Spring 2019

# Forget me, forget me not: unlearning incorrect associations in word learning

Tanja Charlotte Roembke
*University of Iowa*



Follow this and additional works at: https://ir.uiowa.edu/etd

Part of the Psychology Commons

# FORGET ME, FORGET ME NOT: UNLEARNING INCORRECT ASSOCIATIONS IN WORD LEARNING

by

Tanja Charlotte Roembke (Römbke)

A thesis submitted in partial fulfilment
of the requirements for the Doctor of Philosophy
degree in Psychology in the
Graduate College of
The University of Iowa

May 2019

Thesis Supervisor: Professor Bob McMurray

For Tyler.

Willst du immer weiter schweifen?
Sieh, das Gute liegt so nah.
Lerne nur das Glück ergreifen,
Denn das Glück ist immer da.

- Johann Wolfgang von Goethe

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Bob (thebob) McMurray for his mentorship, for (almost) always getting excited about data even when I did not see the forest for the trees and for providing a safe environment where I could grow as a scientist. Moreover, I would like to thank my committee members Susan Wagner Cook, Eliot Hazeltine, Deborah Reed, and Edward Wasserman for their valuable feedback, advice and patience. Special thanks go to Eliot Hazeltine for his advice and chocolate supply. I would also like to thank the University of Iowa's development and (language) learning community, without whom the breadth of my training would not have been possible.

Second, I am grateful to the incredible members and friends of the MACLab for their help and support. In particular, I am thankful to the female scientists that made the basement of Spence Laboratories such a special, compassionate place to work: Effie Kapnoula for her mentoring, friendship, knowledge of smart short-cuts and being an overall role model; Jamie Klein-Packard for wise words, collecting much data and always lending an ear in times of stress; Kayleen Schreiber for her friendship, especially during my first year at the University of Iowa, and representing the importance of following one's own life path; McCall Sarrett for her honesty, ice cream and for introducing me to cheese curds; Sarah Colby for providing insights into the post-PhD life and the magic of cats.

Third, I would like to thank: Marcus Galle (and Keith Apfelbaum) for building the MACLab community; Keith Apfelbaum for his advice on small and big decisions; Claire Goodwin for her research support, willingness to organize lab activities and the best snacks; Michael Seedorff for his statistical expertise and great lunches; Michael Freedberg for keeping me sane through data collection in my second year as a PhD student; Aimee Marino for

recruiting children and broadening my horizon; Francis Smith for the best puns and teaching advice. Of course, data collection would not have been possible without all the dedicated undergraduate MACLab research assistants.

Fourth, my gratitude goes out to Megan Lorenz: Her friendship has been invaluable in navigating graduate school and what comes after. I am also thankful to Valerie Beck for introducing me to sewing, a welcome outlet during my PhD, and my PhD cohort.

Finally, I would like to thank my family and friends that are not affiliated with the University of Iowa. Special thanks go to my husband, Tyler Dryg, who inspires me daily with his idealistic outlook on life. I am thankful for all the early and late rides to/from work, the ice cream, the hugs and never doubting my intellectual capacity. I am thankful for my parents and family back in Germany for their never-ending support. Finally, I would like to express my gratitude to my in-laws for welcoming me into their family and helping me feel at home in Iowa.

ABSTRACT

During word learning, many words are associated with many meanings to build a lexicon. A model by McMurray et al. (2012) suggests that vocabulary acquisition may not only depend on building correct associations, but also pruning incorrect ones. Additional evidence for the importance of pruning comes from a word learning analog in pigeons, where the opportunity for pruning incorrect associations between objects and symbols was manipulated during training (Roembke et al., 2016). To investigate pruning in humans, we conducted six supervised word learning experiments. Participants were first trained to link two objects to each word, and subsequently were tested how quickly these were pruned. Across experiments, association strength was measured by using either eye movements to to-be pruned objects, or a post-training accuracy assessment. Learners showed rapid—though potentially not complete—pruning of incorrect associations, but this depended on whether the symbols were auditory words, orthographic words or non-linguistic symbols. Thus, this dissertation provides first evidence that pruning is operative during word learning. We also examined how newly learned words compete against known words for recognition using eye-tracking and found that despite very high accuracy these words were not strong competitors.

PUBLIC ABSTRACT

People are estimated to acquire tens of thousands of words by the time they reach adulthood. To do this, they have to form associations between words and objects to build a vocabulary. Data from an animal model of word learning as well as data from computational modeling suggest that people not only strengthen associations between words and objects but also prune incorrect associations. This means that people learn not only that the word "ball" maps onto the object "ball" but also that the word "ball" does *not* map onto the objects "bat" or "apple".

As there currently is no behavioral evidence for such pruning in humans, I conducted six experiments in adults in my dissertation. In all experiments, participants were first trained to link two objects to each word, so that each word was associated with two meanings. Subsequently, they learned that each word only mapped onto one object. However, during training, the amount of experience with objects that were available as choices was manipulated, offering the possibility to prune some incorrect associations but not others. Throughout training, participants received feedback as to whether they selected the target object. I measured association strength between words and objects using eye-movements, a post-training accuracy assessment and a yes/no task. Across experiments, I manipulated whether symbols were auditory words, orthographic words or non-linguistic symbols.

Participants showed rapid pruning of incorrect associations, and this process was accelerated when words were auditory. Thus, speed of pruning may be related to ease of acquiring the mappings. Moreover, there was evidence that small associations between symbols and objects remained, and that they could influence subsequent word learning.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

## 1.1 Many-to-many-mappings in cognition

Much of human learning can be characterized by the creation of mappings between individual stimuli or between an individual stimulus and a response. However, in most real learning problems, these individual associations are built within the context of a bigger learning task involving many stimuli and responses. For instance, while learning to type, people map an individual finger response or hand position to a single written letter or key. However, these mappings do not exist in isolation, but are part of a larger assortment of finger-letter-pairs. Similarly, in learning to read, children learn to map sounds onto letters. In English, this requires the mapping between 26 letters and 44 phonemes. In contrast to typing this relationship is not one-to-one in reading: The same letter can map onto several phonemes, the same phoneme can map onto several letters, and so on.

In these examples, people form associations between a sometimes sizeable number of stimuli and/or responses; we will refer to these as many-to-many mappings. The goal of this dissertation is to better understand how such mappings are acquired and a functional *network* is built. Word learning is a perfect domain in which to study the acquisition of many-to-many mappings. People learn to map words (symbols) onto meanings on a large scale. Acquiring a new word involves visually encoding the referent, auditorily encoding its label, and forming an association between the two (McMurray, Horst, & Samuelson, 2012). However, by adulthood, a person typically learns approximately 17,000 base words or word families (Goulden, Nation, & Read, 1990). Moreover, complexity is added as a single word form often maps onto multiple meanings, and a single meaning can map onto multiple word forms. On top of mapping auditory

words to meaning, people also are required to map articulatory movements to meaning and meaning to written words (orthography) as they learn to read.

One obvious property of a well-organized network is strong positive associations between to-be-connected items. A less obvious property is that associations between words and *not*-to-be connected items are minimal (McMurray et al., 2012). Learning a word does not only require "knowing" which meaning maps onto which word (strengthening correct connections), but also to learn which competitor meanings do not map onto that word (weakening incorrect connections). The goal of this dissertation is to better understand how learners minimize the strength and number of incorrect associations within a network.

## 1.1 Word learning as an associative process

The lexicon—the collection of known words in adults—is often conceptualized as an associative network (Elman, 2004). However, word learning as an associative, gradual learning process is more controversial. Much research in children's word learning has been framed around the problem of referential ambiguity, the notion that every single word learning situation is highly ambiguous due to the infinite number of potential referents (Quine, 1960). For instance, if a child hears the word GAVAGAI when seeing a rabbit in a field, they may not know whether this unknown term refers to the rabbit they see, its fur, any other aspect of the environment or even an absent, abstract concept. At its core, the problem of referential ambiguity thus is the consequence of a lack of information as to a novel word's referent in the moment of naming.

Theories of word learning have often suggested that children possess specialized knowledge or biases to overcome this problem and succeed at mapping a word to its referent (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Markman, 1990; Medina, Snedeker,

Trueswell, & Gleitman, 2011). One example of this is the whole object bias: Children are more likely to map a novel word (e.g., COMPUTER) onto an object (e.g., the whole computer) than one of its parts (e.g., the keyboard). Importantly, these theories place the primary burden of learning on the initial encounter with a new word. The learning of a single word is assumed to be quick and one-shot. Evidence for this comes from fast-mapping: When children are presented with three objects, two of which they know (e.g., DOG, BALL) and one they do not, and are asked to pick up the GAVAGAI, they are able to use this set-up to map the novel word they are given onto the object without a name (Carey & Bartlett, 1978). When children quickly map a novel term onto an unknown object, they make use of the mutual-exclusivity-bias based on the assumption that each concept can be named by one word only. In this context, fast-mapping has often been conceptualized as a form of one-shot, all-or-none learning.

However, many alternative theories of vocabulary acquisition argue that words are not learned in one trial or after one exposure, but rather that information is accumulated across learning situations (McMurray, 2007; McMurray et al., 2012; Medina et al., 2011; Smith & Yu, 2008; Yu & Smith, 2007). An important source of evidence for this comes from re-investigations of children's fast-mapping behavior: Horst and Samuelson (2008), for instance, found that 24-months-old children failed to retain newly fast-mapped terms after a short delay of five minutes. These data suggest that even though young children easily pick out the novel object within a typical fast-mapping experiment, one trial may not be enough to result in long-term learning (see also Bion, Borovsky, & Fernald, 2013; Vlach & Sandhofer, 2012).

This conclusion is also supported by work on the hippocampus. Data from patients with amnesia suggest that a well-developed medial temporal lobe (MTL) may be necessary to take full advantage of single trial learning opportunities (Bauer, 2005; Warren & Duff, 2014). Sharon,

Moscovitch, and Gilboa (2011) found successful retention after fast-mapping in patients with amnesia. They argue that patients were able to learn associations through fast-mapping despite hippocampal damage. However, a strict replication of Sharon et al.'s fast-mapping design (2011) with a unique group of amnesic patients failed to observe long-term learning (Smith, Urgolites, Hopkins, & Squire, 2014). Moreover, Warren and Duff (2014) taught patients with amnesia word-object-mappings under two encoding conditions: explicit encoding during which a word was paired with its referent (e.g., "This is a GAVAGAI") or fast-mapping (a two-alternative forced choice version of the classic paradigm described previously). Patients with amnesia were found to show the fast-mapping of novel terms without long-term retention—similarly to the behavior observed in children (Horst & Samuelson, 2008). In addition, Vargha-Khadem et al. (1997) found that three patients with congenital bilateral hippocampal pathology were able to acquire average vocabularies despite never having functioning hippocampi. Together, this suggests that one-shot learning (fast-mapping) requires a functional hippocampus, whereas long-term word learning does not (Warren & Duff, 2014). Consistent with this, the MTL is not well-developed until relatively late in childhood (Bauer, 2005; Overman, Pate, Moore, & Peuster, 1996). Thus, if a hippocampus is necessary to be successful in single trial word learning, it is unlikely to be the driver of early vocabulary acquisition. Instead, word learning may be better characterized by a process that relies on repeated exposures and integration over time, though more inferential mechanisms may exist on top (McMurray et al., 2012; Samuelson & McMurray, 2017).

Evidence for word learning as a slow process also comes from computational modeling. McMurray et al. (2012) developed a connectionist model in which networks learned to map word forms onto visual referents. In their model, gradual associative learning was quasi-independent

of real-time processes such as the ability of the network to identify the referent of a word in-the-moment. Whereas the latter dynamic referent selection allowed the system to engage in more inferential processes in-the-moment, the co-occurrence of words and objects were tracked at the same time via building associations. This approach allowed the model to simulate a number of developmental phenomena, such as the ability to fast-map a novel word onto an unknown object, a sudden acceleration in vocabulary growth (also called vocabulary spurt or vocabulary explosion) or children's improvements in familiar word recognition with development. Importantly, even as dynamic competition is necessary to allow for in-the-moment processing, word learning at its core remains associative and gradual.

## 1.2 Cross-situational word learning

One novel way to investigate the associative component of word learning is to put people in a context in which every trial by itself is ambiguous, but across multiple situations, sufficient statistical information is included to support word learning (Siskind, 1996; Yu & Smith, 2007). This stands in contrast to the previously described, more traditional approaches to word learning, where participants are only given a small number of non-ambiguous trials.

In a typical cross-situational word learning experiment, participants see several novel objects and hear one novel word. They then select the object they believe maps onto the word. As the novel word is consistently paired with one target object (they co-occur on 100% trials), participants can use these co-occurrence statistics across trials to extract the correct word-object-mappings. Foil objects are randomly selected from a pool of objects, which in turn are the target for other novel words. Data from numerous studies have shown that people as well as children as young as infants can learn words under these circumstances (Dautriche & Chemla, 2014; Fitneva

& Christiansen, 2015; Roembke & McMurray, 2016; Roembke, Wiggs, & McMurray, 2018; Smith & Yu, 2008; Trueswell, Medina, Hafri, & Gleitman, 2013; Yu & Smith, 2007).

One critical prediction of an associative account is that when a word is paired with an object, associations between the two are strengthened; thus, over time, the associative strength between an object and a word should increase if they are paired consistently. Thus, a word and object can be associated by low, sub-threshold associations, even as this may not be evident in people's overt behavior. Moreover, the increase of associative strength between a word and an object is not limited to one mapping: Associations between one word and multiple objects may be built during initial acquisition. For instance, a young learner may form an association between the word FORK and the object fork, but may also possess a weaker association between the word FORK and the object spoon due to their frequent co-occurrence during meals. Thus, the associative strength between a word and a number of semantic concepts is not static, but rather may be sensitive to the statistical surroundings of a word.

Yurovsky, Fricker, Yu, and Smith (2014) investigated whether there is evidence for gradual, partial accumulation of knowledge in word learning, as predicted by an associative account of learning. To do so, they first trained adult participants on a small number of word-object-mappings within a cross-situational word learning paradigm. Subsequently, they identified the words that had not been acquired, and these words received additional training in a second learning session. The second phase also included a set of new word to use as a comparison. By doing so, they tested whether participants learned any information about words (e.g., statistical co-occurrence), even as they showed no evidence of having learned them in the first training session (they selected incorrect referents for them). Yurovsky et al. (2014) found that words that had been newly introduced in the second phase were more easily acquired if they

were paired with non-learned words of the first phase. Thus, despite the absence of any measureable learning by the end of the first training phase, participants must have retained some partial knowledge about the original words. Yurovsky et al.'s (2014) data are support of the notion that people maintain subtle associations between words and objects, even if these connections do not necessarily result in readily observable accuracy benefits.

One way to measure these sub-threshold associations more directly is to track participants' eye movements as they complete a word learning task in a version of the visual world paradigm (VWP; Magnuson, Tanenhaus, Aslin, & Dahan, 2003): Even as a person clicks on one object, they might be more likely to look at a competitor if it is associated with the word heard. This captures partial activation for latent associations. Using eye-tracking in the VWP, Roembke and McMurray (2016) trained participants on eight word-object-mappings in a cross-situational word learning experiment. Each word was always paired with its target object across trials, and foil objects were pseudo-randomized. However, at the same time, each word was also paired with a high-co-occurrence (HC) competitor on 60% of trials as well as a low-co-occurrence (LC) competitor. Unrelated foil objects co-occurred with a word at approximately 20%. Over the course of training, participants were more likely to look at the HC competitor than a randomly chosen foil. Importantly, this was even the case as they simultaneously clicked the correct object, indicating participants' maintenance of multiple word-object-mappings at the same time. Moreover, limited evidence from that study suggested that the relative amount of looks to the HC competitor increased with time (Roembke & McMurray, 2016); this suggests that these secondary associations are enhanced as learners encounter the words-object-pairs.

Evidence for an associative mechanism of word learning also comes from trial-by-trial-analyses. Trial-by-trial-analyses were first used by Medina et al. (2011) and later extended by

Trueswell et al. (2013). In these analyses, accuracy on a current trial is predicted by considering trial characteristics the last time the same novel word was heard. Thus, these analyses are a form of autocorrelation. Originally, trial-by-trial-analyses were used as evidence in favor of more propositional, one-shot learning theories. Accuracy on the most recent same-word-trial was considered particularly insightful (Medina et al., 2011; Trueswell et al., 2013): If a participant clicked on the correct object the last time they encountered the word, they should also be more likely to select it on a current trial. However, if they made an incorrect selection on the most recent trial with the same word, they should be at chance. This is consistent with the notion that participants only maintain one word-object-mapping at the time; this mapping is then verified on later exposures ("propose-but-verify" account). Both Medina et al. (2011) and Trueswell et al. (2013) found that participants were at chance on a current trial, if they had selected an incorrect object on a previous trial, but that accuracy was above chance after a correct last-trial choice.

However, both studies only included a small number of trials, thus potentially leading to an under-appreciation of more gradual processes (Roembke & McMurray, 2016; Wasserman, Brooks, & McMurray, 2015). Since then, studies by Dautriche and Chemla (2014) and Roembke and McMurray (2016) have used similar trial-by-trial-analyses to accumulate more evidence that (cross-situational) word learning is a gradual, associative process. Roembke and McMurray (2016), for instance, investigated the importance of associative learning by adding how often participants had encountered a particular word over and above the contributions of last-trial-accuracy (target count; this was part of the previously described co-occurrence manipulation and eye-tracking experiment). Thus, target count was a straightforward measure of the accumulated statistical co-occurrence evidence for a mapping: According to an associative account, how often a word had been experienced should predict accuracy independently of whether one had selected

the correct object on a previous trial or not. Roembke and McMurray (2016) found that this was indeed the case: Participants' accuracy was predicted by both last-trial-accuracy as well as trial count. In addition, these two factors interacted, and further inspection of the interaction suggested that the accuracy benefit of having been correct on a previous trial is the result—not the driver—of learning. These trial-by-trial analyses have also been replicated in 6-to-8-year-old children (Roembke, Wiggs, et al., 2018).

While there is still debate around the exact nature of the underlying mechanism of cross-situational word learning (Berens, Horst, & Bird, 2018), this abundance of data from converging paradigms suggests that associative learning is at least partially the core of cross-situational word learning. This then has implications for the type of network one may expect: During learning, as correct associations are built over time, it may be unavoidable that spurious associations occur as well.

### 1.3 Negative associative learning

Whereas there is ample evidence that multiple word-object-mappings are built and maintained during vocabulary acquisition, one missing piece of the puzzle is how incorrect associations between words and objects are minimized during learning. Associative learning has been hypothesized to include both positive (the strengthening of correct associations) and negative (the pruning of incorrect associations) forms (Hearst, Besley, & Farthing, 1970; Rescorla & Wagner, 1972; Spence, 1937; Thorndike, 1898). Nevertheless, most research has been concerned with how correct associations are built over time (Mackintosh, 1974; Newport, Wallis, Temple, & Siebeck, 2013; Stout & Miller, 2007), leaving unclear how critical the elimination or pruning of incorrect associations is for learning. Importantly, many studies of

associative learning are not complex enough to capture the dynamics of many-to-many-mappings in word learning. Indeed, converging evidence from a number of domains suggests an important role for pruning over development, although it is unclear if this relates to many-to-many learning.

### 1.3.1 *Pruning in synaptic/cellular development*

Pruning on the neural level is well-documented: Early cellular development is characterized by the rapid formation of synapses during prenatal and infant development. Subsequently, unused synapses are eliminated (e.g., Changeux & Danchin, 1976; Huttenlocher, 1979; Huttenlocher & Dabholkar, 1997; Rakic, Bourgeois, Eckenhoff, Zecevic, & Goldman-Rakic, 1986). One example of synaptic pruning is the formation of ocular dominance columns: Connections between the retinas and the visual cortex are pruned, resulting in columns of cells that respond to one eye only. This process is supported by decorrelated signals from each eye (Weliky & Katz, 1997), which emerge as a result of exposure to visual input and/or spontaneous neural activity (Shatz, 2002). Thus, the synaptic pruning of incorrect connections between the retinas and the visual cortex—in this case a nonspecific, though experience-expectant process (Greenough, Black, & Wallace, 1987)—is critical for the development of ocular dominance columns. These data, though clearly from different research area than word learning, underline the scientific reality of pruning as an important process in building a functional representational space at a cortical or neural level.

### 1.3.2 *Pruning in theories of development*

Negative associative learning has also been suggested to be of theoretical importance in theories of development (Siegler, 1989). For example, in the development of face and speech perception, it has been found that infants are first able to distinguish a broader range of features, but that ability decreases as they are exposed to the relevant face/speech features (Pascalis, De Haan, & Nelson, 2002; Werker & Tees, 1984), thus losing or pruning not needed distinctions.

Pruning is also theorized to be important in motor development: For example, the development of complex motor behavior is characterized by multi-joint twitching during sleep. Some components of these twitches have been found to be gradually pruned, thus resulting in the elimination of less organized patterns (Blumberg, Coleman, Gerth, & McMurray, 2013). In addition, during the development of mathematical cognition, children learn to prune less efficient mathematical strategies (e.g., for counting or adding; Siegler, 1999; Siegler & Opfer, 2003). It is not clear what factors contribute to this process or whether its underlying mechanism is similar to synaptic pruning; nevertheless, this suggests that negative associative learning is a mechanism that is at work in many different areas of learning.

### 1.3.3 *Pruning in connectionist/neural modeling*

Similarly, pruning is an essential theoretical prerequisite of most connectionist models: Learning processes are modeled by changing the associative strength between input, hidden and output layers based on statistical co-occurrences. This allows neural networks to investigate and mimic a wide range of learning phenomena in humans (e.g., Rumelhart, Hinton, & Mcclelland, 1986). Typically, association weights in these models start out at non-zero levels (small random

values), requiring the reduction of incorrect associations subsequently (e.g., Rumelhart et al., 1986).

In general, this aspect of neural/connectionist networks has received little theoretical consideration, but was rather a side product of the computational methods employed. However, the previously described model of word learning by McMurray et al. (2012) suggests that the pruning of incorrect associations may play an important role in the acquisition of word-object-mappings. For instance, how quickly a word was activated was more dependent on how well-pruned incorrect associations were, in contrast to how strong the connection between the correct object and word was. Similarly, the model's ability to map a novel term onto a novel object (fast-mapping) was dependent on which spurious connections between objects and words had been pruned or not (McMurray et al., 2012).

## 1.4 Pruning in many-to-many learning paradigms

In word learning, the focus of this dissertation, support for negative associative learning comes from the previously described connectionist model by McMurray et al. (2012), showing that how pruned incorrect associations are may be critical in predicting word learning phenomena (e.g., fast-mapping, rapid word recognition). Importantly, these modeling data indicate the possibility that word-object-mappings are pruned based on one's experience, and that negative associative learning is critical for forming a functional network of words.

However, direct empirical evidence for this is missing. Lab-based studies of word learning are often limited to teaching smaller vocabularies of pseudo-words to adults. As a result, it can be difficult to isolate any purely associative components to learning because adult humans can engage higher-level, inferential mechanisms to accelerate learning. In contrast, word learning

experiments in young children often includes ostensive teaching of a small number of words. For example, a child might be presented with a single object that is repeatedly labeled (e.g., “Look, it’s a TOMA! That’s the TOMA”) in the absence of any object competitors (e.g., Ferguson, Havy, & Waxman, 2015; Namy, Campbell, & Tomasello, 2004; Woodward & Hoyne, 1999).

Thus, to investigate associative processes during vocabulary acquisition in a set of words, Roembke, Wasserman, and Mcmurray (2016) adapted an animal model of word learning for pigeons (developed by Wasserman, Brooks, and McMurray, 2015). In this task, pigeons were trained on 16 symbol-object-mappings: A visual symbol (a so-called pexigram) was paired with one object. Pigeons were trained on these mappings in a supervised learning paradigm; on each trial, one object was presented with two pexigrams, the target and a randomly selected foil (the

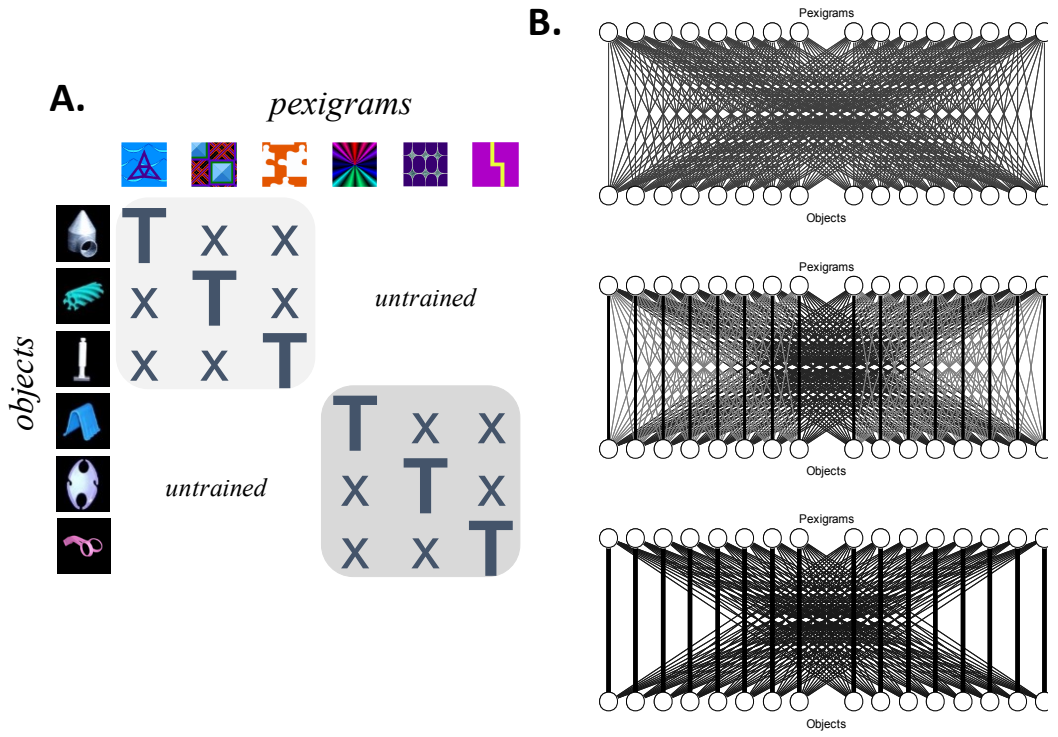


Figure 1: A) Overview of the design used in Roembke et al. (2016). T refers to target (the pexigram that is correct for that object) and X refers to foil (a pexigram that is incorrect for that object). B) Predicted associative matrix over the course ass assuming that pruning plays a role in learning. This Figure is adapted from Roembke et al. (2016).

incorrect response). Pigeons received food if they pecked the correct response choice. If they selected the foil, they had to repeat the trial until a correct response was made.

