type of competition we observe with inter-correlational density compared to name agreement.

In picture naming studies using name agreement, stimuli are specifically chosen so some stimuli have multiple accurate responses (e.g. *couch/sofa)*. Pictures that elicit fewer dominant responses are thought to capture aspects of semantic competition caused by high semantic overlap between a set of highly active competitors. In the case of inter-correlational density, the semantic overlap estimated is more general in nature and reflects not just relationships between very close alternatives, but also overlap with objects that share broader aspects of semantic information, such as context, function, and form. Because of this, the competition predicted by name agreement likely represents very high competition between a small set of closely related competitors while inter-correlational density represents a weaker, but more general type of semantic overlap. Looking at Figure 2, we can see the name agreement of a picture and the inter-correlational density for the associated concept word share a weak relationship for both young ($r = -0.239$) and older adults ($r = -0.244$), supporting the idea that the two measures are capturing different aspects of semantic overlap.

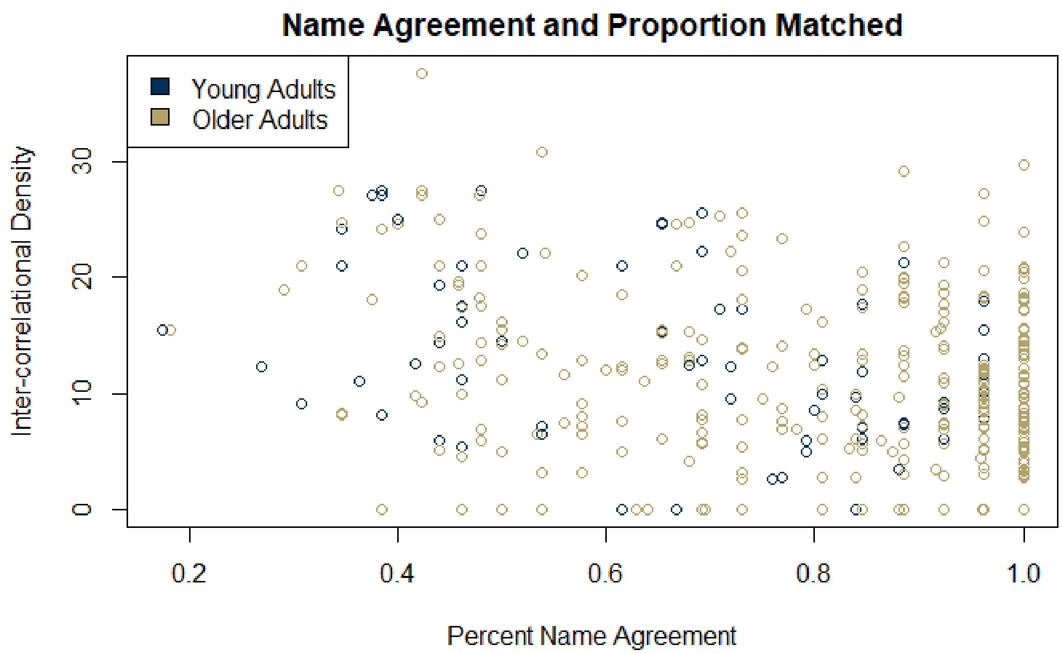


**Figure 2 – Picture name agreement and inter-correlational density.**

**Percent name agreement by square root inter-correlational density of matched concepts in younger and older adult samples with familiarization.**

If older adults show greater competition during lemma selection due to increased noise, we might expect competition between lemmas with very similar levels of activation to be particularly difficult to resolve in older adults. When older adults are naming pictures with medium name agreement (e.g. *couch/sofa*), increased noise in the system would make distinguishing a winner between two lemmas with very close activation levels more difficult. In the case of inter-correlational density, which captures semantic overlap more generally, we would expect greater differences between the activation levels of active lemmas. With greater distinctions between competitors, increased noise in older adult processing would impact competition to a limited degree and competition would look more similar to younger adults.

# CHAPTER 6. CONCLUSION

Through a series of experiments, we attempted to address our primary question of whether older adults are more impacted by semantic competition during lexical selection as predicted by a semantic density measure. To that effect, these data did not provide support for this hypothesis. However, this may be due to differences in the levels of semantic competition predicted by variations in name agreement versus inter-correlational density. In the process of asking this question, we were able to replicate the observation of inter-correlational density as a semantic predictor of competition, at least in the context of picture naming with pre-exposure to stimuli and matched concept words. Without pre-exposure, semantic competition was not clearly observed in young adults. When pre-exposed to the stimuli and matched responses, though, both young and older adults showed clear semantic competition during picture naming, predicted by inter-correlational density. It may be the case that without pre-exposure, responses were too variable, reducing our ability to detect a relationship. We also observed differential influences of word familiarity across age groups during picture naming.

While we failed to provide support for our original hypothesis, several limitations may be informative for future work. Inter-correlational density represents the semantic feature overlap for a limited corpus of words and as such limits our ability to use unmatched responses. By using a broader semantic density measure, we would be better able to include responses which do not correspond to the matched concept word. Additionally, because of how inter-correlational density is calculated, there are groups of objects which share many features and thus have high-inter-correlational density but can be can easily described using category or superordinate labels and are thus excluded. Using a semantic density metric not limited by a small set of concept words may allow researchers to better select items from a range of semantic density but avoid over selection from a specific category. One additional concern with these data is our failure to account for participant differences, specifically the level of education. Given that our younger adult sample is made up of college-age undergraduate students, we can make some reasonable assumptions about their education level. In our older adult sample, we recruited from the Atlanta community with little control on such participant characteristics. In future work, to the extent possible, matching groups on participant characteristics may lead to better group comparability for the process of interest. Future work should take efforts to better consider participant differences as well as the limitations of using inter-correlational density as a semantic overlap metric and consider alternatives for selection of items or predictors.