During training trials, mappings were separated into two clusters (cluster A and B), so that the foil (the incorrect response) for a trial was selected from the items in the same cluster as the target object (see Figure 1 for clarification of the design). This experimental manipulation guaranteed that associations between the object and symbols of the opposite cluster could not get pruned: They never co-occurred, so the birds never had the opportunity to select one of these incorrect pexigrams and get negatively reinforced after pigeons reached a learning asymptote at around 80% correct, they were given testing trials (mixed with training trials), in which the foil pexigram was selected from the opposite cluster. On these trials, a food reward was provided independently of whether the correct choice was selected or not. Pigeons' accuracy was found to be lower on testing trials. If learning these mappings had only depended on how strong the association between a target pexigram and an object were, pigeons should have performed equally well in both training and testing trials.

In addition, pigeons were tested on trials in which the target pexigram was not present (so-called no-correct testing trials). Here, the responses consisted of a within-cluster (pruned) pexigram and an out-of-cluster (unpruned) pexigram. In no-correct testing trials, pigeons were more likely to select the untrained/out-of-cluster foil than the trained/in-cluster one. If incorrect associations had not been pruned, pigeons should have selected the within-cluster pexigram, which had co-occurred with the object during training.

These findings indicate that pigeons prune incorrect associations as they acquire symbol-object-mappings, and that this process affects in-the-moment performance. Importantly, the pruning of incorrect associations is experience-dependent, and happens in response to specific

content (the association between target objects and foil pexigrams). This stands in contrast to broader neural pruning where unused neurons are eliminated after previous overproduction.

Even though these results are suggestive that negative associative learning is instrumental in the acquisition of symbol-object-mappings, there are clear differences between human word learning and the animal model. As argued by Wasserman et al. (2015), using pigeons allows for the isolation of associative learning, as pigeons lack higher-level executive processes. Thus, at this point, it is unclear how critical pruning is in a more complex, cognitively developed species such as humans. For instance, it is possible that humans can use executive functioning or strategies to quickly overcome spurious associations that may have formed; even though this process would also include pruning, the mechanism at work may be very different.

In addition, there are other clear differences between word learning and pigeons' learning: People use language and words every day, whereas pigeons do not require the acquisition of symbol-object-mappings, and therefore do not have a network (lexicon) into which newly learned words can be integrated into. The competition between newly learned and existing words, however, may lead to higher levels of spurious associations, thus potentially leading to an *increase* in the need for pruning in humans.

Moreover, pigeons' acquisition of symbol-object-mappings is more cumbersome than in people: Pigeons required months to learn the mappings (Roembke, Wasserman, & Mcmurray, 2016; Wasserman, Brooks, & McMurray, 2015), instead of the typical one hour length of an adult word learning experiment (e.g., Roembke & McMurray, 2016). Because of this, pigeons slept between learning sessions, a process that has been shown to be central to memory consolidation (e.g., Stickgold, 2005) and thus could be meaningful for pruning and/or strengthening associations. The latter is consistent with data indicating that (word) learning is

supported by sleep (Borquez, Born, Navarro, Betancourt, & Inostroza, 2014; Dumay & Gaskell, 2012, 2007; Henderson, Powell, Gaskell, & Norbury, 2014; McGregor et al., 2013); this will be discussed in more detail as part of Chapter 6. Alternatively, it is possible that high levels of difficulty (not sleep per se) result in a need for pruning, a question that cannot be easily disentangled in a non-human species (most non-human learners will always require at least multiple days to acquire a small network of symbol-object-mappings).

Importantly, it should be underlined that all supervised experiments also include unsupervised statistics that could alter learning. Even as human participants receive feedback as to the correctness of their response, they may still form associations between the words they hear and the objects they encounter. In addition, the fact that two items do not appear together might also carry important information, even if no feedback is given that the two do not map onto each other. Indeed, this is the essence of unsupervised cross-situational learning. However, pigeons are not able to acquire the symbol-object-mappings in the absence of feedback. Thus, it is currently unclear how supervised and unsupervised statistics interact during word learning and unlearning.

Finally, the pigeon study by Roembke et al. (2016) relied on the assumption that there are low levels of associations between all words and objects before training that subsequently are pruned. In pigeons, these may potentially arise from the fact that pigeons were been exposed to all stimuli during a pre-training in isolation. These baseline levels of associations may be too subtle to be revealed in human word learning in which people may employ inference and executive processes to resolve competition between foils (rather than just relying on associations). Thus, it may be possible that the pruning of incorrect associations in humans would not be revealed by designs such as that of Roembke et al. (2016). Instead pruning may be

more adequately described as unlearning: More specifically, pruning may be necessary after a word was mapped onto the wrong concept after misleading evidence (and must later be corrected), not because of initial spurious associations across the lexicon.

## 1.5 Significance and remaining questions

To summarize the previous section, negative associative learning is important in theories of learning and development (Hearst et al., 1970; McMurray et al., 2012; Rescorla & Wagner, 1972; Siegler, 1989, 1999; Spence, 1937; Thorndike, 1898). Despite this, there is little behavioral evidence that it matters in the learning of more complex many-to-many mappings, such as would be needed in categorization, face perception or word learning (in humans). Thus, the overall goal of this dissertation is to investigate whether the pruning of incorrect association is involved in human word learning.

These findings may have implications for wide ranging research fields. For our understanding of basic theories of associative learning, it is critical to investigate whether the experience-dependent pruning of incorrect associations is also evident in human learning, and not limited to animals with less advanced higher-level functioning. In this context, questions that are addressed in this dissertation proposal will significantly inform our currently limited understanding of under what circumstances negative associative learning takes place.

Additionally, the results of the proposed experiments are also significant for theories of word learning and reading. Despite increased evidence that associative processes are core to word learning, there is still a lack of understanding of how associations between words and objects change over time based on people's experience. Moreover, pruning as a functionally meaningful aspect of word learning and processing is predicted by McMurray et al.'s

computational model (2012); thus, behavioral evidence in favor of pruning will be of theoretical importance to validate this model and theories of word learning more generally. Similarly, investigating how pruning may participate in how novel written forms are added to the lexicon may increase our understanding of how people achieve automatic activation of words during reading.

This dissertation will address the following four questions:

1) Determine if connections between objects and words are not just strengthened but also pruned.

2) Determine if pruning differs depending on the type of information that is acquired.

3) Determine if sub-threshold incorrect associations influence subsequent word learning.

4) Investigate how quickly newly acquired word-object-mappings are activated relative to known words. This aim is exploratory, and will be introduced in detail as part of Chapter 6.

These questions will be introduced in the sections below along with a more specific review of the literature relevant to each one.

### 1.5.1 *Incorrect associations during learning*

As previously pointed out, the best empirical evidence for the importance of pruning incorrect spurious and/or pre-existing associations comes from the pigeon study by Roembke et al. (2016). Importantly, even though the pigeons were more likely to select the unpruned competitor, it should be highlighted that their overall accuracy was above chance: Including unpruned foils resulted in a reduction of performance, but not a complete reversal. This is equivalent to Question 1 of the dissertation: Under what circumstances are spurious (subtle), pre-existing associations pruned?

This question is not just of theoretical merit. If a word is presented in a consistent setting, this may lead to the formation of incorrect, spurious associations (e.g., the words FORK and KNIFE). In fact, this is indeed what was shown by Roembke and McMurray (2016): People were sensitive to subtle co-occurrence statistics between words and referents, and activated competitors accordingly. To test how the presence of a high co-occurrence competitor influenced learning, a control experiment was conducted where all foil objects were equally likely to occur. It was found that learning was slower when words were paired with a more frequent competitor (in addition to their target object) than when there was no single more frequent competitor.

These findings are similar to the general notion that variability in irrelevant dimensions aids learning of the relevant one (e.g., Apfelbaum, Hazeltine, & McMurray, 2013; Apfelbaum & McMurray, 2011; Gómez, 2002; Lively, Logan, & Pisoni, 1993; Rost & McMurray, 2009, 2010): For example, Apfelbaum, Hazeltine and McMurray (2013) found that presenting vowels in variable consonant frames (e.g., MAID, RAIN, SAIL) but not in similar ones (e.g., MAIL, MAIN, SAIL) improves first graders' vowel reading. In both cases, presenting the relevant association (e.g., between the word and the target object) with irrelevant ones (e.g., the word and foil objects) helps isolate the correct association, and may allow for the elimination of incorrect connections.

Why should consistency in visual competitors (i.e., increased co-occurrence between words and objects lead to strengthening under some circumstances but pruning under others? One possibility is that the presence of feedback facilitated pruning: The experiment by Roembke and McMurray (2016) was unsupervised (no feedback). In contrast, pigeons were given feedback after each response selection (they are not able to acquire the symbol-object-mappings otherwise). Of course, it is also conceivable that cross-species differences between humans and

pigeons lead or add to dissimilarities in how associations are formed and maintained (or not) over time.

However, it has also been argued that context—operationalized as semantic relation among referents—would facilitate word learning. Richer context could provide additional structure, which in turn might facilitate word learning; learners may use what context a word was uttered in as extra information to constrain potential meanings of a word. For example, if a novel word was heard repeatedly at a zoo, a child may use higher order properties of that context (e.g., a zoo contains many animals) to deduce meaning, even as the term is encountered in a different context. In a study by Dautriche and Chemla (2014), participants acquired word-object-mappings in a cross-situational word learning paradigm. In the first learning instance, all referents either belonged to the same natural category (e.g., animals) or a learned category. Subsequently, competitor and target objects were randomly mixed (as in typical cross-situational word learning experiments). They found better performance when the first trial in which a novel word was encountered provided constant visual competitors. Importantly, this result is more consistent with the notion that repeated exposure with the same objects might have led to the decrease in their activation (pruning) and thus in better learning (as observed in Roembke and McMurray (2016) when no high-co-occurrence competitors were included).

Why did constant visual competitors facilitate learning in some human adults (Dautriche & Chemla, 2014) but not others (Roembke & McMurray, 2016) under similar learning conditions (unsupervised cross-situational word learning experiment? Importantly, the context manipulation in Dautriche and Chemla (2014) was relatively limited: Even though semantic relations between objects (visual referents) were established in the beginning of the experiment, participants also experienced trials in which competitors were randomly mixed. This might have

resulted in the advantages of consistent competitors (i.e., establishing a smaller set of possible target referents) without the disadvantages (i.e., creating spurious associations between words and competitors due to increased co-occurrence). In contrast, high co-occurrence competitors (60% co-occurrence with words across experiments) in Roembke and McMurray's (2016) might have resulted in too many spurious associations to take advantage of a semantic relations/context effect. Consistent context (or, the alternative, variability in competitors) matters, and may be a factor that could be easily manipulated to facilitate word learning in children or during second language acquisition. Thus, one goal of this dissertation is to investigate under what circumstances associations are strengthened and/or pruned during word learning (Question 1).

One factor that might influence how easy it is to prune incorrect associations could be the type of information that needs to be associated; this will be explored as part of Question 2. More specifically, I will investigate if negative associative learning occurs when learning to map written words to referents, or when mapping arbitrary visual symbols to referents. One possibility is that the main difference in how pruning differs across domains (e.g., oral and written word learning) is that people are better at acquiring some mappings than others: A recent study by Roembke et al. (2018) found that adult participants were better at learning word-object-mappings than sound-object-mappings, where sounds were non-linguistic, non-environmental beeps. Thus, it is possible that the higher learning rate for words was the result of or the cause of better negative associative learning.

### 1.5.2 *Unlearning incorrect associations*

Another way to conceptualize the process of getting rid of previously established incorrect associations is unlearning. Unlearning, in comparison to pruning, might be less subtle,

as it could refer to not just eliminating sub-threshold or partial information but also fully established mappings. This is relevant, as children sometimes map a word onto an incorrect object. Naming errors are typically conceptualized as overgeneralizations (Clark, 1973) or retrieval errors (e.g., Huttenlocher, 1974). Overgeneralization errors are typically similarity-based, and could be caused by a faulty hypothesis for the meaning of a word and/or limitations in children's existing vocabulary. Retrieval errors are considered mistakes that are made despite knowing the correct word for a concept, due to momentary difficulties in recalling it. Famous examples of naming errors are so-called overextension and underextension errors, where a concept is mapped onto a broader category than correct (e.g., the word BALL refers to all round objects, including the moon) or a more limited category than correct (e.g., the word BALL refers only to a soccer ball but not others; Gershkoff-Stowe, 2001), respectively.

Naming errors are most frequent in children who have between 50 and 150 words in their lexica, a period of rapid vocabulary growth (Gershkoff-Stowe, 2001), and they tend to be less frequent before and after (though not absent). This developmental trajectory suggests that naming errors may be most adequately described as retrieval errors, where a mapping between a word and a concept is established but may not be strong enough to outweigh other existing associations or recently activated words (Gershkoff-Stowe, 2001).

Such understanding of retrieval errors is based on the assumption that mappings between words and objects are always one-to-one. As a result, if a naming error occurs even though the correct term has been used before, this must be the consequence of a processing error. However, if several latent associations between words and several meanings exist, naming errors might in fact reflect spurious associations between word-object-mappings. This suggests that a critical part of word learning is "unlearning" or pruning these connections to refine the lexical network.

Unlearning is clearly a part of word learning and likely also happens in adults (though may not be as frequent and/or obvious). Under what circumstances are incorrect associations unlearned? How functional are small sub-threshold associations in influencing participants' acquisition of word-object-mappings? This will be investigated as part of Question 3, and introduced in more detail in Chapter 5.

### 1.5.3 *Incorrect associations and word processing*

Finally, incorrect associations might also remain after word-object-mappings have been established; this could influence how efficiently words are processed in-the-moment (McMurray et al., 2012). Speed of word recognition increases with development: In a study by Fernald, Pinto, Swingley, Weinbergy, and McRoberts (1998), children were found to make rapid gains in the efficiency of word recognition over their second year of life. Using an eye-tracking paradigm, they found that 15-month-old infants did not look to the correct picture until after the target word was spoken. However, 24-month-olds were able to make eye movements to the target before the end of the word was uttered. Finally, two-year-olds' eye movements were even more similar to the ones observed in adults, activating words as soon as possible (even if no sufficient information is available to predict the identity of the spoken word with certainty).

These findings can be captured by the connectionist model of word learning by McMurray et al. (2012). Moreover, this model suggests that the speed of word recognition is better predicted by whether incorrect associations have been pruned than by the strength of the correct associations. These simulations suggest that negative associative learning may not just influence the initial acquisition of word-object-mappings, but may also influence in-the-moment processing. Arguably, the presence of incorrect associations during in-the-moment processing

may interact with how information is acquired (c.f., McMurray et al., 2012). In fact, some data suggest a clear predictive relationship between the two abilities: First, Fernald, Perfors, and Marchman (2006) found that speed of word recognition at 25 months predicted lexical and grammatical development over the coming year. Second, in a follow-up study with the same cohort of children, Marchman and Fernald (2008) discovered that word recognition skills at approximately two years of age was also related to cognitive and language outcomes at eight years.

Even though the presence or absence of incorrect associations after a word has been acquired is an important possibility in how small incorrect associations may matter in word learning, it should be highlighted that it was not the main goal of this investigation to examine this question. Nevertheless, some exploratory analyses will be completed as part of Question 4.

# CHAPTER 2: EXPERIMENT 1

## 2.1 Question 1: Overview

The experiments in this chapter address my first question: whether connections between words and objects are not just strengthened during learning, but also pruned. For this purpose, I conducted three experiments (Experiments 1-3). All three were highly inspired by Roembke et al.'s (2016) work in pigeons. More specifically, these experiments used the same clustering design, where some foils never co-occurred with targets during training to prevent pruning of incorrect associations. As the experiments in pigeons, participants received feedback during training, thus allowing for: (1) The quicker acquisition of the word-objects-mappings (than would be possible using purely unsupervised cross-situational learning). This is a necessary condition for the eye-tracking design used. (2) An explicit error signal whenever an incorrect foil was selected. This can be used for pruning this incorrect association.

In preparation for investigating pruning mechanisms in human word learning, I completed numerous pilot studies based on adaptions of this design. Iterations included designs with both accuracy and eye movements to the trained and untrained foils as dependent measure(s). These pilots found little evidence for a role of negative associative learning in this type of vocabulary acquisition; that is, there was no evidence for a difference in performance when foils were from the target's cluster or the opposite one.

If we take these null effects at face value, what are we to make of the differences between humans and pigeons? One possibility is that the pigeon paradigm taps pre-existing, small associations between words and objects (similar to the initial state of a connectionist model). If so, then the human work would suggest that these associations are too subtle to result in measurable performance reductions. Alternatively, it should be noted that pigeons are trained on

symbols and objects in isolation before their training on the symbol-object-mappings started. This potentially allowed for the formation of associations between symbols and objects before the onset of training. In this case, the pruning results in pigeons is a form of unlearning in which specific associations learned during pre-training are later suppressed during the operant procedure.

The experiments in this chapter thus adopted a new approach. First, adult participants were deliberately trained on a mapping between the word and object in isolation in a pre-training phase (similar to Vouloumanos, 2008). However, this pre-training also trained participants on one spurious association. During this phase, all trials were unambiguous: On each trial, only a single object and a novel word were presented. However, across trials, each word was presented with its target on six trials and a featured competitor on four trials.[1] Thus, by the end of training, two associations were established for a given word, one slightly stronger than the other. Previous research has shown that people are sensitive to slight differences in co-occurrence (Vouloumanos, 2008).

After this initial pre-training, participants completed a more traditional supervised learning paradigm. On each trial, they heard one word with four potential objects on the screen (the target and three foils). On some trials, one of these foils was the featured competitor that had been reinforced during pre-training. Participants then responded, and received feedback based on the object they selected.

[1] The exact number of how often each word co-occurred with its target and its featured competitor differed across experiments.

During training, targets and foils were clustered as in Roembke et al. (2016). The result of this was that some foils received opportunities for pruning whereas others did not. However, crucially, this interacted with the featured competitors in systematic ways. For half of the words, the featured competitor belonged to the same cluster during training (e.g., object 2 for word 1, see Table 3); it is therefore termed *in-featured*. It was hypothesized that for *in-featured* competitors, participants would have the opportunity to prune incorrect associations that were built during pre-training, since these items would appear as foils for the target and would consequently have the opportunity for negative reinforcement.

Table 1: Overview of clustering design. Please note that this overview only includes eight word-object-mappings (instead of 16) to facilitate presentation. "T" indicates the target object, "F" indicates the featured competitor.

| | | Word | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | 1 | T | F | | | | | | |
| | 2 | F | T | | | | | | |
| | 3 | | | T | | F | | | |
| | 4 | | | | T | | F | | |
| Object | 5 | | | F | | T | | | |
| | 6 | | | | F | | T | | |
| | 7 | | | | | | | T | F |
| | 8 | | | | | | | F | T |

In contrast, for the other half of the words, the featured competitor was assigned to the opposite cluster (e.g., object 5 for word 3, see Table 1). This foil is referred to as the *out-featured* competitor. For an *out-featured* object, it was predicted that people would *not* have the opportunity to prune incorrect associations with the word, as they never co-occurred during training.

To measure the extent of pruning during training, participants' eye movements were tracked as a more sensitive measure than accuracy. More specifically, as a variant of the visual world paradigm, eye movements to foil objects were compared on trials that participants had selected the target object (Magnuson, Tanenhaus, et al., 2003; Roembke & McMurray, 2016). It was predicted that people's looks to the *in-featured* competitor would be higher than to baseline foils in the beginning of training (since these had been reinforced during pre-training). However, this would decrease over the course of training as these were pruned. For the *out-featured* competitors, no training prediction was made, as they never co-occurred with the target.

Subsequently, participants completed a testing session without feedback. As in training, participants heard one word and saw four objects on each trial. Again their eye movements were tracked. Here, foils could be both selected from within the cluster or not. Importantly, test trials always consisted of two within-cluster objects (one of which was the target) and two out-cluster objects (which had not been seen with the target before). This pairing of two and two was used to counteract biases to select one type of foil over the other by cueing participants to the identity of the target before the word was heard (for example, if three of the foils had been seen with the target and one had not).

Of experimental interest were the test trials that included either the *in-featured* or the *out-featured* competitor. Here, I used the difference in looking between the featured competitors and the baseline foils as a measure of how strongly they were still associated. For *in-featured* test trials, it was predicted that looks should be equal to the *in-featured* competitor and baseline foils, as the increased association between the featured competitor and word should be pruned. For *out-featured* test trials, it was hypothesized that looks to the *out-featured* competitor would be

higher than to baseline foils, as the established spurious association between the *out-featured* competitor and the word had no opportunity to be pruned during training.

Finally, an additional short test session was conducted, in which the target object was not included as an option (no-correct testing trials; Roembke et al., 2016). On these trials, participants were instructed to just "make their best guess". Half of the trials included a word's featured competitor. I used these to ask if this featured competitor was selected more often than baseline foils. Half the words had an *in-featured* and half had an *out-featured* competitor. Across trials, I investigated if participants were more likely to select an unpruned competitor over a specific foil (as in the pigeons). Here, I predicted they would select the *out-featured* competitor at greater than chance levels, but the *in-featured* would be selected at chance level. It was hypothesized that *out-featured* competitors should be more likely to be selected as targets than *in-featured* ones in the no-correct trials because (1) they have a history of high associative strength with the word, and (2) there is no negative evidence that they are not the correct object (as they never co-occurred with the word during training).

To summarize, these experiments tested whether subtle, pre-existing associations between words and objects are pruned during word learning. For this purpose, incorrect associations were built during a pre-training. Subsequently, participants were trained on the correct word-object-mappings. Eye-tracking was used to assess participants' retention of the incorrect associations. Experiment 1 was conducted over two days, which was done as an exploratory approach to determining whether sleep is critical for pruning or not. Experiment 2 and 3 are closely matched one-day studies.

## 2.2 Method

### 2.2.1 *Participants*

Participants were forty monolingual, native English students at the University of Iowa with normal or corrected-to-normal vision. Thirty-five received course credit as compensation. Five participants received $37.5 (2.5-hour experiment, $15/ hour). Participants underwent informed consent in accord with an IRB approved protocol. For one participant, it was impossible to get an adequate eye-track on day 2; thus, eye movements were not tracked and their data are excluded from analysis (accuracy data were still collected and analyzed in the no-correct testing session).

Table 2: List of novel words used in Experiment 1.

| Written word form of nonword | IPA of nonword |
|---|---|
| Bure | /buɹeɪ/ |
| Dimu | /dImu: / |
| Fatei | /fæteɪ/ |
| Goba | /goubɑ/ |
| Haito | /haɪto/ |
| Jifei | /dʒifeɪ/ |
| Kepoi | /kepɔɪ/ |
| Lubo | /lubo/ |
| Mefa | /meɪfɑ/ |
| Naida | /nɑɪdɑ/ |
| Pacho | /pætʃo/ |
| Razi | /ræzi/ |
| Sheku | /ʃeku/ |
| Sipa | /sIpɑ/ |
| Tagu | /tægu/ |
| Zati | /zæti/ |

### 2.2.2 *Stimuli*

Sixteen novel objects were used as referents. Referents were novel objects, presented on a white background. Sixteen nonwords (Table 2) were constructed; these will be referred to as words (as this paradigm is meant to simulate the learning of real words). Words were two-syllable, CVCV pseudo words, which were phonologically legal words in English. Phonological overlap at onset was minimized to guarantee that looks to foil objects were not driven by similarity between the word forms.

Auditory stimuli were recorded by a female, native speaker of English in a neutral carrier phrase (e.g., “He said...”). Subsequently, five exemplars of each stimulus were selected and edited to remove extraneous elements that were not part of the stimulus (e.g., jaw clicks).

Finally, auditory stimuli were normalized and 50 msec of silence was added to their beginning and end.

### 2.2.3 *Design*

Table 3: Overview of Experiment 1's design.

| Featured Type | Training Cluster | Word | Target Object | Featured Object |
|---|---|---|---|---|
| In-featured/ Pruned | A | 1 | 1 | 2 |
| | A | 2 | 2 | 1 |
| | A | 3 | 3 | 4 |
| | A | 4 | 4 | 3 |
| | B | 5 | 5 | 6 |
| | B | 6 | 6 | 5 |
| | B | 7 | 7 | 8 |
| | B | 8 | 8 | 7 |
| Out-featured/ Unpruned | A | 9 | 9 | 10 |
| | B | 10 | 10 | 9 |
| | A | 11 | 11 | 12 |
| | B | 12 | 12 | 11 |
| | A | 13 | 13 | 14 |
| | B | 14 | 14 | 13 |
| | A | 15 | 15 | 16 |
| | B | 16 | 16 | 15 |

Participants learned 16 word-object mappings. The mapping between each word and its referent object was randomized for each participant. In addition, each word was assigned a featured competitor by creating eight word pairs and selecting the other's word target object as the featured competitor. For instance, word 1's target object was object 1 and their featured target was object 2, whereas word 2's target object was object 2 and their featured one was object 1 (see Table 3).

Two clusters were created pseudo-randomly for each participant: Half of the words were assigned to cluster A, and the other half was assigned to cluster B (see Table 3). In doing so, half of the featured word pairs (e.g., word 1 and word 2) was assigned to the same cluster, whereas the other half was assigned to different clusters. As a result, for half of the words, the featured competitor was in the same cluster as the target object (the *in-featured* competitor). For example,

in Table 3, the word pair word 1 and 2 are both in cluster A. As a result, word 1's featured competitor (object 2), word 2's target object, is also in that cluster. For the other half of the words, the featured object was in the other cluster (the *out-featured* competitor). For example, word 9 and 10 are a word pair (word 9's target object is word 10's featured competitor, and vice versa). Word 9 is assigned to cluster A, whereas word 10 is in cluster B. As a result, word 9's featured competitor, object 10, is not in cluster A but cluster B.

The experiment was conducted over two subsequent days (see Figure 2 for an overview of Experiment 1's structure). Even though most participants were scheduled at the same time on both days, this was not possible for a small subset of participants. On day 1, participants completed a pre-training, three training blocks and one testing session (which took up to 1.5 hrs in total). On day 2, they completed one testing session, two training blocks, a second testing session and a no-correct session (adding up to 1hr).

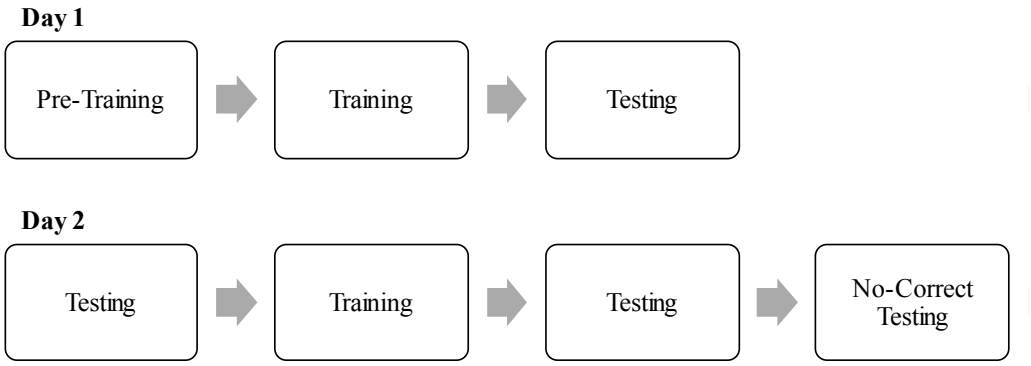


Figure 2: Overview of the basic structure of Experiment 1.

During pre-training, each word was paired with a single object on a trial. Across trials, words were either paired with their assigned target (6/10 trials/word) or their featured competitor (4/10 trials/word) to build spurious (incorrect) associations with the latter. Pre-training consisted of 160 trials (10/word). The word was presented in the center of the screen, offset to the top.

Training on day 1 consisted of three blocks of 112 trials, resulting in an overall number of 336 trials. Each trial during training included four objects (presented in the four corners of the screen), one of which was always the target object (see Figure 3 for a schematic). Foils were randomly selected (without replacement) from the cluster the word had been assigned to. More specifically, foil objects for one word were the target objects of other words in the same cluster. Thus, for words with an *in-featured* competitor, the *in-featured* object was present as a foil on 9/21 training trials per word. For words with an *out-featured* competitor, the *out-featured* object never co-occurred with its target word during training. Importantly, all words and all objects were trained on their own referent. Thus, while both *in-featured* and *out-featured* were controlled when appearing as foils, they also appeared as targets paired with a different word. The training session on day 2 was identically constructed as the one on day 1. However, it only included two blocks of 112 trials, resulting in a total number of 224 trials.

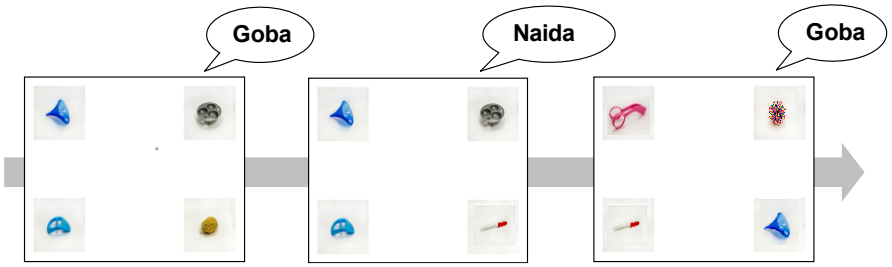

Figure 3: Depiction of training/testing trials in Experiments 1-3.

Each testing session consisted of one block of 160 trials (10 trials/ word). Testing trials consisted of several trial types. For all trial types, two objects were chosen from the same cluster as the word, and two were chosen from the other cluster. Thus, participants could not guess what word they would hear based on the distribution of response options (e.g., if three objects were from the same cluster, this might bias participants to select a response option from the over- or underrepresented cluster). *In-featured* testing trials included the target object and the *in-featured*

competitor and two baseline foils hat were randomly chosen from the other cluster (*out-foils*). *Out-featured* testing trials included the target object, the *out-featured* competitor and two baseline foils, one of which was from the same (*in-foil*) and one of which was from the other cluster (*out-foil*). Control testing trials were made up of the target object and three baseline foils (two *out-foil*s and one *in-foil*). There were three *in-featured* or *out-featured* testing trials per word during testing, and seven control testing trials per word. All foils (except for the featured object) were randomly chosen without replacement from the respective cluster.

Testing sessions were identically constructed across days. However, they were separately randomized, meaning that even though the same types of trials existed across the three testing sessions (one on day 1, two on day 2), trials were not made up of the same exact foil objects.

At the end of day 2, participants completed a no-correct testing session of 32 trials (2 trials/ word). During this session, trials never included the target object as a response option. Instead, four foil objects were presented as possible responses, two out of each cluster. For each word, one trial included the featured object. For trials including the *in-featured* object, baseline foils were one *in-foil* and two *out-foils*. For trials including the *out-featured* object, baseline foils were two *in-foils* and one *out-foil* to complete the set. The remaining trial for each word always included two baseline foils from each cluster. In the no-correct session, one of the two trials for each word was assigned to either the first (trial 1-16) or second half (trial 17-32). The order of trials was randomized within those halves. Participants did not receive feedback during the no-correct session.

Trials were always randomized within block, unless otherwise noted. The location of each object was randomized across the four possible locations on each trial (with the exception of pre-training where there was only one possible location). A summary of all trials of Experiment 1 is provided in Table 4.

Table 4: Design overview of Experiment 1.

| Day | Session | Trials | Number/Blocks | Trials/Word | Trials/ Word and Target | Trials/Word and Featured | |
|---|---|---|---|---|---|---|---|
| 1 | Pre-training | 160 | 1 | 10 | 6 | 4 | |
| | | | | | | In | Out |
| | Training | 336 | 3 | 21 | 21 | 9 | 0 |
| | Testing | 160 | 1 | 10 | 10 | 3 | 3 |
| 2 | Testing | 160 | 1 | 10 | 10 | 3 | 3 |
| | Training | 224 | 2 | 14 | 14 | 6 | 0 |
| | Testing | 160 | 1 | 10 | 10 | 3 | 3 |
| | No-correct | 32 | 1 | 2 | 0 | 1 | 1 |

2.2.4 *Procedure*

The experiment was conducted in a sound attenuated room. The display was a 19" monitor operating at 1280 × 1024 resolution, and sounds were presented via high quality headphones at a volume comfortable to the subject.

Participants were told that their task was to discover which object goes with what word, and that on each trial, they were to indicate their best guess by clicking on that picture. On day 1, participants were informed that they would first complete a pre-training with only one object presented, before advancing to a training session with feedback and a testing one without. Before starting training, participants heard the two types of sounds that would be used for feedback. On day 2, participants were instructed of the order of sessions. In addition, they were instructed

before completing the no-correct testing session that the upcoming trials would not include the target object and that they had to indicate their best guess. This instruction was added after running subject 1, as they showed confusion at the sudden absence of the target object during debriefing and always clicked the same location during the no-correct testing session, rendering the data meaningless.

During pre-training, participants were presented with a small blue circle at the center of the screen. In addition, one object offset to the top of the blue circle appeared. Participants were given 1050 msec to inspect the object. Afterwards, the circle turned red, cueing the participant to click on it to hear the word. When the participant clicked on it, the red circle disappeared and the target word was played. The auditory stimulus was randomly selected to be one of the five exemplars of the target stimulus (with replacement). To move on to the next trial, participants had to click on the presented object.

During training trials, participants were presented with the small blue circle and four objects in the corners of the screen (see Figure 3). As before, the blue circle turned red after 1050 msec, and the auditory stimulus was played. Participants then clicked on the picture corresponding to the word. Finally, they received feedback indicating if their response had been correct or incorrect. Feedback consisted of either a high tone (‘Bing!’), indicating a correct response, or a low tone (‘Mep!’), indicating incorrect ones. The display turned white when participants made a response and stayed so until the end of feedback; feedback was played after approximately 150 msec. Trials were never time-limited.

Testing and no-correct trials were identical to training ones with the exception that participants did not receive any feedback. At either the beginning or end of day 2, participants completed a questionnaire that asked about how much sleep they received the night before, how

much sleep they usually require to feel well-rested and whether they consumed alcohol the night before.

### 2.2.5 *Eye-tracking recording and analysis*

Eye movements were recorded during both days throughout the experiment using an SR Research Eyelink 1000 chinrest mount eye-tracker operating at 250 Hz. No eye-tracking was recorded during pre-training (as only one object was present during pre-training, eye movements were not meaningful). Both corneal reflection and pupil were used to obtain point of gaze whenever possible, though for some participants only good pupil readings could be obtained. At the start of the experiment, participants were calibrated with the standard 9-point display. Every 30 trials, a drift corrections procedure was conducted to check adequate calibration and account for any drift in the track. Fixations were automatically parsed into saccades and fixations using the default "psychophysical" parameter set. Adjacent saccades and fixations were combined into a single "look" (starting at the onset of the saccade and ending at the offset of the fixation as in prior studies; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Samelson, Lee, & Tomblin, 2010). To account for noise in the eye-tracking record, the ports containing the objects were extended by 100 pixels when computing the point of gaze. No overlap between the objects resulted from this.

## 2.3 Results

### 2.3.1 *Overview*

Results will be presented separated by day and loosely in the order in which participants completed each phase of the experiment. First, I analyzed participants' response selections on trials that included the featured competitor in block 1 of training. Thus, I tested if participants

had associated words with the featured competitor as a result of pre-training. It was predicted that participants would be more likely to select the featured competitor than a baseline foil. Subsequently, proportion of looks during featured trials of training were analyzed. This analysis tested whether participants activated the featured competitor more than a baseline foil, even as they clicked on the target object. It was predicted that looks to the featured competitor would be significantly higher than looks to baseline foils in the beginning of training, but that this "preference" would disappear as training proceeded.

Finally, I examined eye movements during testing trials using a similar approach. Testing, included both *in-featured* and *out-featured* testing trials. For *in-featured* testing trials, looks to the featured competitor were compared to the averaged looks to the baseline foils. It was predicted that participants would look equally to the *in-featured* competitor and baseline foils, as training allowed for the pruning of incorrect associations with the *in-featured* competitor. For *out-featured* testing trials, looks to the featured competitor were compared to a baseline foil from the same cluster and one from the different cluster in two contrasts. It was predicted that looks to the *out-featured* competitor were significantly higher than ones to either baseline foil, as participants had no opportunity to prune incorrect associations with the *out-featured* competitor during training.

For no-correct testing trials, it was predicted that participants would be more likely to select the *out-featured* competitor than a baseline foil. At the same time, it was predicted that this would not be true for the *in-featured* competitor due to the pruning of its incorrect association with the word.

### 2.3.2 *Training day 1*

As can be seen in Figure 4, average accuracy climbed above 90% by block 2 of training (chance was at 25%, as there were four objects on the screen, one of which was the target). This high overall level of performance was observed despite considerable individual differences in performance in block 1 of training. While people ranged from 39% to 98% correct in block 1, by block 3, all were above 79% correct. Overall, accuracy during the training section suggests that participants were able to acquire word-object-mappings easily even in the presence of the misleading associations included in pre-training.

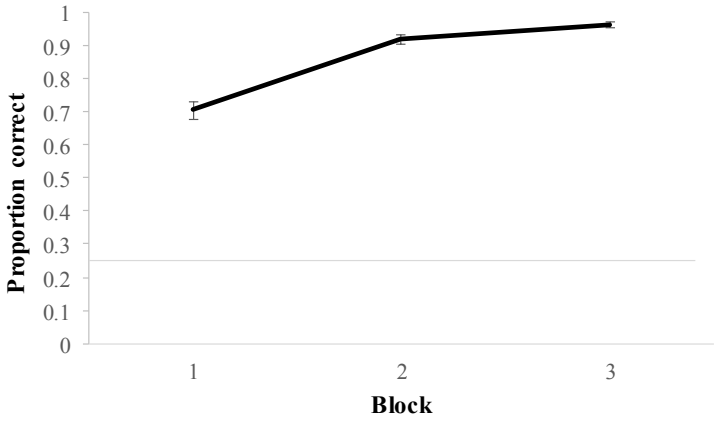

Figure 4: Proportion correct during training blocks of Experiment 1. Error bars indicate standard error of the mean.

To test the validity of the pre-training manipulation, I asked whether participants were sensitive to the pre-training statistics. If so, during training, they should be more likely to select the featured object (that had been paired with the word on 40% of all trials during pre-training) than a randomly selected foil whenever it was a response option. To investigate this, response choices were examined during training trials that included the *in-featured* object, along with two other foils. These were examined on block 1 (accuracy was too high in the remaining blocks for this measure to be meaningful): On these trials, participants select the target object on 66% of trials. The *in-featured* foil was chosen on 16% of all trials, and a baseline foil was selected on 8%. To investigate whether this difference between *in-featured* and baseline foil selection was significant, I computed the log-odds ratio of the likelihood of choosing an *in-featured* or baseline

foil choice for each participant. This ration was then tested against zero, using a two-tailed one-sample t-test: Participants chose the *in-featured* foil significantly more often than the baseline foil ($t(39)= 3.56$, $p < 0.001$), suggesting that they had learned and retained the pre-training statistics.

I next analyzed eye movements on the training trials. For this analysis, only correct trials were included. Again, participants' behavior during *in-featured* training trials was investigated, testing whether they were more likely to look at an *in-featured* object than a baseline foil (the average of the two baseline foils) even when they had selected the correct object (c.f., Roembke & McMurray, 2016).

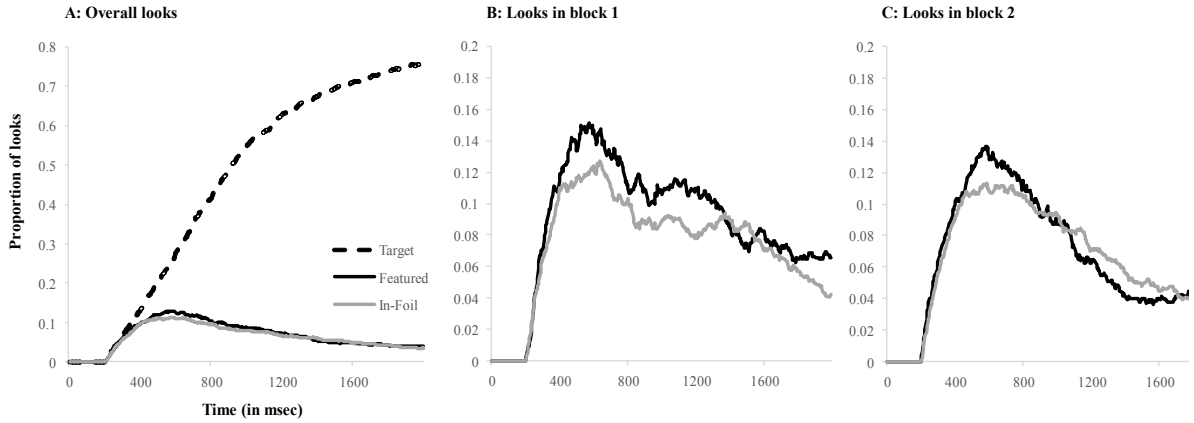


Figure 5: Panel A depicts overall looks during training blocks on day 1. Eye movements in block 1 (Panel B) and block 2 (Panel C) are shown in the subsequent panels. In Panels B-C, looks to the target are removed to zoom in on foils.

Panel A of Figure 5 depicts proportion of looks averaged across training on correct *in-featured* trials. Participants' looks quickly converged on the target; this is not surprising given only trials were included where participants clicked the target and the absence of phonological overlap between words. Moreover, participants appear to look slightly more to the featured competitor than the baseline foil. Increased looks to the featured competitor appear to be more pronounced early on in training (block 1, Figure 5B) than later (block 2, Figure 5C).

For analysis, the looks for both the *in-featured* object and the averaged baseline foils were calculated between 300 msec and 2000 msec for each participant (the area under the curve); looks that were generated right before or after this window were excluded (see Figure 5). The resulting values were analyzed using a mixed effects model in R (version: Ri386; R Core Team, 2014) with a fixed effect of block (1-3; centered) and object type (transformed; *in-featured* = 0.5, *in-foil* = -0.5). To select the random effects, model comparisons were used to find the model with the most complex random effect structure needed to fit the data (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Chi-square difference test was used for nested models to determine the most appropriate random effect structure. Here, I will report chi-square statistics for the comparison of the most complex model that reached significance and the previous model, or chi-square statistics for the second least complex model evaluated if none of the more complex models were significantly better than the least complex model tested. The model with only subject as a random intercept and no random slopes was found to best capture the data ($\chi^2(1) = 12.70$, $p = 0.100$).

In this model, the main effect of block was significant (B = -0.02, SE < 0.01, $t(1816) = -9.18$, $p < 0.001$), indicating that participants' looks to both types of objects decreased as the experiment progressed. This is a typical eye-tracking result; participants learn the position of the objects on the screen and make less eye movements in later trials.[2] In addition, no main effect of object type was observed (B = 0.01, SE < 0.01, $t(1813) = 1.55$, $p = 0.121$): Across the whole training phase, participants were equally likely to look to the *in-featured* object and the *in-foil*s

[2] It should be noted that trials never included foils that mapped onto phonological competitors (e.g., cohorts that overlapped at onset) during training; as a result, participants knew after hearing the first phoneme which object the word mapped on, facilitating making eye movements to the target.

(see Figure 5A). However, there was a significant interaction of block and object type (B = -0.01, SE < 0.01, $t(1813) = -2.43$, $p = 0.015$).

To investigate this interaction further, the data were split by block and the analysis repeated (the fixed effect of block was dropped). The main effect of object type was significant in block 1 (B = 0.02, SE < 0.01, $t(544.7) = 2.69$, $p = 0.007$; see Figure 5B), but not in block 2 (B < -0.01, SE = 0.01, $t(588.94) = -0.37$, $p = 0.714$; see Figure 5C) or block 3 (B < -0.01, SE < 0.01, $t(599) = -0.19$, $p = 0.853$). This pattern is consistent with pruning: In the beginning of training (right after participants completed pre-training), incorrect associations between the word and *in-featured* competitor still exist; this is reflected in participants' increased probability to look at the *in-featured* competitor in comparison with a randomly selected *in-foil* during block 1 of training.

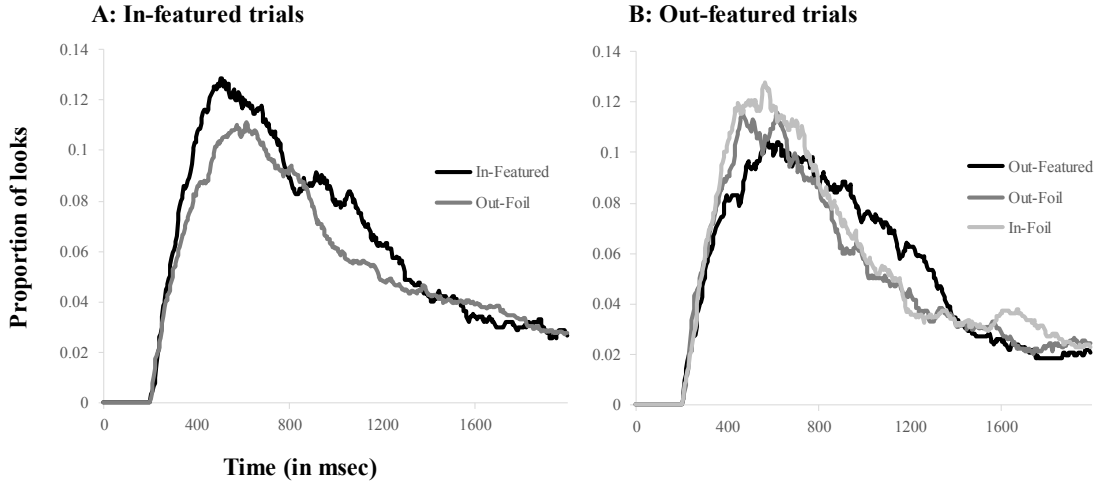


Figure 6: Eye movements during testing on day 1. Panel A shows looks during in-featured testing trials (looks to out-foils were averaged to create the baseline). Panel B displays looks during out-featured testing trials.

However, as they are exposed to feedback as to what the correct word-object-mappings are and/or learn the correct mappings, participants stop to consider the *in-featured* competitor (block 2 and block 3).

### 2.3.3 *Testing day 1*

During the testing section at the end of training, listeners were highly accurate (M = 98%). Thus, only eye movement data were analyzed. Eye movements were analyzed separately for each trial type (recall the trial types differed on which types of competitors were present on the screen). During *in-featured* testing trials, similarly to *in-featured* trials during training, looks to the *in-featured* competitor were compared to the looks to the baseline foil by calculating area under the curve between 300 and 2000 msec. However, in contrast to *in-featured* training trials, the baseline was now two *out-foils* from the other cluster (looks to the two *out-foils* were averaged to find the baseline). As can be seen in Figure 6A, participants looked more to the *in-featured* competitor than *out-foils*, though differences were small.

This was investigated statistically in a mixed effects model that included fixed effects of block (centered) and object type (transformed; *in-featured* = 0.5, *out-foil* = -0.5). The model that best fit the data ($\chi^2(1) = 2.52$, $p = 0.112$) included a random intercept of subject but no random slopes. Participants did not significantly look more at one object type than the other (B = 0.01, SE < 0.01, $t(599) = 1.54$, $p = 0.124$). This is consistent with the finding from block 3 in training, indicating that participants previously established preference for the *in-featured* competitor was gone by the end of training and thus also during testing.

In addition, I asked whether participants looked more to the *out-featured* competitor than a baseline. To investigate this, looks to the *out-featured* competitor were compared to looks to the *in-foil* and *out-foil*. Figure 6B indicates that participants' looks to the *in-foil* and the *out-foil* were similar in shape. Eye movements to the *out-featured* competitor had a slightly lower peak and were harder to suppress, though differences between foil types again were very small. Theoretically, associative strength should be lowest between the target word and the *out-foil*

(they did not co-occur during training and were not paired during pre-training), but stronger between the word and the *out-featured* (they were paired during a minority of trials during pre-training) and the *in-foil* (they co-occurred with the word during training, though they also received negative feedback). To investigate this pattern, I used a mixed effects model with looks between 300 and 2000 msec as the dependent variable. Fixed effects included two contrasts, comparing looks to the *out-featured* competitor and looks to the *in-foil* as well as looks to the *out-featured* competitor and the *out-foil*. The model that best fit the data included a random intercept of subject but no random slopes ($\chi^2(1) = 1.36$, $p = 0.243$). Neither the comparison of *out-featured* object and *in-foil* (B < 0.01, SE < 0.01, $t(918) = 1.54$, $p = 0.582$) nor the comparison of the *out-featured* competitor and *out-foil* (B < -0.01, SE < 0.01, $t(918) = -0.95$, $p = 0.341$) reached significance.

### 2.3.4 *Training day 2*

Participants' performance remained very high throughout training on day 2 (average accuracy block 1 = 99%; average accuracy block 2 = 99%). To analyze participants' proportion of looks, the same statistical analysis was used as for day 1 training, with a fixed effect of object type and block. The model that best fit the data was the one with a random intercept of subject and stimulus but no random slopes ($\chi^2(1) = 4.87$, $p = 0.027$). There was no significant effect of block (B < 0.01, SE < 0.01, $t(1223) = 0.21$, $p = 0.835$), indicating that participants' looks did not decrease over training. This may have been the consequence of training on day 2 only consisting of two blocks (instead of three blocks on day 1). Moreover, there was no main effect of object type (B < 0.01, SE < 0.01, $t(1223) = 0.72$, $p = 0.475$) nor a significant interaction of block and object type (B < 0.01, SE < 0.01, $t(1223) = 0.46$, $p = 0.648$): Participants were equally likely to

look at the *in-featured* competitor and the randomly selected *in-foils*. This suggests that pruning (as well as associative building) was completed by day 1, and that participants did not benefit from sleep and additional training.