# REFERENCES


Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. Language and Cognitive Processes, 24(5), 713-734.

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. Behavior Research Methods Instruments & Computers, 36(1), 140-155.

Au, R., Joung, P., Nicholas, M., Obler, L. K., Kass, R., & Albert, M. L. (1995). Naming ability across the adult life span. Aging, Neuropsychology, and Cognition, 2(4), 300-311.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 59(4), 390-412.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. Memory & cognition, 29(4), 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 Section a-Human Experimental Psychology, 50(3), 560-585.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. Journal of Statistical Software, 67(1).

Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: evidence from blocked and continuous manipulations of semantic context. The Quarterly Journal of Experimental Psychology, 66(11), 2135-2160.

Berg, T., & Levelt, W. J. M. (1990). Speaking: From Intention to Articulation. The American Journal of Psychology, 103(3), 409.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. British Journal of Psychology, 93(1), 89-114.

Britt, A. E., Ferrara, C., & Mirman, D. (2016). Distinct Effects of Lexical and Semantic Competition during Picture Naming in Younger Adults, Older Adults, and People with Aphasia. Frontiers in Psychology, 7, 813.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior research methods, 41(4), 977-990.

Burke, D. M., & Mackay, D. G. (1997). Memory, language, and ageing. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 352(1363), 1845-1856.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? Journal of Memory and Language, 30(5), 542-579.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93(3), 283-321.

Donderi, D. C. (2006). An information theory analysis of visual complexity and dissimilarity. Perception, 35(6), 823-835.

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. Journal of Memory and Language, 35(5), 708-723.

Forsythe, A. (2009, July). Visual complexity: is that all there is?. In International Conference on Engineering Psychology and Cognitive Ergonomics (pp. 158-166). Springer, Berlin, Heidelberg.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. Cognition, 100(3), 464-482.

Hultsch, D. F., MacDonald, S. W., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57(2), P101-P115.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word-Frequency Effects in Speech Production - Retrieval of Syntactic Information and of Phonological Form. Journal of Experimental Psychology-Learning Memory and Cognition, 20(4), 824-843.

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. Journal of Verbal Learning and Verbal Behavior, 23(1), 39-66.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13).

LaGrone, S., & Spieler, D. H. (2006). Lexical competition and phonological encoding in young and older speakers. Psychology and Aging, 21(4), 804-809.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22(1), 1-38; discussion 38-75.

Llorens, A., Trébuchon, A., Riès, S., Liégeois-Chauvel, C., & Alario, F. X. (2014). How familiarization and repetition modulate the picture naming network. Brain and language, 133, 47-58.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. Annual review of neuroscience, 19(1), 577-621.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(3), 503-535.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. Behavior Research Methods, 37(4), 547-559.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: processing is slowed by near neighbors and speeded by distant neighbors. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(1), 65-79.

Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. Language and Speech, 48(2), 123-156.

Nicholas, M., Barth, C., Obler, L. K., Au, R., & Albert, M. (1997). Naming in Normal Aging and Dementia of The Alzheimer's Type. In Goodglass & Wingfield (Eds.), Anomia: Neuroanatomical and Cognitive Correlates (pp. 166-188): Academic Press.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. Cognition, 114(2), 227-252.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from https://www.pstnet.com.

R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. Cognition, 146, 240-244.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. Cognition, 42(1-3), 107-142.

Ronnlund, M., & Nilsson, L. G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. Intelligence, 36(3), 192-209.

Salthouse, T. A. (2016). Aging Cognition Unconfounded by Prior Test Experience. Journals of Gerontology: Psychological Sciences, 71(1), 49-58.

Santiago, J., MacKay, D. G., Palma, A., & Rho, C. (2000). Sequential activation processes in producing words and syllables: Evidence from picture naming. Language and Cognitive Processes, 15(1), 1-44.

Schaie. (2013). Developmental influences on adult intelligence: The Seattle Longitudinal Study (2nd ed.). New York, NY: Oxford University Press.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the Time Course of Lexical Access in Language Production - Picture-Word Interference Studies. Journal of Memory and Language, 29(1), 86-102.

Schroeder, D. H., & Salthouse, T. A. (2004). Age-related effects on cognition between 20 and 50 years of age. Personality and Individual Differences, 36(2), 393-404.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. Behavior Research Methods Instruments & Computers, 28(4), 516-536.

Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: Naming in the picture-word interference task. Psychology and Aging, 17(4), 662-676.

Thomas, J. C., Fozard, J. L., & Waugh, N. C. (1977). Age-Related Differences in Naming Latency. The American Journal of Psychology, 90(3), 499.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. The Quarterly Journal of Experimental Psychology, 67(6), 1176-1190.

Verhaeghen, P. (2003). Aging and vocabulary scores: a meta-analysis. Psychol Aging, 18(2), 332-339.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale–Third Edition and Wechsler Memory Scale–Third Edition technical manual. San Antonio, TX, The Psychological Corporation.