### 2.3.5 *Testing day 2*

There were two testing sessions on day 2, one right before training (the first section of day 2) and one right after. Accuracy was high in both testing sessions (testing 1 M = 98%; testing 2 M = 99%). I used the same statistical approach to examine the fixation data as for testing on day 1, with object type as the only fixed effect. The model that best captured the data for *in-featured* testing trials of testing 1 included random intercepts for subject and auditory stimulus but no random slopes ($\chi^2(1) = 5.12$, $p = 0.024$). A significant effect of object type was observed (B = 0.01, SE < 0.01, $t(571.50) = 2.30$, $p = 0.022$), indicating that looks to the *in-featured* competitor were higher than baseline foils. This was surprising, as there had not been a significant effect of object type at the end of day 1 (during training or testing). For *out-featured* testing trials of testing 1, the model with a random intercept of subject but no random slopes best fit the data ($\chi^2(1) < 0.01$, $p = 1$) and there was no significant difference between looks to the *out-featured* competitor and *in-foil* objects (B < 0.01, SE < 0.01, $t(915.05) = 0.76$, $p = 0.450$) as well as to the *out-featured* competitor and *out-foil* objects (B < 0.01, SE < 0.01, $t(915.05) = -1.50$, $p = 0.133$).

For the second testing session of day 2 (after training), the model that best fit the data included a random intercept of subject but not random slopes ($\chi^2(1) = 0.49$, $p = 0.483$). No significant effect of object type was observed (B < 0.01, SE < 0.01, $t(599) = 0.34$, $p = 0.734$). Similarly, for *out-featured* testing trials, the model with only a random intercept of subject but no random slopes fit best ($\chi^2(1) < 0.01$, $p = 1$), and participants did not look more to the featured competitor than the *in-foil* (B < 0.01, SE < 0.01, $t(915) = 0.36$, $p = 0.720$) nor the *out-foil* (B < -0.01, SE < 0.01, $t(915) = -0.95$, $p = 0.344$). This is consistent with the data from training on day 2 and the completion of pruning.

All eye-tracking results are summarized in Table 5. Overall, the majority of results on day 2 were null effects (with the exception of the first testing phase on day 2).

Table 5: Overview of eye-tracking results of Experiment 1. '-' indicate results with p > 0.05. '*' indicate results with p < 0.05.

| Day | Session | Block | Object Type | Block × Object Type |
|---|---|---|---|---|
| Day 1 | Training | * | - | * |
| | Testing | na | *In-featured*: -<br>*Out-featured*: - | na |
| Day 2 | Testing 1 | na | *In-featured*: *<br>*Out-featured*: - | na |
| | Training | - | - | - |
| | Testing 2 | na | *In-featured*: -<br>*Out-featured*: - | na |

### 2.3.6 *No-correct testing*

For no-correct testing analyses, data from the first participant were excluded because the participant did not understand the instructions: He or she did not know that they had to make a best guess in the absence of the target, and just selected the same location repeatedly. After noticing this, instructions were changed to specify this more clearly for the rest of the participants. In addition, an exclusion criteria was adopted for all other experiments: If a

participant selected the same response location eight times in a row, this was seen as evidence that s/he did not comply with instructions. Participants that showed this response pattern were excluded. This was only true for the first participant, leaving 39 data sets for analysis.

Data were analyzed separately by trial type (control trial, *in-featured* or *out-featured* trial). As before, if a trial included more than one object of the same object type (e.g., control trials included two *in-foils* and two *out-foils*), the average of how often that foil type was selected was calculated. As a result, percentages of response choices do not always add up to 100%. Subsequently, ratios of percentages were calculated, asking whether one competitor type was selected more than another. For example, a ratio was calculated of how often the *in-featured* competitor was selected over a baseline foil. Finally, ratios were log scaled to achieve normality. The resulting number was then tested against 0, using a one-sample t-test. If the t-test was significant, this indicated that the numerator of the ratio (e.g., the featured competitor) was selected more often than the denominator of the ratio (e.g., the baseline foil).

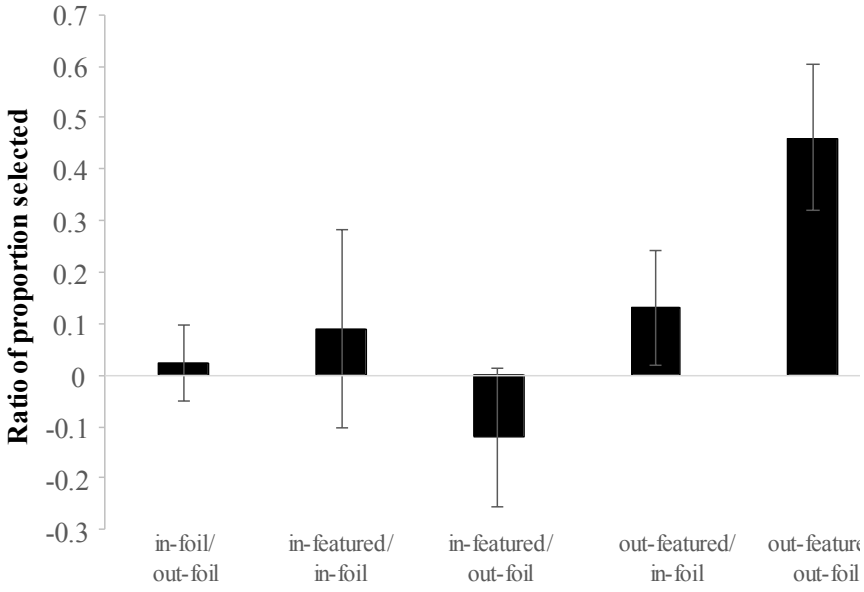


Figure 7: Means of ratios computed as part of the no-correct testing session of Experiment 1. Error bars represent standard error of the mean.

Ratios are plotted in Figure 7. For control trials (two *in-foils*, two *out-foils*), participants did not select one object type more than the other ($t(38) = 0.31$, $p = 0.378$). This was also true for *in-featured* trials (one *in-featured* foil, one *in-foil*, two *out-foils*); the *in-featured* competitor was not more likely to be selected than the *in-foil* ($t(38) = 0.47$, $p = 0.321$) or the *out-foil*s ($t(38) = -0.89$, $p = 0.810$). For *out-featured* trials (one *out-featured* object, one *out-foil*, and two *in-foils*), participants were not more likely to select the *out-featured* foil than the *in-foil* ($t(38) = 1.17$, $p = 0.125$). However, they were significantly more likely to choose the *out-featured* object than the *out-foil* ($t(38) = 3.26$, $p = 0.001$): The *out-featured* object was selected 30% of all trials, whereas the *in-foils* and *out-foils* were chosen 25% and 20% of all trials, respectively. This suggests that participants may have retained pre-training associations to *out-featured* objects on some level, at least in comparison to the *out-foil*.

### 2.3.7 *Sleep questionnaire*

Participants reported to have slept on average for 7.72 hours (SD = ± 1.76 hours) between day 1 and day 2 of the experiment. They reported that on average they slept for 7.2 hours per night (SD = ± 1.01 hours) and that they needed approximately 7.61 hours (SD = ± 1.07 hours) to feel rested.[3] A small subset of students reported to have drank alcohol the night before (5 out of 40), and only two participants reported that they had more than one drink the night before. Sleep data was collected to describe the sample. No further analyses were completed with participants' sleep information.

[3] The average number was calculated when participants gave a range as a response.

## 2.4 Discussion of Experiment 1

Experiment 1 supports that participants acquired incorrect associations through a short pre-training: Participants were more likely to select the featured competitor than a baseline foil during block 1 of training. This pattern of response choices is indirect evidence that participants must have paid attention and learned to associate words with two referents during pre-training. This is consistent with previous findings that people are sensitive to co-occurrence statistics of words and objects (Roembke & McMurray, 2016; Vouloumanos, 2008), even if words appear frequently with more than one referent. Moreover, the eye movement analysis of training on day 1 is consistent with the hypothesis that participants are able to prune or unlearn these incorrect associations swiftly: Even though looks were increased to *in-featured* competitors in block 1 of training, this preference was no longer observable in block 2 or block 3. Thus, in the beginning, the increased associative strength between *in-featured* objects and words led to an increase in consideration (and thus more looks); as the associative strength between *in-featured* objects and words decreased, *in-featured* objects were treated more similarly to baseline foils.

At testing, looks to the *in-featured* competitor were compared to baseline foils that had been included in the other cluster (*out-foils*). The objects in the other cluster (*out-foils* and the *out-featured* object) never co-occurred with the target word during training. Thus, the *out-foil* had never been paired with the target word and could therefore represent the least likely referent. Alternatively, this characteristic may make objects from the other cluster, and the *out-featured* object in particular, the *best possible* referent, as they never were rejected as incorrect either; this would be consistent with learning being driven by supervised statistics mostly. Data from the testing trials on day 1, however, showed no significant differences between looks to any of the

object types. These findings are consistent with the idea that pruning incorrect associations with both types of featured objects was completed by the end of training.

Importantly, this was true for both *in-featured* as well as *out-featured* competitors: That suggests that incorrect associations with the *out-featured* competitor were pruned despite them never co-occurring with the word during training (as the object was assigned to a different cluster). This is inconsistent with how pruning operated in pigeons: Roembke et al. (2016) reported that pigeons were more likely to select the out-cluster (presumably unpruned) foil than the in-cluster (pruned) foil, thus resulting in lower accuracy levels on testing trials that included out-cluster (*out-foil*) items. The absence of increased looks to the *out-featured* competitor during testing trials suggests that the pruning of incorrect associations is possible, even if no explicit error signal is given.

In addition, participants' learning may have been influenced by their knowledge of real words: Words often map onto one semantic concept only. In fact, children have been observed to use this assumption when fast-mapping novel words onto objects (mutual-exclusivity assumption; Carey & Bartlett, 1978). It is possible that participants underwent a similar process here: Even though *out-featured* objects never co-occurred with the word of the other cluster during training (the one they had been paired with in pre-training), it still acted as the target object for another word in its cluster. For instance, object 10, the *out-featured* competitor to word 9, was also the target of word 10. As participants learned that object 10 also co-occurred with word 10 (and never with word 9 after pre-training), this might have led them to apply a version of the mutual-exclusivity assumption, pruning incorrect associations between object 10 and word 9. Pigeons, in contrast, do not possess prior experience with word-object-mappings (nor do they have the same level of cognitive abilities), thus potentially leading to the differences observed

between the two species. This is consistent with evidence that mutual exclusivity develops (Halberda, 2003; McMurray et al., 2012) and is shaped by children's linguistic experience (Houston-Price, Caloghiris, & Raviglione, 2010; Kalashnikova, Mattock, & Monaghan, 2015).

One limitation to this finding is that eye-tracking cannot assess associative training at a singular moment in time (e.g., one trial), but that rather a number of trials are needed to find the average of looks. Thus, it is possible that participants did consider the *out-featured* object more when they first encountered it again, but that this preference quickly disappeared and thus is not detectable in participants' average eye movements. Consistent with this hypothesis, in the no-correct testing session of pigeons, this preference was gone after a couple of trials (Roembke et al., 2016).

On day 2, there were little differences between how object types were processed/treated. The only significant effect of object type in eye movements was present during testing 1: Participants looked more to the *in-featured* competitor than one of the baseline foils. At the same time, this was not true for the *out-featured* competitor, which was equally considered to the present baseline foils (*in-foil* and *out-foil*). These differences disappeared by the time the second testing session was completed, and there were also no differences between objects during training. In addition, participants were more likely to select the *out-featured* competitor—similarly to pigeons in Roembke et al. (2016)—in the no-correct testing session. This was not true for the *in-featured* competitor.

Why would participants be more likely to consider the *in-featured* competitor the first time they were brought back on day 2, but not the *out-featured* one? One explanation of these results could be that even as the strength of the association between a word and an object was zero, participants still retained some "knowledge" of a previously established connection. Some

evidence for this may come from day 2 of Experiment 1: Participants looked more to the *in-featured* competitor than the baseline foil at the beginning of day 2. This might indicate that memories of previous (spurious/incorrect) associations are retained, even if they cannot be detected by the end of day 1. This is consistent with the study by Yurovsky, Fricker, Yu, and Smith (2014), where people had learned information about word-object-mappings although they performed at chance. In addition, these data highlight the possibility that word-object-mappings consistent with the previously acquired incorrect associations may be more easily learned. This opens up the question: Were the incorrect associations really pruned by the end of day 1 or were participants simply adequate at suppressing them to perform the task more efficiently? It is impossible to draw this distinction from the data, though it is not clear why incorrect associations should be maintained indefinitely in the absence of their support by co-occurrence statistics.

At the same time, participants were found to be more likely to select the *out-featured* object in the absence of the target object (during the no-correct testing session). Implicit behavior can be driven by different processes/knowledge than explicit behavior. Whereas participants' eye movements are without question implicit (they are made to direct participants' motor response), selection in the no-correct testing session may not be. In contrast, participants may select an object in the absence of the target based on explicit strategies or remembering the pre-training session. However, it is also possible that no-correct selections are implicit, with participants selecting a particular referent "at random", even as the featured object is selected higher than chance. Most likely, the no-correct testing is less "pure", with a mix of implicit and explicit processes contributing to the outcome. This mismatch between the two measures in question (eye movements vs. object selection during no-correct testing) may be the reason why the *in-featured* competitor is "preferred" in the beginning of day 2 and the *out-featured* one at the end of day 2.

The other possibility is that the preference for the *in-featured* competitor was seen during the first testing of day 2, whereas the preference for the *out-featured* competitor was not present until the end of day 2. That means, by the time participants completed the no-correct testing session, they had received approximately one additional hour of exposure to the word-object-mappings. This additional training, even though participants were at ceiling throughout, may have resulted in more subtle changes of the associative representation of the mappings in questions.

Together, results from day 1 and day 2 suggest that no sleep is required to allow for the pruning of incorrect associations, as pruning appeared to be completed within the first session. However, the re-emergence of looks to the *in-featured* competitor at the beginning of day 2 could indicate that incorrect associations were reinforced or consolidated with sleep. This might have led to the reactivation of latent associations on the next day.

Overall, these results are not consistent with feedback and supervised learning being the only driver of associative strength. Participants' looks to the *in-featured* competitor on day 2 of testing may indicate that—even though they received much feedback that this is not the correct referent for the just heard word—it still is more likely to co-occur with the target than one of the objects assigned to the opposite cluster. Thus, unsupervised learning based on the simple co-occurrence of words and objects may play an important role, too.

There are differences in the design of this study to the one conduced in pigeons (Roembke et al., 2016) that might contribute to how exposure to the featured competitor affected learning: For pigeons, each trial ended on a correct response, as they were required to complete correction trials after selecting a foil (not a target). For humans, in contrast, participants moved

on after receiving a feedback signal indicating an incorrect response. This difference might have decreased opportunities for pruning in the human experiments.

# CHAPTER 3: EXPERIMENTS 2 AND 3

## 3.1 Experiment 2: Overview

Experiment 1 included first evidence that people prune incorrect associations between words and an incorrect competitor. Importantly, all effects observed in Experiment 1 were subtle (as expected in the absence of phonological overlap between words). Thus, it is important to replicate these findings to be more confident in people's ability to prune incorrect associations; this was the goal of Experiment 2.

As Experiment 1 revealed little significant differences in looks to object type on day 2, Experiment 2 was designed to be a one-day study only. In addition, several smaller design changes were made to optimize the measurement of pruning incorrect associations. For example, objects during pre-training were presented randomly in the four corners of the screen instead of always in the screen center; this change was made to encourage participants to look to the object a word was paired with. In addition, a section at the end of Experiment 2 was added during which real word competitors were included to assess competition between novel words and the existing lexicon as part of Question 4. To do so, nonword items in Experiment 2 were matched with existing English words.

## 3.2 Method

### 3.2.1 *Participants*

Participants were 40 monolingual, native English speakers with normal or corrected-to-normal vision; all were students at the University of Iowa and received course credit for participation. One participant did not complete the experiment due to time restrictions; their data were included when possible. Participants underwent informed consent as part of an IRB approved protocol.

### 3.1.2 Stimuli

Stimuli were based on the stimuli used in Experiment 1; however they were modified in several ways to make them compatible with the second phase of the experiment which compared word recognition in newly learned and known words.

Referents were clipart images of novel objects. The 16 photographs of unusual objects that were used in Experiment 1 were converted into clipart images. Clipart images of objects were developed to allow for the direct comparison of novel objects and clipart images of real words that are typically used in the visual world paradigm with real words.

Table 6: List of novel words used in Experiments 2-4.

| Written word form of nonword | IPA of nonword | Cohorts |
|---|---|---|
| beamler | /bimlɚ/ | beaker, beaver |
| camat | /kæmæt/ | camel, candle |
| dellet | /delet/ | dentist, desert |
| fumlic | /fuːmlɪk/ | fungus, funnel |
| garsit | /gærsɪt/ | garden, garlic |
| hustrim | /hʌstrɪm/ | huddle, hunter |
| jospit | /tʃɔːspɪt/ | jockey, jogger |
| lindle | /lɪndle/ | lizard, liver |
| mougger | /maʊgɚ/ | mountain, mousetrap |
| nocket | /nɔːket/ | nostril, nozzle |
| parshim | /pærʃɪm/ | parcel, party |
| raimmer | /reɪmɚ/ | raincoat, raisin |
| sauble | /sɑːble/ | saucer, sausage |
| shemster | /ʃemztɚ/ | shepherd, shelter |
| turgon | /tɝgɔːn/ | turbine, turkey |
| wammock | /wæmɔːk/ | wallet, water |

Previous iterations of these images have been successfully used in completed studies of word learning and speech perception (Apfelbaum & McMurray, 2016; Roembke & McMurray, 2016).

Words were two-syllable pseudo words (see Table 6). These were different to the stimuli used in Experiment 1, and were created to have a strong real word cohort competitors as required for exploratory Question 4. As a result, nonwords included consonant clusters. As in Experiment 1, none of the nonword words were cohort competitors of each other. Auditory stimuli were again recorded by a female, native speaker of English, and five exemplars were selected for each word. All auditory stimuli underwent the same editing process as described in Experiment 1.

### 3.2.2 *Design*

As in Experiment 1, participants learned 16 word-object mappings. For each participant, words were assigned to an object randomly. Each word was also assigned a featured competitor. Instead of the completely random assignment of words and featured competitors in Experiment 1, assignment of words/featured competitors was completed in four different ways. This was done to avoid semantically close competitors when real words were included in the word recognition session: For instance, the cohorts of the nonword BEAMLER were BEAVER and BEAKER. CAMAT's cohorts included the word CAMEL. As both the words CAMEL and BEAVER describe animals, including both pictures on the same trial might lead to looks driven by their semantic closeness. To avoid this, BEAMLER's featured object never mapped onto CAMAT.

As in Experiment 1, featured competitors were mirrored within a word pair, where the featured competitor of one word was the target for the other, and vice versa. Half of the words

were assigned to cluster A and half to cluster B in a way that half of the word pairs that shared target/featured objects were in the same cluster and half were not.

As in the prior experiment, participants first completed the pre-training. This was followed by three blocks of training, a testing session and the no-correct testing. Last, participants completed a real-time word recognition phase to examine how novel words competed with known words (see Figure 8 for a schematic). All of these sessions were completed within one day, and the experiment overall took approximately 1.5 hrs.

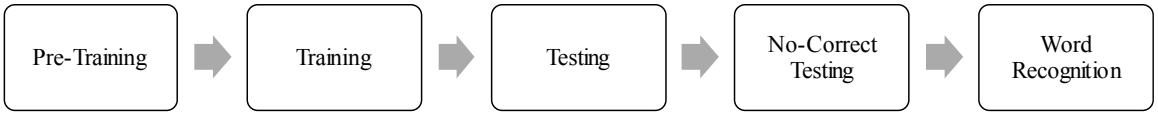


Figure 8: Overview of the basic structure of Experiments 2-5. Experiment 5 did not include a word recognition session due to stimulus type restrictions. Eye movements were tracked from the training session onwards.

As in Experiment 1, pre-training consisted of 160 trials during which only one object was presented on each trial. Each word was paired with its target object on 60% of all trials and its featured competitor on 40% of all trials (there were ten trials per word). In contrast to Experiment 1, the object was randomly presented in one of the four corners (instead of always being in the middle of the screen offset to the top). This change was implemented to encourage participants to pay attention to the object that was paired with each word, as the changing position required participants to make increased eye movements to click the object.

Training was structured identically to Experiment 1: On each trial, participants saw four objects, one of which was always the target, and heard one word. Training consisted of three blocks of 112 trials (7 trials/ word), resulting in an overall number of 336 trials. Feedback as to whether participants had selected the correct object was provided during training. The design of testing was the same as in Experiment 1. Testing consisted of 160 trials (10 trials/ word), and feedback was never provided. Subsequently, participants received instructions to complete the

no-correct testing of 32 trials (2 trials/ word), during which the target object was never present. Again, the same design was used as in Experiment 1.

Finally, participants completed the novel real-time word recognition phase; this will be described in more detail as part of Chapter 6 (Question 4).

### 3.2.3 *Procedure*

Trials were structured similarly to Experiment 1: In the beginning of each trial, participants were given time to pre-screen the present object(s). After 1050 msec, the blue circle in the middle of the screen tuned red, which cued participants to click on it. Clicking the circle triggered the word to be played. A trial ended after a participant selected one of the present objects. During pre-training, only one object was present; nevertheless, participants were required to click on it to advance to the next trial. Auditory feedback (the same as in Experiment 1) was only given during training blocks. Trials were never time-limited.

### 3.2.4 *Eye-tracking recording and analysis*

The same eye-tracking recording apparatus and method of analysis were used as in Experiment 1.

## 3.3 Results

I first analyzed the training data before investigating testing trials and the no-correct testing session. Please note that the results of the real-time word recognition section will be analyzed separately and reported later (see Chapter 6).

### 3.3.1 *Training*

Accuracy was very high in Experiment 2: Participants performed above 80% correct by block 1 and accuracy was close to perfect by the end of training (M = 97%). No participant failed to learn and none of the participants were excluded from analysis as a result. This high level of performance rate is shown in Figure 9, suggesting a higher learning rate than in Experiment 1 (see Figure 4). Importantly, there is no strong reason to believe this was due to the small design changes (the only differences between Experiment 1 and 2 during pre-training and training were the change in auditory stimuli and the presentation location of the object during pre-training); thus, the apparent differences in learning rate may simply reflect random individual differences between the samples.

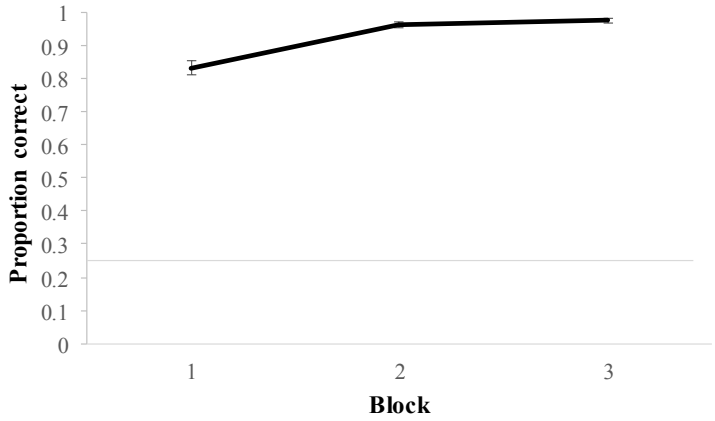


Figure 9: Proportion correct during training blocks of Experiment 2. Error bars indicate standard error of the mean.

I next investigated participants' response choices during block 1 of training, in order to test if participants showed evidence of having acquired the pre-training statistics. Participants selected the target on 79% of the trials, the *in-featured* on 12% of the trials and one of the baseline foils on 8% of the trials. Response selections were converted to log-odds ratios and analyzed using a one-sample t-test. The difference between selecting the *in-featured* over the

baseline was significant ($t(39)= 3.91$, $p < 0.001$), indicating that participants must have encoded the incorrect association between a word and its featured competitor during pre-training.

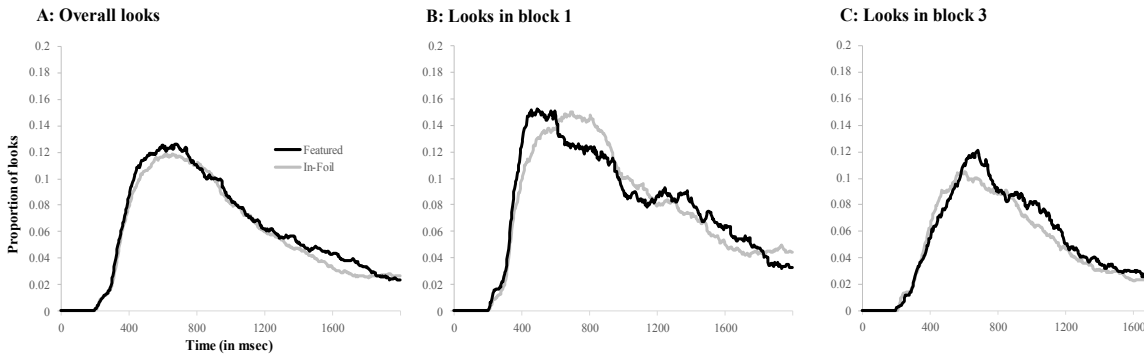


Figure 10: Overall looks during in-featured trials of training in Experiment 2 (Panel A), and during block 1 (Panel B) and block 3 (Panel C) of in-featured trials. Looks to the target are not depicted.

I next examined the eye-movements during training. Surprisingly, as is evident in Figure 10, participants did not look more at the *in-featured* competitor than the *in-foil* object during training. This was even true for looks during block 1 (see Figure 10B and Figure 10C). To confirm this statistically, the same mixed models approach was used as in Experiment 1. The dependent measure was the proportion of looks to the *in-featured* competitor and *in-foil* between 300 and 2000 msec. The fixed effects were object type (transformed; *in-featured* = 0.5, *in-foil* = -0.5), block (centered), and their interaction. As before, I used a chi-square difference test to evaluate nested models. I will report chi-square statistics for the comparison of the most complex model that reached significance and the less complex one before, unless no additional random intercepts and slopes were required. The model that fit the data best included random intercepts of subject and auditory stimulus with a slope of the interaction term on subject ($\chi^2(9) = 18.32$, $p = 0.032$). Using this model, the main effect of block was significant (B = -0.02, SE < 0.01, $t(37.52) = -6.72$, $p < 0.001$). This indicates that participants made less eye movements as the experiment progressed. However, neither the main effect of object type (B = 0.01, SE < 0.01,

$t(39.11) = 1.34$, $p = 0.189$) nor the interaction of block and object type (B < 0.01, SE < 0.01, $t(173.66) = 0.22$, $p = 0.824$) were close to significant. The absence of a main effect of object type or a significant interaction (the latter of which was observed in Experiment 1) suggests that participants must have pruned incorrect associations established during pre-training very swiftly. This likely reflects their better learning.

### 3.3.2 *Testing*

Accuracy at testing was high (M = 99%). For *in-featured* testing trials, the same analysis strategy was used as for training with the exception that the baseline was now looks to *out-foils* (each testing trial always included two objects from the in- and two from the out-cluster). Proportion of looks to the featured and baseline foils was calculated between 300 and 2000 msec on correct trials only. Again, model comparisons were first conducted to find the model that best fit these data; models always included the fixed effect of object type. The model with only a random intercept of subject best fit the data ($\chi^2(1) = 0.30$, $p = 0.585$). Consistent with the training data, participants did not look significantly more at the *in-featured* competitor than the *out-foils* (B < 0.01, SE < 0.01, $t(597) = 1.02$, $p = 0.308$).

For *out-featured* testing trials, a contrast was created for looks to the *out-featured* competitor versus looks to the *in-foil* as well as one for looks to the *out-featured* competitor versus looks to the *out-foil*. Again, the model with only a random intercept of subject best fit the data ($\chi^2(1) = 0$, $p = 1$). However, neither comparison of object type reached significance (*out-featured* vs. *in-foil*: B < -0.01, SE < 0.01, $t(912.06) = -0.77$, $p = 0.443$; *out-featured* vs. *out-foil*: B < -0.01, SE < 0.01, $t(912.06) = -0.40$, $p = 0.687$).

For the no-correct testing, two participants were excluded from analysis as they selected the same response location more than eight times in a row. This suggested that they did not comply with task instructions. After converting percentage selections into ratios, ratios were compared to zero (indicating no difference in frequency of selection) using a one-sample t-test. None of the tested contrasts reached significance (all $M < 0.25$; all $p > 0.100$), suggesting that participants did not "prefer" one response choice over another in the absence of the target.

### 3.4 Discussion of Experiment 2

Results of Experiment 2 were surprising: During training, participants were not more likely to look at the *in-featured* competitor than the baseline *in-foils*. This was despite the very clear parallels with Experiment 1 (similar pre-exposure to featured competitor; identical training and testing regime), which found evidence for increased looks to the *in-featured* competitor during training. In addition, no increased looks to the *in-featured* or *out-featured* competitor were observed during testing. This is consistent with the notion that pruning was completed within training. Moreover, the no-correct testing session did not reveal any selection preferences of participants in the absence of the target object.

There are (at least) two possible explanations for this: Participants might have not acquired the misleading associations during pre-training, thus resulting in featured objects to be equally associated with the target words as other foils. If this were the case, there would have been no need to prune incorrect associations. However, challenging this view, participants' selections during block 1 of training indicate that they were sensitive to the pre-training co-occurrence of words and featured competitors: They were significantly more likely to select the *in-featured* competitor than a randomly selected *in-foil*, though the difference between *in-*

*featured* and *in-foil* selections was smaller than in Experiment 1 (difference Experiment 1: 8%; difference in Experiment 2: 4%). This finding suggests that participants paid attention during pre-training and that pre-training affected their behavior subsequently. Thus, it is likely that the reason why we did not observe increased looks to the *in-featured* competitor during training is the result of the very quick pruning of incorrect associations.

Thus, associations between words and their *in-featured* competitors may have been pruned extremely quickly. If the latter were the case, it would be hard to detect this quick elimination process in an eye-tracking paradigm that relies on the averaging of a number of trials. Why might negative associative learning be swifter in Experiment 2 than in Experiment 1? Looking to participants' accuracy performance across both experiments, it is clear that learning was better in Experiment 2 than in Experiment 1. Learning word-object-mappings in Experiment 2 might have been easier due to the higher phonotactic regularity of words used: As nonword stimuli (e.g., CAMAT) were created to have word competitors (e.g., CAMEL), they were more similar to existing words than stimuli used in Experiment 1. Previous research has shown that novel words that are phonotactically more frequent are more easily acquired (Gonzalez-Gomez, Poltrock, & Nazzi, 2013; Storkel, 2001, 2003; Storkel & Rogers, 2000). Participants' higher learning rate might have facilitated the pruning of incorrect associations. Alternatively, the reduction of incorrect associations might have allowed students to acquire word-object-mappings more quickly. We cannot draw any conclusions on the directionality of the effects, though it is possible that the two are related.

Of course, differences between experiments' results might also be the consequence of random individual differences between the two samples that participated in the studies. In addition, as can be seen in Experiment 1, the investigated effect—differences between looks to

the *in-featured* competitor and *in-foil*—are small and subtle, thus increasing the possibility to detect it in one experiment but not another. However, this also underlines the possibility that results of Experiment 1 (i.e., observing the pruning of incorrect associations) are not strong.

Even though one should be careful drawing strong conclusions, Experiment 2 may indicate that the pruning of subtle associations may be very quick, particularly in the type of learning paradigm employed here: Participants received feedback, thus potentially allowing for the efficient pruning of associations. However, pruning of incorrect associations may be slower/more demanding if spurious associations are more pronounced; this should also facilitate observing the pruning process within the previously used eye-tracking paradigm.

## 3.5 Experiment 3: Overview

The objective of Experiment 3 was to investigate whether it would be possible to observe pruning if learning was more difficult, as one of the apparent differences between Experiment 1 and Experiment 2 was that accuracy was higher in the latter. Thus, I decided to increase the strength of the incorrect associations: In pre-training, words were now paired with its featured competitor on the majority of trials (80%) and with the target object on 20% of trials. It was predicted that participants would take longer to unlearn the incorrect associations to the featured object, resulting in increased looks to the *in-featured* object during training. However, given how quickly participants were found to learn in the past, it was still predicted that pruning would be completed by testing.

## 3.6 Method

### 3.6.1 *Participants*

Forty-two monolingual, native speakers of English with normal or corrected-to-normal vision participated in this experiment. All were students at the University of Iowa and received course credit as compensation for participation.

### 3.6.2 *Stimuli and design*

The same design and materials were used as in Experiment 2 with one exception: During pre-training, each word was paired with their assigned target on a minority (20%) of trials (2/10 trials/word). At the same time, each word was paired with their featured competitor on 80% of all trials (8/10 trials/word). This was done to build a high amount of spurious (incorrect) associations with the featured competitor, thus increasing the need to prune incorrect associations.

### 3.6.3 *Procedure*

The same procedure as in Experiment 2 was utilized.

### 3.6.4 *Eye-tracking recording and analysis*

Eye movements were recorded using the same apparatus and analysis techniques as in Experiment 1 and Experiment 2.

## 3.7 Results

The same analysis approach was used as for Experiment 2. The only difference between Experiment 2 and Experiment 3 was that the word co-occurred more frequently with the featured competitor than its target object during pre-training.

### 3.7.1 *Training*

Three participants were excluded, two because of construction noise during the experiment. The third subject was removed from analysis, as they decided to quit the experiment after two blocks of training, leaving too little data.

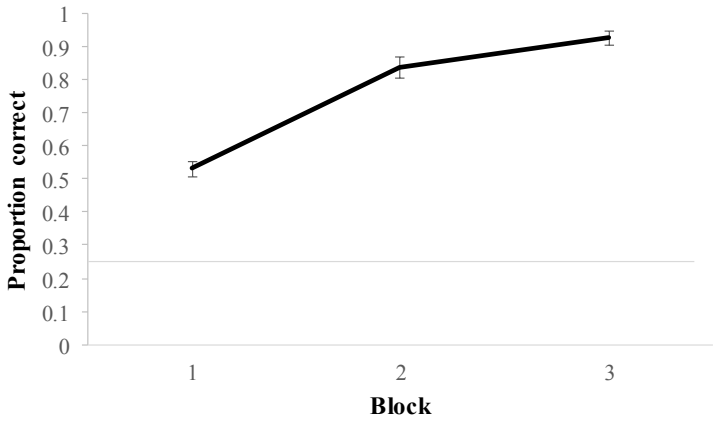


Figure 11: Proportion correct during training blocks of Experiment 3. Error bars indicate standard error of the mean.

Learning performance during training is presented in Figure 11: Participants were much slower to learn the word-object-mappings when pre-training favored the featured competitor instead of the eventual target object (e.g., block 1 M = 53%). First, I analyzed participants' responses on *in-featured* training trials of block 1 to test if the *in-featured* competitor was selected more often than a baseline foil: The target was selected on 45% of trials, the *in-featured* object on 34% and the *in-foil* on 12% of the trials. A similar odds ratio analysis showed that the difference between the rate of selecting the *in-featured* and the *in-foil* was significant ($t(38) = 10.44$, $p < 0.001$). Thus, there was only a slight increase in how often the *in-foil* was selected in comparison to Experiment 2 (8%), but

participants' selection between the *in-featured* competitor and the target object were more equally split. This is further evidence that participants were sensitive to the co-occurrence statistics they were exposed to during pre-training.

Next, eye movements during training were analyzed. Eye-tracking data from training can be seen in Figure 12; this depiction suggests that participants may have retained (relatively strong) incorrect associations throughout training, even as they learned the correct word-object-mappings. Interestingly, looks to the featured competitor appear to be above the *in-foil* even at the end of the trial; this suggests that the featured competitor might never get fully suppressed (during real-time processing).

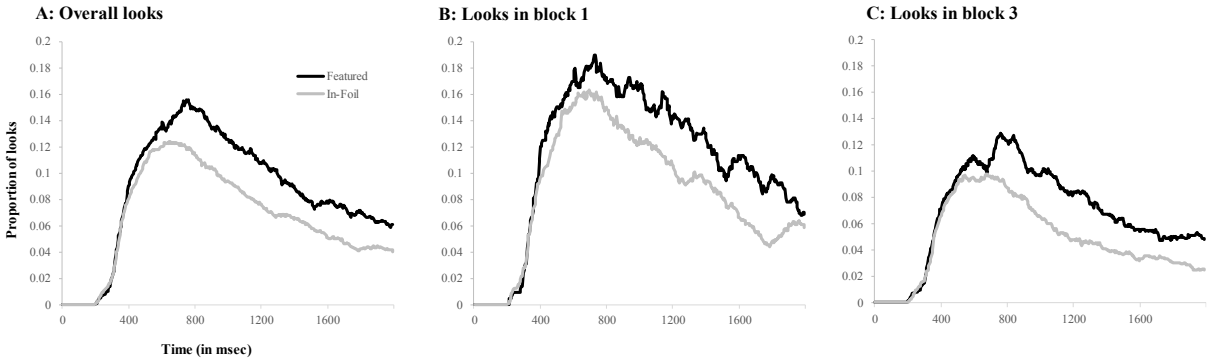


Figure 12: Overall looks during in-featured trials of training in Experiment 3 (Panel A), and during block 1 (Panel B) and block 3 (Panel C) of in-featured trials. Looks to the target are not depicted.

To compare looks to the *in-featured* over-and-above the baseline *in-foils*, I took the same statistical approach as before: The proportion of looks to each object type was calculated on correct *in-featured* trials between 300 and 2000 msec. The fixed effects were object type (transformed; *in-featured* = 0.5, *in-foil* = -0.5) and block (centered). Nested models were compared using the chi-square test, and I will report chi-square statistics of the comparison of the most complex model that reached significance and the less complex one before (when possible). The model that best fit the data included a random intercept for subject with a random slope of the interaction of block and object type on subject only ($\chi^2(9) = 25.96$, $p = 0.002$). It was found

that participants' overall number of looks decreased as the experiment progressed (B = -0.02, SE = 0.03, $t(49.74) = -7.40$, $p < 0.001$), consistent with standard eye-tracking findings. More interestingly, there was also a main effect of object type (B = 0.03, SE = 0.01, $t(36.92) = 4.44$, $p < 0.001$), indicating that participants were significantly more likely to look at the *in-featured* competitor than the *in-foil*, independently of the block number. This suggests that the increased co-occurrence of the featured object and the word during pre-training might have been strong enough to act as a "protector" against quick pruning. The interaction between block and object type did not reach significance (B < -0.01, SE = 0.01, $t(60.09) = -1.27$, $p = 0.209$). This suggests that higher looks to the featured competitor than the baseline foils remained stable across training.

### 3.7.2 *Testing*

Participants performed close to ceiling at testing (M = 96%). Proportion of looks to the featured competitor and baseline foils was calculated between 300 and 2000 msec, and analyzed with linear mixed effect models. For *in-featured* testing trials, the model that best captured the data included an additional random intercept of subject and auditory stimulus as well as a random slope of object type on subject ($\chi^2(2) = 9.30$, $p = 0.010$). This model showed a significant effect of object type (B = 0.03, SE = 0.06, $t(37.21) = 4.46$, $p < 0.001$), indicating that participants were more likely to look to the *in-featured* competitor than the baseline *out-foils*.

For *out-featured* testing trials, two contrasts were formed to compare looks to the *out-featured competitor* with *in-foils* and *out-foils*, respectively. The model with only a random intercept of subject fit the data best ($\chi^2(1) < 0.01$, $p = 1$). Neither the comparison of looks to the *out-featured* competitor and the *in-foil* (B < -0.01, SE < 0.01, $t(886.17) = -1.62$, $p = 0.106$) nor

the comparison of looks to the *out-featured* competitor and the *out-foil* (B < -0.01, SE < 0.01, $t(886.17) = -1.27$, $p = 0.203$) reached significance.

For the no-correct testing, one additional participant's data were excluded, as they did not complete this section due to running out of time. For each trial type, proportion of competitor selection were calculated and transformed into log-odds ratios. These data are presented in Figure 13. Ratios were compared against 0 using a one-sample t-test; values above zero indicate that the numerator of the ratio was selected more often than the denominator. Participants were found to be equally likely to select the *in-foil* and the *out-foil* during control trials ($t(37) = 1.26$, $p = 0.108$). The featured object was more frequently chosen than any other competitor type (with one exception, *in-featured/ in-foil*: $t(37) = 1.47$, $p = 0.075$); this was true for both *in-featured* (*in-featured/ out-foil*: $t(37) = 2.50$, $p = 0.008$) as well as *out-featured* competitors (*out-featured/ in-foil*: $t(37) = 1.82$, $p = 0.038$ ; *out-featured/ out-foil*: $t(37) = 2.98$, $p = 0.002$). These findings are consistent with the results of the eye-tracking results, suggesting that participants did not prune incorrect associations to the featured competitor during training. Surprisingly, this was even true for the *out-featured* competitor,

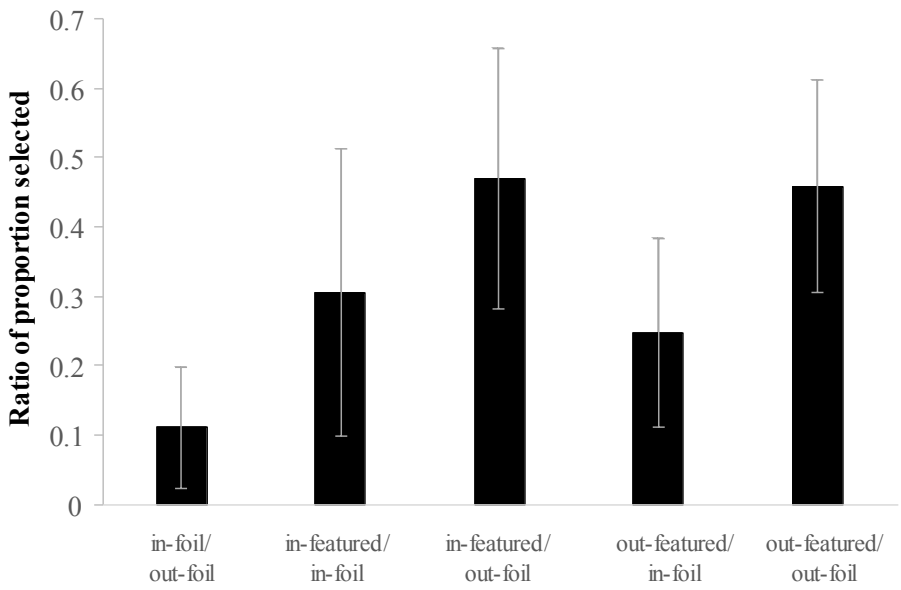


Figure 13: Means of ratios computed as part of the no-correct testing session of Experiment 3. Error bars represent standard error of the mean.

which never co-occurred with the target word during training (and despite the non-significant eye-tracking findings as part of the *out-featured* testing trials analysis).

### 3.8 Discussion of Experiment 3

To summarize the results of Experiment 3, participants kept looking to the featured competitor during training, even as they acquired the correct word-object-mappings. Moreover, participants more strongly activated the *in-featured* competitor in testing trials, and selected both the *in-featured* and the *out-featured* competitor more often than other competitor types during no-correct testing. This stands in high contrast to the findings of Experiment 2, where pruning must have been completed within block 1 of training. Again, the only difference between the two studies was the co-occurrence statistics during pre-training: During Experiment 2's pre-training, the word was presented with its target object on 60% of all trials and with the featured competitor on the remaining 40%. However, in Experiment 3, pre-training involved the word's presentation with the featured competitor on 80% of all trials, thus leaving only 20% of co-occurrence with the target object.

Experiment 3's results suggest two conclusions: First, pruning is not always quick and efficient (as observed in Experiment 2), but may instead be a more prolonged process sometimes. Under what circumstances then are associations pruned, and under which ones are they more "protected" (i.e., under which circumstances is pruning more difficult)? It is possible that participants' explicit (conscious) knowledge of the word-object-mappings played a role (which likely is impacted by the strength of the trained associations) in deciding. After Experiment 3's pre-training, participants might have been aware of which object each word (presumably) mapped onto (i.e., the featured competitor). During training and testing, the incorrect

associations between the word and the featured competitor might have been maintained through participants' explicit memory of what had happened during pre-training (potentially even as associative pruning took place). In Experiment 2, in contrast, participants might have not been consciously aware of the word more often appearing with its target due to the relatively even split between the two objects.

Of course, this explanation is speculative at this point. Nevertheless, it is consistent with ideas from the memory literature that how information is represented is influenced by people's explicit knowledge (or absence of it). Importantly, if pruning is impacted by participants' higher-level processing, it would not be possible to observe such interactions in less-developed species such as pigeons.

Second, Experiment 3 further supports the notion that pruning operates similarly for the *in-featured* as well as the *out-featured* competitor: Either associations with both are pruned (Experiment 1, Experiment 2) or not (Experiment 3). This is surprising, considering the findings in pigeons by Roembke et al. (2016), where pruning was dependent on the negative feedback received after incorrect selections.

Overall, Experiment 3 indicates that pruning may be a more complex process than previously assumed: If a word is presented with an incorrect referent repeatedly, people may struggle with unlearning this alternative "hypothesis", even as they correctly select the target object.

## 3.9 Comparison of Experiments 2 and 3

As pointed out before, the only difference between Experiment 2 and 3 were the co-occurrence statistics during pre-training. Thus, a direct comparison of their accuracy and eye

movement data allow for the testing of the importance of the change in pre-training statistics. More specifically, I asked whether accuracy was significantly lower during Experiment 3 training than Experiment 2's one (as suggested by Figure 9 and Figure 11), and how participants' eye movements changed as a result of the pre-training statistics.

## 3.10 Results

### 3.10.1 *Accuracy*

To investigate differences in accuracy across experiments, data from both experiments' three training blocks were entered into a binomial mixed model. Experiment was contrast coded (Experiment 2: +0.5; Experiment 3: -0.5); block, the second fixed effect, was centered as before. In addition, the interaction term of experiment and block was added. Random effects evaluated were subject, auditory stimulus and target objects (note that the same words and objects were used in both experiments). It was found that the model that best fit the data included random intercepts of subject and stimulus and a slope of block on subject ($\chi^2(2) = 487.54$, $p < 0.001$).

This model revealed a significant main effect of block (participants' accuracy non-surprisingly increased through training; B = 1.69, SE = 0.10, z = 16.11, $p < 0.001$). Moreover, accuracy was significantly lower in Experiment 3 than in Experiment 2 (B = 1.52, SE = 0.05, z = 27.78, $p < 0.001$); there was no interaction between block and experiment ($p = 0.743$). Overall, these results indicate that participants paid attention to the statistics included in the pre-training (at least in Experiment 3), and that they significantly influenced participants' response selections during training. Nevertheless, it is also important to point out that participants in Experiment 3—despite a lower accuracy starting point—still reached close to ceiling levels at the end of training (i.e., they caught up in their learning of the word-object-mappings, even though the interaction

did not reach significance). This may speak to participants' ease of learning symbol-object-mappings when symbols are auditory words. No analyses of the testing accuracy data were conducted due to participants' overall high levels of performance.

### 3.10.2 *Eye-tracking*

A similar analytical approach was used as for comparing accuracy data across experiments. First, training data were analyzed across experiments. To do so, proportion of looks to the featured and baseline foils were computed between 300 and 2000 msec of all correct training trials that included the featured competitor. Experiment was contrast coded (Experiment 2: +0.5; Experiment 3: -0.5). In addition, previous fixed effects (object type, block) and all possible interaction terms were added to analyze eye-tracking training data. The model that best fit included a random intercept of subject and a slope of the interaction of block and object type ($\chi^2(9) = 38.51$, $p < 0.001$).

It was found that there was a significant effect of experiment (B = -0.020, SE = 0.003, $t(3536) = -7.14$, $p < 0.001$). More specifically, participants were found to look more to foil objects in general in Experiment 3; this is likely due to the higher levels of "insecurity" in Experiment 3 than in Experiment 2, where participants' accuracy was high even during block 1 of training. Moreover, there was a main effect of block (B = -0.021, SE = 0.002, $t(44) = -9.37$, $p < 0.001$), indicating a decrease in looks over time, as well as a main effect of object type (B = 0.018, SE = 0.004, $t(37) = 4.17$, $p < 0.001$) as a result of increased looks to the featured competitor (likely mostly driven by Experiment 3). The only interaction that reached significance was the one of experiment and object type (B = -0.026, SE = 0.006, $t(3495) = -4.66$, $p < 0.001$), confirming the results from the previous analyses (for each experiment separately):

The effect of object type is significant in Experiment 3 but not in Experiment 2. Moreover, the interaction between experiment and block was marginal (B = 0.007, SE = 0.003, $t(3468) = 1.90$, $p = 0.057$; all other interactions: $p > 0.200$).

The same statistical approach was taken for participants' eye movement data at testing. For *in-featured* testing trials, the model that best fit the looks in the window of interest (300-2000 msec) included a random intercept of subject with a random slope of object type ($\chi^2(2) = 14.40$, $p < 0.001$). The main effect of experiment was significant (B = -0.034, SE = 0.004, $t(2320) = -8.37$, $p < 0.001$), indicating slightly more looks in Experiment 3 than in Experiment 2 (consistent with what was found during training). Additionally, the main effect of object type reached significance (B = 0.016, SE = 0.005, $t(49.5) = 3.07$, $p = 0.004$): Participants' overall looks were higher to the *in-featured* competitor than the included *out-foils*. Finally, the interaction of experiment and object type reached significance as well (B = -0.025, SE = 0.008, $t(2312) = -3.17$, $p = 0.002$), as the effect of object type was much more pronounced in Experiment 3 than in Experiment 2.

For *out-featured* testing trials, the model that best captured the data was the one including only a random intercept of subject but no random slopes ($\chi^2(1) = 2.04$, $p = 0.153$). As before, object type was split into two contrast: the comparison of looks to the *out-featured* and the *in-foil* as well as the comparison of looks to the *out-featured* competitor and the *out-foil*. The only factor that reached significance was experiment, again the result of overall more looks in Experiment 3 than in Experiment 2 (B = -0.01, SE = 0.003, $t(1873) = -3.83$, $p < 0.001$). In addition, the comparison of looks to the *out-featured* and the *in-foil* was marginal (B < -0.01, SE = 0.002, $t(1833) = -1.66$, $p = 0.096$), indicating a tendency for higher looks to the *out-featured*

competitor than the *in-foil*. This was not true for the comparison of the looks to the *out-featured* competitor and the *out-foil* (B < -0.01, SE = 0.002, $t(1833) = -1.92$, $p = 0.234$).

### 3.11 Discussion of Experiments 2 and 3's comparison

The results from the cross-experimental analyses overall confirm the findings from the analyses including data from Experiment 2 or 3 only. One important addition revealed by this combined analysis is the main effect of object type during training and *in-featured* testing trials. This suggests that participants looked slightly more to the featured object than *in-foil*s (training) or *out-foil*s (testing) in both Experiment 3 as well as Experiment 2, even as this effect was too subtle to be detected in the analysis of Experiment 2 only. That means, although pruning was extremely quick, there may be still small traces of the pre-training in participants' looks. Importantly, this appears to be only true for *in-featured* competitors (increased looks to the *out-featured* competitor at testing are not as robust). The co-occurrence of the *in-featured* competitor with the target word during training appears to be sufficient to "protect" some remaining associations of words and *in-featured* competitors.

This is surprising, and may suggest that pruning of incorrect associations is not as "complete" as previously thought after all. As a result, it may be interesting to further explore how "functional" subtle connections with the featured competitor are after repeated training (including feedback) that it is not the correct mapping. For example, one additional experiment may explore if these (pruned) connections to the featured competitors are more easily re-acquired if necessary (also see Experiment 6 for further exploration of this suggestion).

## 3.12 Question 1: Summary and discussion

### 3.12.1 *Summary/discussion*

The goal of Question 1 was to address whether incorrect associations between words and objects are pruned as the result of supervised and/or unsupervised learning. For this purpose, participants were first trained on two referents for each word, and subsequently tested whether the incorrect meaning was maintained even after receiving consistent feedback that the word only had one correct meaning after all.

Three experiments (Experiments 1-3) were conducted to address this question. They found that pruning of incorrect associations proceeded relatively quickly. By the end of training or even during training, participants no longer activated the incorrect referent that they had been previously trained on, if the secondary meaning was less strongly associated with the word than its primary one (Experiment 1-2). However, at the same time, the secondary meaning was not fully suppressed if it was more strongly associated with the word during pre-training than the primary one (Experiment 3), even as participants learned word-object-mappings within approximately two blocks of training.

At the same time, across experiments, evidence emerged that subtle, latent incorrect associations might be maintained long after they can be detected via eye-tracking. Support for this comes from participants' selection of the *out-featured* competitor in the no-correct testing trials at the end of Experiment 1, and the combined analysis of Experiments 2 and 3, where the main effect of object type reached significance.

Interestingly, pruning was more effective when unsupervised statistics did not support an association between a word and referent: When the featured competitor was assigned to the opposite cluster than the word (the *out-featured* competitor), participants never activated words

at testing. This suggests that receiving feedback was not sufficient for pruning incorrect associations. One caveat in this interpretation might be that—as required for the eye-tracking design—participants' accuracy during learning was often very high. As a result, participants did not receive much feedback after making an incorrect response, but most often received confirmatory feedback after selecting the target object. This stands in contrast to the pigeons from Roembke et al.'s (2016) study, where learning was much slower and negative feedback more frequent.

Moreover, there is tentative evidence that no sleep is required to prune incorrect associations (Experiment 1). However, it is also clear that the one study that addressed this question was not a sleep design that controlled amount of time between learning and sleep. Thus, this conclusion should not be seen as definitive, but rather as a first step in better understanding sleep's potential role in pruning incorrect associations during word learning.

### 3.12.2 *Limitations*

There are a number of limitations that apply to all experiments conducted: First, participants were all undergraduate students. As a result, the learning rate was high (and, as previously described, this was also an important prerequisite for the eye-tracking design employed). Thus, these experiments cannot address how pruning proceeds in children or adults from different backgrounds (e.g., older adults).

Second, it is possible that looks to the featured competitors at testing were not detectable due to the smaller number of trials used for analysis (in comparison to training trials). However, this hypothesis is unlikely, given that numeric differences between looks to the different foil types were very small if they existed at all.

Third, the visual world paradigm and eye-tracking were used to detect subtle differences in how different word meanings were activated during processing. However, this methodology may make it difficult to measure differences in activation strength over shorter periods of time. This is the case, as eye-tracking—as a non-continuous measure of activation—relies on the averaging across many trials. As a result, it is possible that participants looked slightly more to the featured competitor, even if it did not appear in their averaged eye movements. In addition, it should be noted that eye-tracking is a relatively conservative measure of activation strength, as only trials in which the correct object was selected are analyzed. Trials in which the featured competitor was selected—the strongest evidence that the secondary meaning was activated—are not considered (Roembke & McMurray, 2016). This is why participants were found to select featured competitors more than the baseline foils during block 1 of training in Experiment 2, even as no differences in looks were detected.

In summary, it can be concluded that incorrect meaning associations between words and objects are pruned; however, it is still unclear what influences how quickly this process proceeds and how "complete" it is. In addition, converging evidence from a different methodology might be needed to confirm these results, given the limitations of the current design employed.

# CHAPTER 4: EXPERIMENTS 4 AND 5

## 4.1 Question 2: Overview

The goal of Question 2 was to investigate how pruning operates in other learning domains. For this purpose, two experiments were conducted (Experiments 4-5). In Experiment 4, participants were taught mappings between written words and objects, thus addressing how pruning may be a part of reading. In Experiment 5, participants learned mappings between non-linguistic symbols and objects to study the domain-generality of the pruning mechanism. To address the importance of pruning beyond spoken word learning, the design of both experiments was closely matched with the one of Experiment 2.

## 4.2 Experiment 4: Overview

The goal of Experiment 4 was to investigate whether pruning is also operative during the learning of novel written word forms. From mid-childhood onward, hundreds of new words are acquired incidentally though reading (Joseph & Nation, 2018; Nagy, Herman, & Anderson, 1985): Children no longer learn to read, but rather read to learn. They are thought to learn the spellings and meanings of new words through their independent reading (Share, 2011), requiring multiple exposures to develop their full understanding (Joseph & Nation, 2018).

Many theories of written word recognition postulate that as learners become more adept at reading, they directly activate the semantic content of a word instead of activating its phonology first (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996). More specifically, skilled readers—in contrast to beginner readers—are hypothesized to activate a word's meaning directly from its written representation. Thus, during this phase of reading, there is a similar problem of learning

to map input to meanings as in spoken word learning. During this process, pruning of associations with incorrect meanings might be a critical part of vocabulary acquisition.

Pruning may also be important in reading for automatic word recognition. Importantly, automaticity in written word recognition could represent a “bottleneck” of students’ literacy acquisition: Even as they know how to read a word, they may not be able to activate it quickly enough in-the-moment (Roembke, Hazeltine, Reed, & McMurray, 2018; Stanovich, 1980). As posited in spoken language (McMurray et al., 2012), it is possible that connections between words and incorrect meanings need to be minimized to allow for this process.

To investigate pruning of incorrect associations during reading, this experiment paired written words with objects to closely match the design that was used for auditory words. It should be noted that this design does not mimic how written words are acquired during reading, which is usually thought to be more implicit involving text alone. However, there is no reason why the mechanism of learning should be different: Cross-situational/co-occurrence statistics between words and meanings can be tracked even in the absence of visual referents (Siskind, 1996). Moreover, by using the same design for both auditory and written word learning, we can ask if the same mechanism is operative during both.

To summarize, Experiment 4 used the same experimental set-up as Experiment 2. However, words were presented orthographically at the center of the screen. To incorporate eye-tracking, words were covered up with a mask after a short exposure in all parts of the experiments but the pre-training. No mask was employed during pre-training to facilitate the encoding of the novel written words and guarantee that participants acquired the associations between words and objects.

## 4.3 Method

### 4.3.1 *Participants*

Forty-five monolingual native English speakers with normal or corrected-to-normal vision participated in this experiment. They received course credit for participation. Two participants did not complete the full experiment due to time limits; their data were included in all analyses (when possible). The study was completed in accord with an IRB approved protocol.

### 4.3.2 *Stimuli and design*

The same stimuli and design were used as in Experiment 2. The only difference was that words were presented in written form (see Table 6 for overview of words in written form). As a result, there was only one exemplar per word.

Each word was assigned both a target object as well as a featured competitor. The featured competitor of one word was the target of another and vice versa. As before, the experiment consisted of the following portions: pre-training, three blocks of training, testing, no-correct testing and the word recognition section. During pre-training, each word was paired with its target object on 60% and its featured competitor on 40% of trials. In training, words were separated into two clusters, and foil objects could only be selected from the pool of objects that were assigned to the same cluster (i.e., the target objects of the other words assigned to the same cluster). Clusters were created, so that for half of the words the featured competitor could be one of the foils (the *in-featured* competitor) and for half of the words the featured competitor was never one of the foils (the *out-featured* competitor).

At the end of the experiment, participants completed the real-time word recognition section, the results of which will be discussed as part of Question 4. Due to an error in the

original design of this section, a redesigned trial make-up (described in Chapter 6) was used for 20 of the participants; only data of those participants will be analyzed in the word recognition section.

### 4.3.3 *Procedure*

The same procedure was used as in Experiment 2 with the exception of the presentation modality of the word. As before, a trial started with the presentation of a blue circle in the middle of the screen as well as the object(s) presented in the corners of the screen (there was only one object on the screen during pre-training, but four objects during all other parts of the experiment). After 1050 msec to inspect the object(s), the circle turned red to indicate that participants could click on it to see the word. When the participant clicked on it, the red circle disappeared and the target word was presented in the middle of the screen. During pre-training, the written word stayed on the screen until participants clicked on the present object. This was done to allow for participants to encode the word well in order to form robust word-object-mappings.

During all other phases of Experiment 4, the written word was presented for 75 msec only. Subsequently, a mask ('#########') was presented for 100 msec in the center of the screen. This procedure was chosen to guarantee that participants would start making eye movements after the word presentation and to equalize presentation conditions to the auditory modality. Participants wore headphones to receive auditory feedback during training, indicating whether they made a correct or incorrect response.

### 4.3.4 *Eye-tracking recording and analysis*

As before, eye movements were recorded using an SR Research Eyelink 1000 head-rested eye-tracker operating at 250 Hz with the exception of pre-training. The same method of analysis were used as in Experiment 1.

## 4.4 Results

The data of one participant were excluded from analysis, as their accuracy never exceeded 35% (chance was at 25%), suggesting that they did not follow the task protocol. In addition, one participant's eye movements were atypical (e.g., their target fixations did not follow the normal logistic trajectory); thus, their eye movements were excluded from all eye-tracking analyses.

### 4.4.1 *Training*

As seen in Figure 14, participants learned the written word-object-mappings rapidly, and reached levels close to ceiling by block 2 of training. However, performance appears to be lower than when learning auditory words (Experiment 2), where accuracy was above 80% in block 1. I asked whether participants' object selections in block 1 reflected the co-occurrence statistics they were trained on in pre-training. On *in-featured*

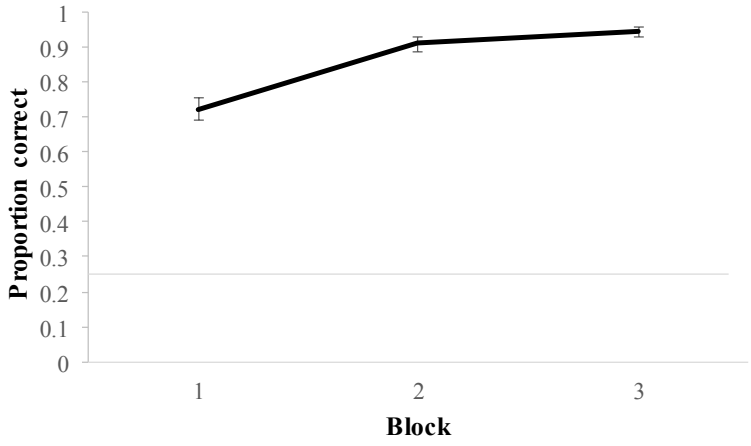


Figure 14: Proportion correct during training blocks of Experiment 4. Error bars indicate standard error of the mean.

trials of block 1, participants selected the target on 70% of all trials, the *in-featured* competitor on 13% of all trials and one of the baseline *in-foils* on 9% of all trials. Proportions were converted into log-odds ratios, asking whether the *in-featured* competitor was selected more

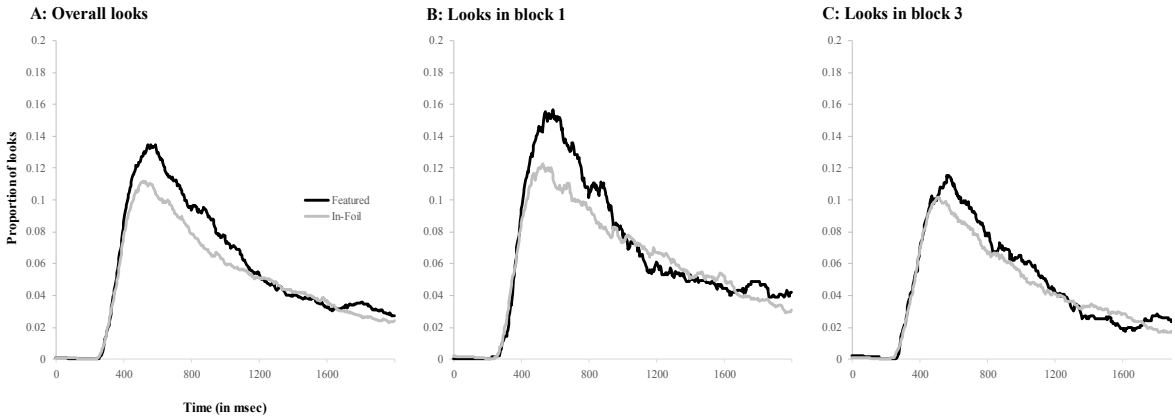


Figure 15: Overall looks during in-featured trials of training in Experiment 4 (Panel A), and during block 1 (Panel B) and block 3 (Panel C) of in-featured trials. Looks to the target are not depicted.

often than a baseline foil. A two-tailed one-sample t-test revealed that participants were significantly more likely to select the *in-featured* competitor than one of the *in-foils* if they did not select the target on *in-featured* training trials ($t(43) = 2.27$, $p = 0.014$). Again, this supports the notion that participants paid attention during pre-training and that they learned the incorrect associations they had been exposed to.

Subsequently, eye movements were analyzed during *in-featured* training trials, on which the target was selected. They are presented in Figure 15: Panel A depicts overall looks during *in-featured* training trials; participants look more the *in-featured* competitor than the baseline foils. Moreover, this appears to also be true for looks in block 1 (Panel B) and block 3 (Panel C) of training, though the overall number of looks was lower at the end of training. To investigate this statistically, proportion of looks between 300 and 2000 msec was calculated. A linear mixed effects model with fixed effects of block (centered), object type (transformed; *in-featured* = 0.5, *in-foil* = -0.5) as well as their interaction term were used to analyze participants' eye movements.

Nested models were compared using a chi-square difference test. We report the chi-square statistics comparing the model that was ultimately used to the next most complex model (the last comparison that would have reached significance). If none of the more complex ones reached significance, we report the comparison of the least complex model tested and the next complex one.

The model that best fit the data and still converged included a random intercept of subject and stimulus with no random slopes ($\chi^2(1) = 12.23$, $p < 0.001$). The main effect of block reached significance (B = -0.01, SE < 0.01, $t(1975) = -6.56$, $p < 0.001$), indicating that participants made less eye movements as the experiment progressed. In addition, there was a significant effect of object type (B = 0.01, SE < 0.01, $t(1975) = 3.15$, $p = 0.002$): Participants looked more to the *in-featured* object than the baseline (*in-foil*) throughout training. Importantly, the interaction of object type and block did not reach significance (B < -0.01, SE < 0.01, $t(1975) = -0.54$, $p = 0.588$), suggesting that this *in-featured* advantage was stable across learning. This is inconsistent with a straightforward account of pruning, which would predict a quick reduction of looks to the *in-featured* object over the *in-foils* as participants' unlearning of the incorrect associations proceeds.

### 4.4.2 *Testing*

For both *in-featured* and *out-featured* testing trials, proportion of looks were calculated between 300 and 2000 msec on correct trials. First, looks to the *in-featured* object in comparison to looks to the *in-foils* during *in-featured* testing trials were analyzed. The only fixed effect included was object type (transformed; *in-featured* = 0.5, *in-foil* = -0.5). The random effect structure that fit the data included a random intercept of subject ($\chi^2(1) < 0.01$, $p = 1$). The main

effect of object type did not reach significance (B < 0.01, SE < 0.01, $t(655.1) = 0.15$, $p = 0.879$), indicating that participants were equally likely to look at the *in-featured* competitor and at baseline objects (*out-foils*) during testing.

For *out-featured* testing trials, the fixed effect of object type was formulized with two contrasts, one of which compared looks to the *out-featured* competitor and the *in-foil* and one of which compared looks to the *out-featured* competitor and the *out-foil*. The model with only a random intercept for subject best captured the data ($\chi^2(1) < 0.01$, $p = 0.934$). Participants were not more likely to look at the *out-featured* competitor than either the *in-foil* (B < -0.01, SE < 0.01, $t(997.98) = -1.16$, $p = 0.247$) or the *out-foil* (B < 0.01, SE < 0.01, $t(997.98) = 0.50$, $p = 0.619$). Together, eye movements during the testing section of Experiment 4 suggest that participants may have pruned all incorrect associations by block 4.

Second, for the no-correct testing section, one participant was excluded from analysis, as they did not comply with instructions (i.e., s/he selected the same location on repeating trials more than eight times). For each trial type (control, *in-featured* and *out-featured*), response ratios were calculated and log scaled to attain normality. Subsequently, they were

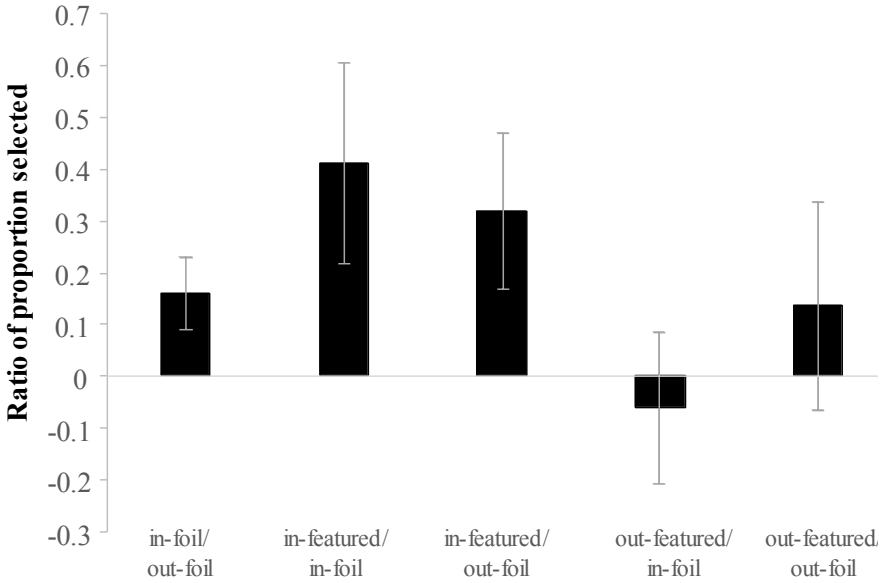


Figure 16: Means of ratios computed as part of the no-correct testing session of Experiment 4. Error bars represent standard error of the mean.

tested against zero using a one-sample t-test, thus asking whether the numerator of the ratio was more often selected than the denominator. The resulting data are presented in Figure 16: Means tended to be higher when the numerator of a ratio was an in-cluster foil (*in-foil* or *in-featured*). The one-sample t-test showed that participants were more likely to select the *in-foil* than the *out-foil* in control trials that included neither an *in-featured* nor an *out-featured* competitor ($t(38)= 2.28$, $p = 0.014$): Participants selected one of the *in-foils* on 54% of the trials and one of the *out-foils* on 46% of all trials. Moreover, there were significant differences in the *in-featured* trials as well; participants preferred the *in-featured* competitor over the *in-foil* ($t(38)= 2.13$, $p = 0.020$) as well as the two included *out-foils* ($t(38)= 2.11$, $p = 0.021$). Participants were not more likely to choose the *out-featured* competitor than either of the *in-foils* ($t(38)= -0.41$, $p = 0.659$) or the *out-foil* ($t(38)= 0.68$, $p = 0.250$). Together, these results suggest that participants were sensitive to co-occurrence statistics, but that they used them to *build* associations between words and objects instead of pruning them (thus explaining the preference of even the *in-foil*).

## 4.5 Discussion of Experiment 4

In Experiment 4, participants looked more to the *in-featured* competitor than a baseline foil throughout training blocks. Thus, these results indicate that participants' pruning of incorrect associations was less robust than pruning in spoken words. This contrasts with Experiment 1 and 2, where we found no increased looks to the *in-featured* competitor by the second block of training. This is potentially due to participants' lower overall learning rate of the written word-object-mappings. Visual comparison of Experiment 4 where participants learned to map written words onto objects to Experiment 2 where participants were required to map auditory word forms onto objects suggests that the learning of the former may have been harder: Performance

on block of 1 in training was at 83% in Experiment 2 but at 72% in Experiment 4. This is consistent with the idea that the pruning of incorrect associations is correlated with the strength of correct associations and/or might be facilitated by participants' acquisition of the correct mappings.

The potential difference in learning rate between Experiment 1, 2 and Experiment 4 by itself is surprising: Human adults have had extensive experience with auditory as well as written words. If anything, it could be argued that written words might be more easily mapped onto a referent because they are more stable in time whereas auditory information is fleeting, though this advantage was limited due to masking of the written words. Nevertheless, participants had an easier time acquiring auditory word-object-mappings. One of the reasons for this may be that written nonwords are less automatically recognized, thus slowing down the mapping process to an object.

As mentioned previously, the results of Experiment 4 suggest that participants may have *built* associations between words and "to be pruned" competitors. Support for this comes from the no-correct testing session: Participants were significantly more likely to select the *in-featured* competitor over the included *in-foil* and *out-foils*. At first glance, one might argue that this finding is best explained by associations between a word and its *in-featured* competitor never were completed pruned. However, participants were also more likely to select an *in-foil* over *out-foils* in control trials of the no-correct testing session. In contrast to the *in-featured* competitor, *in-foils* did not co-occur with a word during pre-training—but associations with both of these items (*in-foil* and *in-featured*) should have been pruned by training.

Importantly, the *in-featured* competitor as well as *the in-foil* co-occurred with the word during training. Even the supervised training trials contain cross-situational statistics that could

aid learning. This is because each foil object in the same cluster had an approximately equal chance of being present on each trial. As there were only eight objects in each cluster (one of which was the target) and four response options on each trial, each foil object co-occurred with the word on three out of seven trials.

Moreover, participants did not choose the *out-featured* competitor more often than other baseline objects during no-correct testing trials: Even though the *in-* and the *out-featured* competitor were both paired with their respective word during pre-training (the only difference between them is what training cluster they were assigned to), the *in-featured* competitor also co-occurred with the word during training, whereas the *out-featured* competitor did not. Thus, the incorrect association with the out-featured competitor was not supported by unsupervised statistics.

Together, these findings indicate that unsupervised co-occurrence statistics influenced the strength of word-object-associations. The building of associations indicates that participants may (equally) weigh information from supervised training (feedback) as well as unsupervised statistics (which words and objects co-occur). It should be noted that both the *in-* and the *out-featured* competitor were the target for a separate word, and co-occurred with it on all training trials.

Overall, Experiment 4 is further evidence for less robust pruning both in the presence (*in-featured*) and absence (*out-featured*) of feedback. Moreover, these data further highlight that interactions between supervised and unsupervised learning may be more complex than previously thought, with unsupervised statistics out-weighing supervised ones at least under some circumstances.

## 4.6 Overview Experiment 5

The goal of Experiment 5 was to investigate how pruning operates outside a word learning task: Is pruning operative in a completely non-linguistic task? Is pruning less robust if there is no existing network in which novel mappings have to be integrated?

In Experiment 5, participants were taught non-linguistic mappings between abstract visual symbols (which were used as the equivalent to the words; see Figure 17) and novel objects. Abstract visual stimuli were selected as symbols (in contrast to other options, such as non-linguistic sounds) to match the experimental set up used by Roembke et al. (2016) in pigeons. It was predicted that people would also be able to prune incorrect associations between non-linguistic symbols and objects, supporting the notion that pruning is a component of a domain-general learning mechanism.

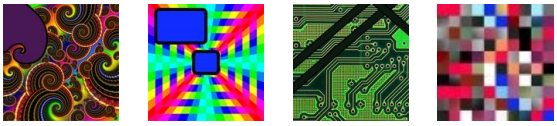

Figure 17: Examples of non-linguistic visual symbols used in Experiment 5 as well as the visual mask (last picture to the right).

## 4.7 Method

### 4.7.1 *Participants*

Forty-one monolingual, native speakers of English with normal or corrected-to-normal vision were recruited to participate. All were students at the University of Iowa and received course credit for participation. Participants underwent informed consent and the study was conducted in accord with an IRB approved protocol.

### 4.7.2 *Stimuli and design*

The same novel objects were used as in Experiments 2-4. Instead of words, novel visual symbols (see Figure 17) were developed to act as word-equivalents based on the previously used pexigrams with pigeons (see Roembke et al., 2016; Wasserman, Brooks, & McMurray, 2015). The symbols were edited to minimize nameability to discourage paired associative learning (i.e., participants naming symbols and associating that label with an object). There was only one exemplar per visual symbol. In addition, a visual mask was created by scrambling patches of pixels of all symbols to create an effective masker to the previously presented stimulus (Figure 17). Experiment 5 did not include a real-time word recognition section at the end of the experiment, as there are no phonological or orthographically competing words for visual symbols. Apart from the change in stimuli, all other design choices were the same as in Experiments 2 and 4.

### 4.7.3 *Procedure and eye-tracking recording and analysis*

The same procedure was used as in Experiment 4. The same eye-tracking apparatus and analysis method was used as in all previously reported experiments.

## 4.8 Results

The data of two participants were excluded from analysis, as their accuracy remained at chance throughout the experiment (their average training accuracy remained below 30%).

4.8.1 *Training*

Participants' training accuracy was plot (see Figure 18): It appears that accuracy was lower when symbols were visual and not auditory: Average accuracy on block 1 was 54%, and participants did not reach a learning plateau at the end of training (block 3 M = 90%). In contrast, average accuracy in Experiment 2 was 83% and 72% in Experiment 4.

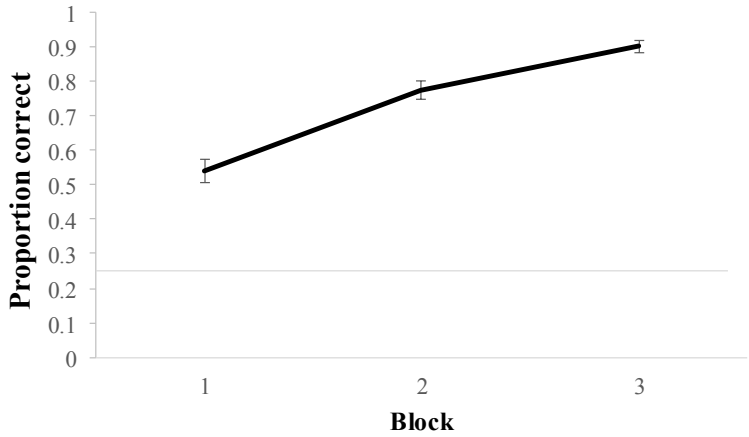

Figure 18: Proportion correct during training blocks of Experiment 5. Error bars indicate standard error of the mean.

To determine if people were sensitive to the pre-training statistics, response selections on featured trials of block 1 of training were analyzed: The target was selected only on 52% of *in-featured* training trials, the *in-featured* competitor was selected 21% of the time and one of the *in-foils* 14% of the time. To investigate the difference between *in-featured* competitor and *in-foil* statistically, log-odds ratios were calculated and compared against zero using a one-sample t-test. As anticipated, the higher selection of the *in-featured* competitor than the *in-foils* was significant ($t(38) = 3.10$, $p = 0.002$). This suggests that participants were sensitive to the pre-training statistics.

Figure 19A presents eye movements during correct *in-featured* trials of training. It can be seen that the participants made more looks to the *in-featured* competitor the *in-foil* early on during processing. Moreover, increased looks to the *in-featured* competitor remained present both during block 1 as well as block 3 of training (Panels B and C of Figure 19). Eye movements during *in-featured* training trials were investigated by calculating the proportion of looks to the

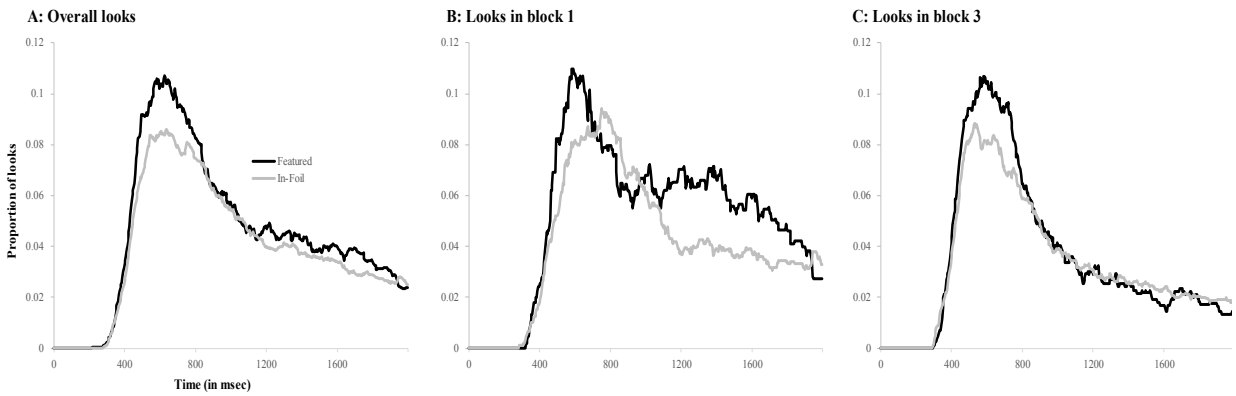


Figure 19: Overall eye movements during training (Panel A), block 1 (Panel B) and block 3 (Panel C) of Experiment 5's training.

*in-featured* and baseline foils between 300 and 2000 msec for correct trials. The fixed effects of block (centered) and object type (transformed; *in-featured* = 0.5, *in-foil* = -0.5) were included. Comparisons of nested models were completed, and the chi-square difference test was used to select the least complex random effect structure that was needed to fit the data. A model with random intercepts for subject and stimulus (no slopes) best fit the data ($\chi^2(1) = 20.03$, $p < 0.001$). There was a significant main effect of block (B = -0.01, SE < 0.01, $t(1603) = -4.82$, $p < 0.001$); this is consistent with participants' number of looks decreasing as the experiment progressed. Importantly, there was also a significant effect of object type (B = 0.01, SE < 0.01, $t(1603) = 2.32$, $p = 0.021$), whereas the interaction of block and object type did not reach significance (B < -0.01, SE < 0.01, $t(1603) = -1.29$, $p = 0.199$). This indicates that participants had not significantly pruned incorrect associations to *in-featured* competitors by the end of training (see

also Panel A of Figure 19). On top of that, the absence of a significant interaction may suggest that participants did not prune at all, as increased levels of looks to the *in-featured* object remained stable throughout training.

### 4.8.2 *Testing*

Figure 20A shows higher looks to the *in-featured* than baseline foils during *in-featured* testing trials. Similarly, there also appears to be more looks to the *out-featured* competitor than either baseline foil (Panel B of Figure 20). To investigate eye movements during testing statistically, proportion of looks were calculated between 300 and 2000 msec on correct *in-featured* and *out-featured* testing trials. Models always included a fixed effect of object type; this was transformed (*in-featured* = 0.5, *in-foil* = -0.5) for *in-featured* testing trials.

For *in-featured* testing trials, the model that best fit the data included random intercepts of subject and stimulus without any random slopes ($\chi^2(1) = 5.46$, $p = 0.019$). There was no significant effect of object type (B < 0.01, SE < 0.01, $t(566) = 1.51$, $p = 0.133$): Despite the small

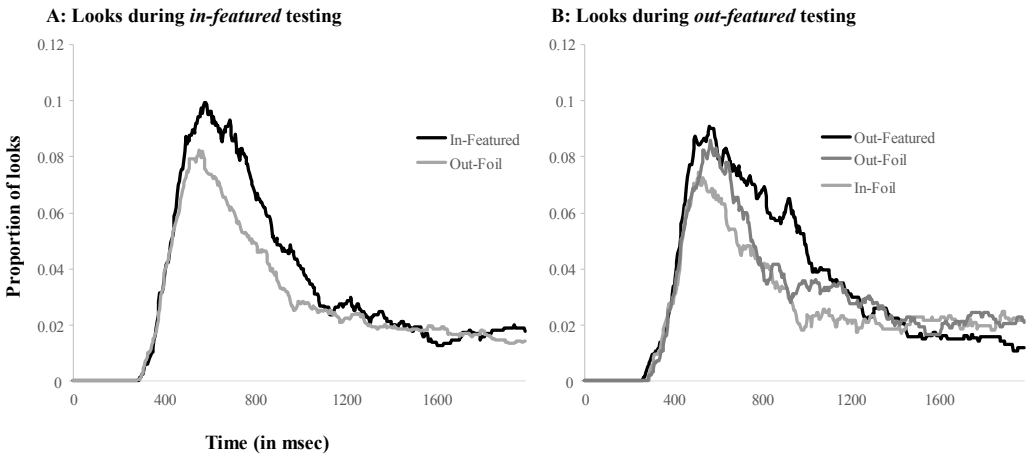


Figure 20: Eye movements during in-featured (Panel A) and out-featured (Panel B) testing trials of Experiment 5. Looks to the target object are omitted.

difference present in Figure 20A, participants equally looked to the *in-featured* competitor and the *out-foils* (used as the baseline).

For *out-featured* testing trials, the same contrast coding scheme was used as before with one comparison of *out-featured* and *in-foil* as well one comparison of *out-featured* and *out-foil* looks. The model that best fit the data included a random intercept for subject but no random slopes ($\chi^2(1) = 2.54$, $p = 0.111$). Participants looked marginally more to the *out-featured* object than the *in-foil* (B < -0.01, SE < 0.01, $t(886.1) = -1.89$, $p = 0.059$), but this slight *out-featured* advantage was not evident when comparing looks to the *out-foil* (B < -0.01, SE < 0.01, $t(886.1) = -0.11$, $p = 0.910$). This suggests that *out-featured* and *out-foils* were treated similarly, even as only the out-featured competitor had been paired with the word during pre-training.

For no-correct testing, no participant selected the same response location more than eight times; thus, all participants were included for the no-correct testing analyses. Ratios between response selections were created separately for each trial type. Subsequently, they were converted into log-odds ratios and compared to zero using a one-sample t-test. The resulting means are plotted in Figure 21. The only contrast that reached significance was within trials that

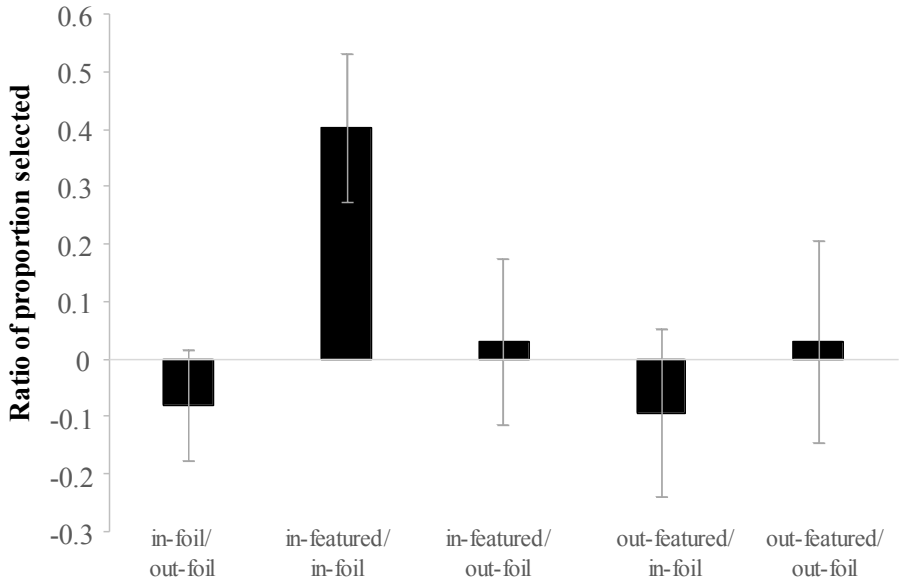


Figure 21: Means of ratios computed as part of the no-correct testing session of Experiment 5. Error bars represent standard error of the mean.

included the *in-featured* competitor: Participants were significantly more likely to select the *in-featured* object than an *in-foil* ($t(38)= 3.12$, $p = 0.002$); however, they were not more probable to choose the also included *out-foil* ($t(38)= 0.20$, $p = 0.422$). In addition, participants were not more likely to select the *out-featured* competitor than the *in-foil* ($t(38)= -0.65$, $p = 0.739$) or the *out-foil* ($t(38)= 0.16$, $p = 0.436$).

## 4.9 Discussion of Experiment 5

The goal of Experiment 5 was to investigate pruning of incorrect associations in a non-linguistic task. There were two major findings: First, learning in Experiment 5 was slower than in all other experiments that used the same pre-training statistics (Experiment 1, Experiment 2, Experiment 4; based on visual comparison). Second, pruning was less robust than in an auditory word learning task (Experiment 1, Experiment 2). Moreover, as in Experiment 4, participants activated the *in-featured* competitor more than another *in-foil* during training; however, this increased consideration was no longer observable at test. Nevertheless, participants were more likely to select the *in-featured* competitor than any other foil type during no-correct testing, indicating that—even though increased activation was no longer evident during eye-tracking in testing trials—the associative history between the *in-featured* competitor and the word was somehow retained.

Overall, these findings are consistent with what was observed in previous experiments, again highlighting the importance of unsupervised learning in combination with supervised one. This supports the notion that similar, domain-general processes may be at work during the learning of linguistic as well as non-linguistic mappings, particularly as results were similar across Experiment 4 (written words as symbols) and Experiment 5 (abstract non-linguistic

symbols). Any differences in the patterns of pruning/strengthening associations may be more driven by the learning rate of the mappings rather than the linguistic status of the materials per se.

## 4.10 Question 2: Summary and discussion

### 4.10.1 *Summary and discussion*

The goal of Question 2 was to determine whether pruning of incorrect associations operated differently, when mappings were not between auditory word forms and objects. For this purpose, I conducted two experiments, testing pruning of incorrect associations when symbols were written words (Experiment 4) or non-linguistic symbols (Experiment 5).

Learning rates for both types of mappings was lower than when symbols were auditory words, though performance was lowest for non-linguistic symbols (based on visual inspection of accuracy across Experiments 2, 4 and 5). Interestingly, even as learning was lower for non-linguistic symbols than for written words, overall findings for both experiments were comparable: Pruning was less robust than when mappings included auditory words as symbols. As a result, participants looked to the featured competitor above baseline at the end of training. At the same time, looks to the featured competitors no longer were significantly different from baseline at testing—suggesting that incorrect associations between words and objects were pruned or suppressed eventually.

Based on these findings, one may conclude that pruning of incorrect associations proceeds more quickly when learning is easier *or* that learning is easier when the pruning of incorrect associations happens more quickly. Of course, these conclusions remain tentative at this point.

### 4.10.2 *Limitations*

One important limitation of this section is that because the design used did not test the strength of incorrect associations after pre-training, it is unclear whether participants had learned the incorrect mappings (i.e., connections with the featured competitor) to the same extent in Experiments 4 and 5 as in Experiment 2 (the most direct comparison). It is possible that, because learning was lower for non-auditory word-object-mappings, this also applied to pre-training (also see Roembke, Wiggs, & McMurray, 2018). From participants' selections during block 1 of training, it is clear that pre-training influenced their performance in Experiments 4 and 5 (as the featured competitor was more often selected than a baseline foil); however, one cannot draw conclusions about the original overall strength of incorrect associations this way (as it is confounded with participants' learning rate).

In summary, Experiments 4 and 5 offer additional evidence that people are slower at learning mappings that do not include auditory words as symbols (Roembke, Wiggs, et al., 2018). This is surprising in the case of written words, as adult participants have extensive experience with these representations. Moreover, these results offer tentative evidence that there is a relationship between the amount of learning and the extent of pruning, though this relationship may not depend on whether material is linguistic but rather on how easily symbols are processed.

# CHAPTER 5: EXPERIMENT 6

## 5.1 Preface/disclaimer

Experiments 1-5 suggest a couple of clear trends. First, after training, within-cluster featured competitors (i.e., *in-featured* competitors that co-occurred with the word during training) were more likely to be considered than featured competitors in a different cluster. Evidence for this comes from no-correct testing (Experiments 3, 4 and 5), where participants tended to select the *in-featured* competitor above baseline, but not the *out-featured* competitor. This was surprising in that the *in-featured* competitors had ample opportunity for negative reinforcement, whereas the *out-featured* competitors did not.

However, the *in-featured* competitor also co-occurred with the word during training; this was not the case for the *out-featured* competitor. As a result, incorrect associations between the *in-featured* competitors and words might also get strengthened by the unsupervised learning of co-occurrence statistics (a form of cross-situational word learning: Siskind, 1996; Yu & Smith, 2007): More specifically, the fact that the target word and its *in-featured* object co-occurred might be enough to build an association between the two, even if a different object was rewarded. In contrast, a word and object not co-occurring (e.g., the word and the *out-featured* competitor) might result in the pruning of incorrect associations, even if the association never directly received negative feedback. Thus, pruning by feedback may not be sufficient to allow for the complete removal of incorrect associations. More broadly, this suggests both unsupervised and supervised mechanisms may work in parallel (and sometimes at crossed purposes).

Second, there was little evidence of increased looks to either of the incorrect competitors during testing. However, by the end of the experiments, participants tended to be more likely to

select a featured competitor than a baseline foil in the no-correct testing session. This raises the question of how no-correct testing session behavior is shaped by incorrect associations. There are three possibilities: (1) Incorrect associations with the featured competitor were not pruned by the end of training after all. Instead, the eye-tracking measures used to test increased activation of incorrect competitors were not sensitive enough to detect small, sub-threshold associations. This seems unlikely, as eye-tracking is widely assumed to be a more sensitive measure of participants' knowledge than their accuracy response. (2) Associations with incorrect competitors were pruned, but small traces remained that can influence future behavior. Such small traces might either exist in the primary association between a word and its featured competitor (i.e., the association that was pruned during trained) and/or in a secondary association. The secondary association could be a different form of associative memory, such as an episodic memory, that was preserved. (3) Incorrect associations were not pruned but rather might have been inhibited in-the-moment. As a result, they may be more likely to reappear at a later time point.

Unfortunately, one of the limitations of the previous experiments is that participants' knowledge of incorrect associations was measured very differently in different phases: Participants' acquisition of the incorrect associations was deduced from their response selections in block 1 of training. Pruning of incorrect associations was measured with a variant of the visual world paradigm, where the absence of a difference in eye movements to the featured competitor and baseline foils indicated a reduction in associative strength. Finally, in no-correct testing, participants' response selection when no target object was included was used to estimate associative strength of incorrect associations at the end of the experiments. As a result, it is hard to draw inferences about what the participant knows at each phase of the experiment. A common

measure at each phase of the experiment is needed to test how strong incorrect (and correct) associations are between words and objects.

More importantly, if sub-threshold associations do remain after pruning, the question becomes: How functional are they? Do these remaining incorrect associations facilitate relearning of a previous mapping (e.g., if the featured competitor must now be mapped to the word)? Or are they epiphenomenal to learning and reflect some other form of memory?

Thus, the goal of Experiment 6 was to address how functionally useful these sub-threshold associations may be, by using an improved experimental design. For this purpose, several methodological changes were made. During pre-training, words were equally paired with its eventual target word and the same incorrect competitor (the featured competitor; 50/50 split). As a result, each word was associated with two objects, but participants did not know which of the objects was correct. This change allowed us to test participants' pruning of incorrect associations in the absence of a "favorite". In addition, this scenario may be comparable to children being exposed to a consistent context (e.g., a kitchen) where items repeatedly occur together (e.g., the objects FORK and SPOON).

Most importantly, a new direct test of associative strength was introduced. This was administered at multiple phases (e.g., after pre-training, during training, after training, and at the end of the experiment). For the novel test, we used a yes/no paradigm in which each trial included one word and object. Participants were asked to indicate if the two matched (yes button) or mismatched (no button). This design allows people to indicate that multiple objects could go with a specific word (across trials) while avoiding biasing participants toward one answer or another. By testing incorrect associations directly, there was no need to deduce participants' acquisition of incorrect associations after pre-training. Similarly, results from the direct test can

be used after learning to see exactly how much these incorrect associations were suppressed or after relearning to see if they were reinstated.

The same training procedure was adopted as in previous experiments of this dissertation. However, participants were trained on only half of the words that were included in pre-training. This allowed us to assess the impact of decay: To what extent does training cause the strength of incorrect associations to decrease? In contrast, how much does associative strength change due to decay over time alone?[4] During training, featured competitors could appear as a foil for trained words. Participants completed two tests in the middle and right after training; this allowed for the direct comparison to participants' selections at pre-test for both trained and untrained words.

Subsequently, we tested if latent associations between the featured competitors and words can still shape future word learning. To do so, a novel relearning phase was added. In this phase, participants were taught to re-acquire the previously pruned associations (between the target and the featured competitors). Participants learned to map the featured competitor (test condition) or a baseline foil (control condition) on each word. This allows us to test how well incorrect associations are unlearned during the initial phase of training: If associations are not completely eliminated, they should be more easily re-acquired (compared to the control condition). Alternatively, associations might be too weak to influence word learning rate, or might even hinder relearning if they are below baseline.

There are several advantages of this design over the one used previously (Experiments 1-5 of this dissertation): All measures clearly map onto a prediction. This stands in contrast to the sometimes more difficult to interpret measures that were set up in Experiments 1-5, as

[4] A third route by which associations could be unlearned or lost are interfernce processes; however, this possibility will not be explored here.

Experiment 6 will combine a standard learning design with more explicit tests of the strength of word-object-associations.

It should be noted that the rest of this chapter is written such that it can be read independently of Experiments 1-5; this is the case, as I plan to publish Experiment 6 (in addition to a follow-up experiment that is not part of this dissertation) by itself. Due to the small effects observed in Experiments 1-5 and some design issues, I may or may not publish those results and if so, they will be published independently of this work. As a result, the following introduction will discuss a similar background as did Chapter 1, the general introduction.

## 5.2 Introduction

To acquire a new word, one has to learn its phonological form, its meaning, and an association between the two. Importantly, words-object-mappings are not isolated from each other but are part of a network: Words are not just associated with their meaning; they prime semantically related words (Apfelbaum, Blumstein, & McMurray, 2011; Huettig & Altmann, 2005; Yee, Blumstein, & Sedivy, 2008), they inhibit similar sounding words (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Paul A. Luce & Pisoni, 1998), and they must be linked to other knowledge such as spelling, articulation and syntax.

However, even if we consider only the word form-to-meaning portion of this network, such links are rarely one-to-one. During word learning, people may acquire associations between words and incorrect but frequently co-occurring referents or meanings. For example, a child might mistakenly associate the word FORK with both the kitchen items FORK and SPOON. Of course, these secondary (incorrect) associations (meanings) are suppressed or lost over

development/time. But how does this process—which we will refer to here as unlearning[5]—take place? Even after such unlearning, are incorrect associations between words and meaning maintained in some latent form? Do associations stick around (in some form) to influence later behavior, or do they fully disappear if no longer supported by the learning environment?

This issue of unlearning has been examined in the memory literature. In a seminal study, Ebbinghaus (1913) found that previously acquired, but subsequently forgotten lists of nonwords were later more easily relearned. A similar process has been described in the context of associative learning: A learned association between a stimuli and a response that has undergone extinction is more easily acquired in a retraining (Rescorla, 2001). In both cases, subsequent behavior was influenced by latent, sub-threshold associations that were previously forgotten (Ebbinghaus, 1913) or unlearned (Rescorla, 2001) to the point at which they were no longer measurable (also see Shiu & Chan, 2006).

The goal of this study is to investigate this unlearning and relearning process in the context of word learning. We ask how incorrect word-referent-mappings are lost, and whether latent, sub-threshold associations are maintained that can influence subsequent behavior. To address this question, we will first train participants on two meanings for each word; subsequently, only one of those meanings will receive positive feedback. In the relearning phase of the experiment, participants will be retrained on mappings they had been exposed to, thus addressing whether and, if so, how latent associations impact word learning.

[5] Unlearning refers here to the general process of eliminating an incorrect association. Thus, it should be considered synonymous to what was described as pruning in previous dissertation chapters.

### 5.2.1 *Frequency of incorrect word-meaning-associations*

Children make mistakes when learning novel words. Most frequently, naming errors are either overgeneralizations (Clark, 1973) or retrieval errors, and made by toddlers who know between 50 and 150 words (Gershkoff-Stowe, 2001). Whereas overgeneralization errors tend to be similarity-based (e.g., overextending the word BALL to all round objects), retrieval errors might indicate that a mapping between a word and a concept was not strong enough to suppress other existing associations (Gershkoff-Stowe, 2001). Unlearning such incorrect associations thus might be a critical part of vocabulary development.

Acquisition of partial word meanings, either correct or incorrect, is not limited to early vocabulary learning. In older children and adults, most word learning is from written text. In a study by Fukkink, Blok, and De Glopper (2001), children were read short passages including novel abstract and concrete words whose meaning was implied, but not explicitly stated. They found that participants' definitions of novel words often included false attributes, particularly definitions given by younger children and for abstract words. In addition, words' meanings, as understood by the children, were often tied to the context the item had been heard in. As a result, definitions were often partial, and did not generalize to other contexts the word could appear in. Similarly, analyzes of the types of errors that were made by children with language learning disability indicate that they had more difficulty gaining partial word knowledge during reading tasks than a control group (Steele, 2012). Together, these data suggest that unlearning incorrect associations between words and meanings may happen often during people's vocabulary acquisition.

### 5.2.2 *Evidence for latent associations in word learning*

Direct laboratory evidence for the maintenance of incorrect word meanings comes from cross-situational word learning, where information across trials is combined to deduce the correct word meaning (Siskind, 1996; Yu & Smith, 2007). In a typical cross-situational word learning experiment, participants are presented with one word and a number of objects on each trial. Whereas each trial by itself is ambiguous, words and their referents consistently co-occur across trials. Thus, by tracking word-object-co-occurrences, it is possible to acquire word meanings, even in the absence of feedback (Dautriche & Chemla, 2014; Roembke & McMurray, 2016; Roembke, Wiggs, et al., 2018; Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Yu & Smith, 2007; Yurovsky & Frank, 2015).

Recent data from cross-situational word learning suggest that, at least under some circumstances, several meanings or "hypotheses" are maintained for each word. For example, Roembke and McMurray (2016) manipulated co-occurrence statistics of words and objects so that each word consistently appeared with its target, but appeared also at above baseline level with a second object, its "featured" competitor. In a control experiment, no featured competitors were included. In addition to tracking participants' accuracy across blocks, eye-tracking was used as a more sensitive measure of weak associations between words and objects. Here, if people were maintaining multiple hypotheses, they may show increased fixations to the featured competitor, even while they are clicking on the target. Results showed exactly this pattern. In addition, learning (measured by accuracy) was poorer when words had a competitor that co-occurred with the word at above baseline frequency, than in the control condition where they did not (Roembke & McMurray, 2016). This suggests that people track multiple hypotheses per word and activate those alternatives in-the-moment. Moreover, the presence of such alternative

associations seems to slow overall learning. However, it is less clear how such competing meanings are unlearned or inhibited.

Such associations might be sub-threshold; that is, they can exist even in the absence of overt mistakes (e.g., despite high accuracy). Evidence that sub-threshold associations are maintained comes from a study by Yurovsky, Fricker, Yu, and Smith (2014), investigating participants' partial knowledge in the acquisition of word-object-mappings. In their study, participants were first taught a number of word-object-mappings using cross-situational word learning. Subsequently, for each subject, words were identified that were not acquired, and participants were given additional training on those. Learning of those words was compared to participants' acquisition of a novel set of words. Thus, this experiment asked whether participants were able to learn word-object-mappings more quickly if they had received prior training on them, even as their accuracy from that session indicated little learning. Yurovsky et al. (2014) found that in session 2, words that had not been acquired in session 1 were more easily learned than control words without prior training. These results suggest that not only do participants acquire sub-threshold associations but they can take advantage of them when given additional trials for learning.

The study by Yurovsky et al. (2014) provides evidence that sub-thresholds associations are a part of building correct associations. However, it does not answer whether sub-threshold associations persist after unlearning—and, if they do, to what extent they can drive behavior. During learning, it is easy to conceive how small associations between words and referents might be built, and how such connections could benefit later learning. It is much less clear if this would be the case if incorrect associations are actively unlearned.

Why might incorrect sub-threshold associations matter beyond misunderstanding a word's meaning? As indicated by Roembke and McMurray's (2016) data, alternative meanings are activated in-the-moment; this may cause slower learning and less efficient processing of words. Evidence for the latter comes from a connectionist model of word learning by McMurray, Horst, and Samuelson (2012). They found that speed of word recognition in the model was better predicted by the weakness of incorrect associations than the strength of correct ones. That means, unlearning incorrect meanings might be critical in allowing for speedy vocabulary development and language processing.

### 5.2.3 *Mechanisms of unlearning words*

What happens when an alternative association is lost (e.g., its effects are undetectable in a behavioral response)? One possibility is that old meanings are fully unlearned. That means, there no longer is any trace of the previous word-object-mapping. Under such account, one might expect that subsequent training on old (unlearned) word-object-mappings will not lead to a different performance (faster learning) than an entirely new set of mappings. Alternatively, an incorrect association could simply be suppressed—but not actually disappear—to allow for the acquisition of new meanings. Under such account, incorrect associations could be suppressed in-the-moment via real-time inhibition, even as they are retained. If this were the case, relearning incorrect meanings might be easier, as the association between word and object was never unlearned. Finally, incorrect associations could be unlearned to the extent that their associative strength is below baseline levels. As a result, subsequent relearning of incorrect word-object-mappings might be slower than if no previous exposure existed.

It should be noted that this study will not make strong assumptions as to the nature of the underlying mechanism of unlearning and/or how previously undetected associations reappear: Incorrect associations could be pruned, inhibited or unlearned by another mechanism. In contrast, the goal of the conducted experiment was to capture how unlearning proceeds in vocabulary acquisition, and whether remaining associations might continue to influence people's behavior, even as they are actively rejected as potential referents. In this context, we will conceptualize differences in learning (i.e., how well is a word known?) as differences in associative strength.

### 5.2.4 *The current study*

The current study addressed the following questions:

1. How quickly are incorrect associations unlearned in word learning?
2. How functional are sub-threshold associations in influencing subsequent word learning?

The experiment consisted of three phases: *pre-training, training* and *relearning*. Pre-training was a simple exposure learning task to teach two associations for each word. One of these associations constituted the later target object, and one the so-called featured competitor. Training used a supervised learning paradigm to build one of those associations (the one with the target object) and unlearn the other (the one with the featured competitor). In relearning, participants completed a supervised paradigm, where they learned a new association for each word that could either match or not match the one that was previously unlearned during training. We manipulated two factors: the identity of the new target object during relearning and the inclusion of a word during training. Both of these manipulations were within-subject but between-word.

First, the identity of the new target object (*target-identity*) was manipulated as follows: During relearning, a word was now either mapped onto the competitor that it had co-occurred with during pre-training (the featured condition) or it mapped onto a foil object that it was not paired with during pre-training (the baseline condition). This manipulation tested the extent of unlearning that had occurred during training: If unlearning of the incorrect association between a word and featured competitor was completed during training, learning of the new mapping should be equal across the featured and baseline condition. However, if small traces of the incorrect association remained, relearning of the mapping between word and the featured competitor should be easier than learning one between the word and a new foil object. Finally, if unlearning of the incorrect association resulted in below-baseline levels of the association between word and featured competitor, relearning in the featured condition should be lower than in the baseline one.

Second, we manipulated the types of training in which a word participated. Not all words participated in all three phases of the experiment. For this purpose, words were separated into three *clusters*: trained, untrained and control. Trained words were included in all phases of the experiment. Our primary test of relearning occurred by examining the effect of *target identity* within these words.

The other two conditions offer additional insight. Untrained words were included in all phases but training. Control words were only included during training but no other phases. These three clusters allow two additional comparisons of interest. First, we can assess the role of the training phase (versus decay) in unlearning the association with the featured object by comparing incorrect associations of trained and untrained words. Comparing the learning rate of trained and control words during training allowed for a test of pre-training's impact on subsequent learning.

In addition, inclusion of a word during training (trained/untrained words)[6] was crossed with target identity during relearning (featured/baseline foil). This allowed us to compare differences in relearning for words that had undergone unlearning and should only be strongly associated with their target objects (trained words) and ones that did not undergo unlearning and should be equally associated with their target objects and featured competitors (untrained words).

The strength of correct and incorrect associations between words and objects was assessed using a yes/no test, asking participants if a given word "went with" a given object. This test was administered at four points in the experiment: after pre-training, in the middle of training, after training and after relearning. The training and relearning phases used a supervised learning paradigm to guarantee high learning rates during training and relearning, thus facilitating the strengthening of the correct associations between words and objects and the unlearning of incorrect ones.

[6] This was not true for control words because they were not included in the relearning phase.

## 5.3 Method

### 5.3.1 *Participants*

Fifty-five native speakers of English with normal or corrected-to-normal vision participated in this experiment. All were students at the University of Iowa and received course credit for participation. Participants underwent informed consent, and the study was approved by the University of Iowa's IRB.

### 5.3.2 *Stimuli*

Participants learned 24 word-object-mappings. Words were two-syllable, phonologically legal CVCV pseudo words. Phonological overlap between words was minimized to facilitate participants' learning (see Table 7 for a list of words). All words were recorded by a male native speaker of English, and went through standard minor auditory editing (e.g., removal of clicks, addition of 50 msec silence in the beginning and end of each stimulus). For each word, five exemplars were selected (Roembke & McMurray, 2016). Images were photographs of highly infrequent objects that were presented on a white background.

Table 7: List of novel words used in Experiment 6.

| Written form | IPA |
|---|---|
| Bosha | /boʊʃɑ/ |
| Bure | /buɹeɪ/ |
| Chodu | /tʃoʊdu/ |
| Dimu | /dɪmuː/ |
| Fatei | /fæteɪ/ |
| Gichi | /giːʃi/ |
| Goba | /goʊbɑ/ |
| Haito | /heɪto/ |
| Jifei | /dʒifeɪ/ |
| Kepoi | /kepɔɪ/ |
| Lubo | /lubo/ |
| Mefa | /meɪfɑ/ |
| Naida | /nɑɪdɑ/ |
| Pacho | /pɑtʃo/ |
| Qufe | /kufeɪ/ |
| Razi | /ræzi/ |
| Sheku | /ʃeku/ |
| Sipa | /siːpɑ/ |
| Tamu | /tæmuː/ |
| Thidai | /θiːdeɪ/ |
| Viku | /viːkuː/ |
| Wouvo | /waʊvo/ |
| Yami | /ˈjɑmi/ |
| Zati | /zæti/ |

### 5.3.3 *Design*

Word-object-mappings were created by randomly assigning words to objects for each participant. For each participant, words were randomly separated into three clusters of eight words: trained, untrained and control. Trained words were included in all phases of the experiment. Untrained words were included in all phases of the experiment but training. By not including untrained words in training, we could contrast strength of incorrect associations between words and featured competitors that did (trained words) or did not (untrained words) undergo unlearning. Control words were only included during training (see Table 8). During training, learning rate of control words provided a baseline against which we could compare the learning rate of trained words. Thus, training cluster is a within-subject, between-word manipulation.

Table 8: Overview of Experiment 6's design.

| | | Word cluster included | | | Num. trials / | |
|---|---|---|---|---|---|---|
| Order | Phase | Trained | Untrained | Control | Reps/word | Design notes |
| 1 | Pre-training | ✓ | ✓ | | 256 / 16 | Word appears equally with target +featured competitor |
| 2 | Test (Prelim) | ✓ | ✓ | | 80 / 5 | 2AFC; yes/no task |
| 3 | Training | ✓ | | ✓ | 224 / 14 | 4AFCm supervised |
| 4 | Test (Mid Trn) | ✓ | ✓ | | 80 / 5 | 2AFC; yes/no task |
| 5 | Training | ✓ | | ✓ | 224 / 14 | 4AFC; supervised |
| 6 | Test (Post trn) | ✓ | ✓ | | 80 / 5 | 2AFC; yes/no task |
| 7 | Relearning | ✓ | ✓ | | 224 / 14 | 4AFC; supervised |
| 8 | Test (Post relrn) | ✓ | ✓ | | 80 / 5 | 2AFC; yes/no task |

For trained and untrained words, featured competitors were created by pairing words: The target object of one object served as the featured object of the other word, and vice versa: For instance, object O1, the target of word W1, was the featured competitor of word W2,

whereas W2's target, object O2, was W1's featured competitor. Word pairs always belonged to the same cluster (i.e., trained or undertrained).

*Pre-training.* During pre-training, words in the trained and untrained clusters were presented with a single object on each trial. Words were presented 16 times, and equally often presented with their target object and their featured competitor (eight trials with the target object, eight trials with the featured competitor; for a total of 256 trials). To avoid direct repetitions of the same item, words were randomized in sets of 16, thus creating 16 blocks. Blocks were re-randomized if two subsequent blocks ended and started with the same word. Direct repetitions were prevented to not further underline that each word was paired with two objects.

*Tests.* During all tests (test after pre-training, mid-training, post-training test, and post-relearning), participants were presented with one word and object at the same time. Words belonged either to the trained or untrained cluster. Each word was repeated five times (one trial with the target object, one with the featured competitor, two with baseline foils, one random), resulting in 80 trials overall. On each trial, participants had to indicate if the word and object belonged together. This yes/no task design was selected to not bias participants to reject one type of referent more than another and to allow for the acceptance of more than one referent per word across trials. Foils were selected to be either from within the same experimental cluster (trained/untrained; within-cluster foil) or not (out-cluster foil). For the random trial, one of the three foil categories was randomly selected (target, featured competitor, or baseline foil). Random trials were included, so that participants were not able to pick up on the foil structure during testing. As for pre-training, words were pseudo-randomized within blocks of 16 trials (one repetition of each word) to avoid direct repetitions of a word.

*Training.* Participants completed 448 trials of supervised training (separated into four blocks of 112 trials) on trained and control words. Participants completed a test of their knowledge of word object mappings after completing two blocks of training and at the end of training.

On each training trial, participants were presented with four objects (one of which was the target) and heard one of the words. Foil objects were always from the same cluster (trained/control) as the target word, and were selected without replacement to prevent high spurious associations between words and objects. As a result, each foil from the same cluster (including the featured object) had an equal probability of occurring with a word. Thus, for each trained word and training block, there were three trials that included the featured competitor and four trials that did not. For control words, no trials included a featured competitor because control words were not included in pre-training and thus did not possess one. Trials were randomized within block.

*Relearning.* The relearning phase of the experiment consisted of 224 trials, separated into two blocks of 112 trials. In this phase, we manipulated whether a word was retrained on a featured competitor or a random one; this was a within-subject, between-word manipulation.

For this purpose, half of the untrained and half of the trained words (the featured condition) were assigned their featured competitor as the new target. To do this, pairs of word-objects-mappings were respected: For example, if object O2 was the featured competitor of word W2, it now was its target. At the same time, object O1 became the target of word W2. For the other half of trained and untrained words (the baseline condition), the new targets were the within-cluster foils that were also used during tests (also see Table 8 for an overview of the design). These were not featured, and so were expected to have low baseline associations that did

not need to be unlearned during training. As in the featured condition, these were symmetrical. For example, assume that during training word W5 mapped onto object O5 and word W6 mapped onto object O6. In addition, during training, word W7 mapped onto object O7 and word W8 mapped onto object O8. During relearning, object O7 became word W5's new target and object O8 was word W6's new target. At the same time, object O5 became word W7's new target and object O6 was word W8's new target (see Table 9 for a visual representation). As a result, in the baseline condition, new targets were never the previous featured competitor of its assigned word. Thus, the only difference between featured and baseline conditions was that by the time of relearning, participants had received training on the mapping with the featured competitor during pre-training but not the baseline foils. This reconfiguration resulted in 16 novel word-object-mappings.

During relearning, four objects were presented on each trial, one of which was the novel target. Foils were pseudo-randomly selected from all possible objects within the same cluster (without replacement to rule out unintended increased co-occurrence of a word and an object). Thus, the old target acted now as one of the foil objects. In addition, for the baseline condition,

Table 9: Overview of relearning design. Please note that this overview only includes eight word-object-mappings of experimental interest (instead of 16) to facilitate presentation. "T" indicates the target object established during training (the old target); "F" indicates the featured competitor established during pre-training; and "R" indicates the target established during relearning (the new target). Light gray shading signifies words in the featured condition; dark gray shading signifies words in the baseline condition.

| | | Word (W) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Object (O) | 1 | T | F, R | | | | | | |
| | 2 | F, R | T | | | | | | |
| | 3 | | | T | F, R | | | | |
| | 4 | | | F, R | T | | | | |
| | 5 | | | | | T | F | R | |
| | 6 | | | | | F | T | | R |
| | 7 | | | | | R | | T | F |
| | 8 | | | | | | R | F | T |

the old featured competitor could be a foil object (this did not apply to the featured condition because the old featured competitor was the new target). Moreover, the old target and the old featured competitor were never present simultaneously on the same trial in the baseline condition. This was done to allow us to select trials that were equated for difficulty between the baseline and the featured condition: we could exclude baseline trials with the old featured competitor during analysis (since the featured condition would not have something analogous).

For each word, eight trials did not include the old target, whereas six did. For a word in the baseline condition, six out of those eight trials without the old target included the old featured competitor as a foil. Trial order within block was random.

### 5.3.4 *Procedure*

Participants received instructions that they would learn a set of novel word-object-mappings. They completed the experiment phases in the following order: pre-training, preliminary test, first half of training, mid-training test, second half of training, post-training test, relearning, and the post-relearning test (see Table 8).

On each pre-training trial, participants first viewed the object and blue center dot for 1050 msec on a 19" monitor operating at 1280 × 1024 resolution. The object was sized at 300 x 300 pixels, and was presented in one of the four corners of the screen (randomized across trials) to increase participants' probability of encoding the item a word co-occurred with. Subsequently, the blue center dot turned red. This signaled participants that they could play the word by clicking on the dot. When clicking the dot, the auditory stimulus was played via headphones. A trial ended when participants selected the object.

During testing trials, participants again saw one object and a blue center dot for 1050 msec. The object (300 x 300 pixels) was always presented in the top middle of the screen. In addition, two buttons were present to indicate a match (a green check mark; yes) or mismatch (a red cross; no). These were located in the left (match) and right (mismatch) bottom corners of the screen. After 1050 msec, the dot turned red, cueing participants to press on it to play the word. Participants had to select one of the buttons before advancing to the next trial. Participants never received feedback during testing.

Training and relearning trials followed an identical procedure. On each trial, participants saw four objects in the corners of the screen and the blue center dot for 1050 msec. Subsequently, the blue dot turned red, and participants clicked on it to hear the word. To advance to the next trial, participants had to select the referent of the word they just head. Subsequently, they received feedback on whether their response was correct (high tone) or not (low tone). Feedback was given immediately after participants' responses, and object pictures remained present during feedback.

For relearning, participants were not told that they had to learn novel word-object-mappings; instead, instructions were identical to what participants were given in the beginning of training. Overall, the experiment lasted approximately 1.5 hours and included 1248 trials.

## 5.4 Results

### 5.4.1 *Inclusion criteria/test after pre-training*

Seven participants did not finish the experiment due to time limitations; their data was included in all analyses when possible.

The experiment's design allowed me to test what participants learned during pre-training. This could be used to determine if pre-training statistics had been acquired (if they had not been learned, there would be nothing to unlearn [or relearn]). For this purpose, participants' sensitivity (d'), was calculated separately for target and featured objects. To do so, participants' hit and false alarm rates were calculated. A hit was defined as a trial, in which participants indicated a match between the word and the object when the target/featured object was present ('yes' response). A false alarm was defined as a trial in which participants indicated that the word and object matched, when the object was a foil. For this purpose, d' values at floor or ceiling were adjusted with a standard correction: If participants' hit rate/false alarm rate was at 1 or 0, half a hit/false alarm was subtracted/added, taking the number of relevant trials into account. For example, if there were 20 trials and a participant's false alarm rate was 0, this was adjusted to be 1/40 (0.025) instead. Similarly, if the same participant's hit rate was 1 and there were again 20 trials, 0.025 would be subtracted from 1 (adjusted hit rate = 0.975).

Participants were excluded if their d' for the target or the featured object was not significantly above 0 (null sensitivity). To do so, the one-signal significance test developed by Marascuilo (1970) was used, resulting in an approximate d' cut-off point of 0.23. Null sensitivity was true for eight participants,

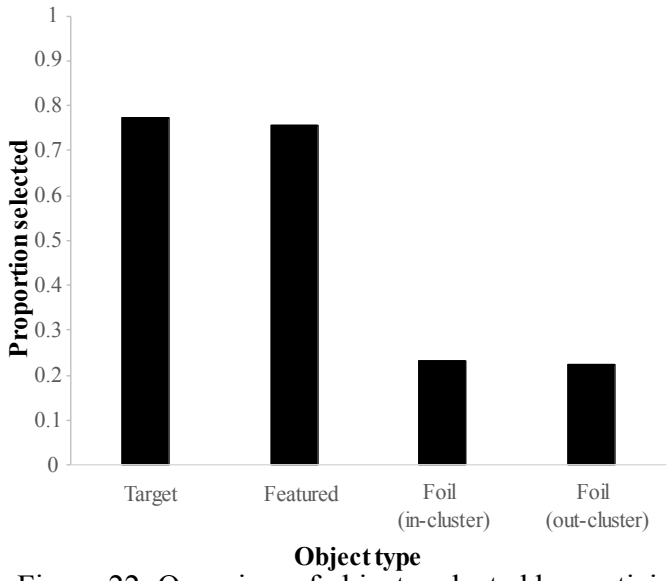


Figure 22: Overview of objects selected by participants at the test after pre-training and before converting values to d' (sensitivity) in Experiment 6.

leaving 47 participants for analysis. Seven out of the eight excluded participants failed because one of their d's was not significantly above 0; for one participant, this was true for both target and featured objects. For the remaining participants, average d' was 1.79 for targets (SEM = 0.17) and 1.77 for featured objects (SEM = 0.17).

For the remaining subjects, performance on the preliminary test reflected the statistics of pre-training (see Figure 22): Participants accepted the target or featured object at above 70% of trials; at the same time, foil objects were rejected as referents on the majority of trials (> 70% trials). This shows that participants were able to acquire two associations for each word.

### 5.4.2 *Question 1: Unlearning of incorrect associations*

We first analyzed participants' training data by comparing performance on trained and control words. Subsequently, associative strength of correct and incorrect associations was analyzed from the tests administered in the middle and at the end of training.

*Training*

For both trained und control words, participants' average accuracy reached ceiling (above 90% correct) by the end of training (see Figure 23). Training data were analyzed to address

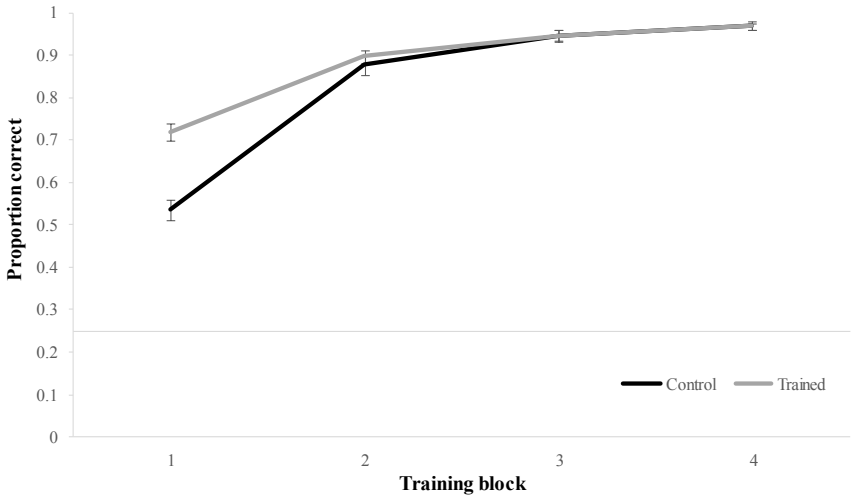


Figure 23: Overview training performance for trained and control words. Error bars indicate standard error of the mean.

two questions i: (1) did participants learn the word-object-mappings they were trained on? and (2) was this learning different for trained and control words?

Training was divided into four blocks of 112 trials. To investigate differences across the four training blocks, data were entered into a binomial mixed model implemented in R (version i386 3.4.3). Cluster (trained/control) was contrast coded (+/- 0.5), and block (1-4) was centered. Nested models were compared using chi-square tests of model comparison (Matuschek et al., 2017). For this, we report the chi-square statistics comparing the most complex model that reached significance and the next simpler model. The model that best fit the data and still converged included random intercepts of subject, stimulus and target object but no random slopes ($\chi^2(1) = 38.96$, $p < 0.001$).

This model found a significant effect of block (B = 1.20, SE = 0.03, z = 44.36, $p <$ 0.001), indicating participants' increase in accuracy as the experiment advanced. In addition, cluster reached significance (B = 0.24, SE = 0.06, z = 3.81, $p < 0.001$): Performance was higher for trained than for control words—but only in the beginning of training (interaction of block and cluster: B = -0.41, SE = 0.05, z = -7.79, $p < 0.001$); this can also be seen in Figure 23. The latter was confirmed when splitting the data in half by block, where cluster was significant in the first (B = 0.61, SE = 0.05, z = 12.24, $p < 0.001$), but not in the second half of training (B = 0.15, SE = 0.11, z = 1.42, $p = 0.155$). These results indicate that pre-training on both the correct and incorrect association facilitated subsequent acquisition of the correct mappings.

*Mid-training and post-training tests results*

To analyze performance on the mid-training test and the post-training test, d' was again calculated for target and featured competitors separately. For this purpose, "yes" responses on

target and featured trials were counted as hits; whereas, “yes” responses on both types of foil trials were counted as false alarms. The same correction (for floor and ceiling) was applied as for preliminary test analyses. As can be seen in Figure 24, sensitivity to the target and featured object remained largely unchanged for untrained words, but changed for trained words, with sensitivity for target words increasing and featured words decreasing.

Subsequently, we examined d’ values for each participant from test in the middle of training and the one right after training in a linear mixed effects model. We also used d’ values for target and featured objects on the preliminary test as a covariate. In addition to the covariate, the following fixed effects were added: object type (featured/target), cluster (untrained/trained) and test phase (mid-training test, post-training test) as well as their interaction terms. Object type and cluster were contrast coded (+/- 0.5); test phase was converted into subsequent numbers (2, 3) and then centered. The model that best fit the data is specified in (1) in simplified R language ($\chi^2(9) = 152.15$, $p < 0.001$); more complicated models were oversaturated:

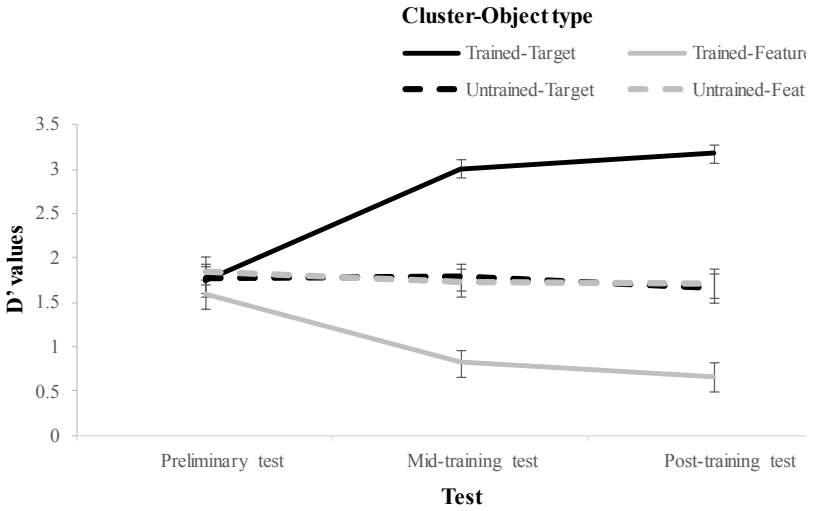


Figure 24: Overview of d’ for both experimental clusters (trained/untrained) in the preliminary test, mid-training test and post-training test. Error bars indicate standard error of the mean.

d’ ~ d’ preliminary test + object type * cluster * test phase +

(object type * cluster |subject) (1)

The covariate, d' on the preliminary test, accounted for significant variance (B = 0.54, SE = 0.05, $t(86.67) = 11.47$, $p < 0.001$). The main effect of object type was significant as well (B = 1.16, SE = 0.09, $t(46.02) = 13.26$, $p < 0.001$), indicating participants' d' was higher for target than featured objects. In addition, the main effect of cluster was significant (B = 0.27, SE = 0.12, $t(42.86) = 2.31$, $p = 0.026$) as well as the interaction between object type and cluster (B = 2.24, SE = 0.17, $t(46.31) = 13.50$, $p < 0.001$). Moreover, the three-way interaction of object type, cluster and test phase reached significance (B = 0.42, SE = 0.16, $t(184.00) = 2.63$, $p = 0.009$).

Unsurprisingly, when the data were split by cluster and the same model was ran as before (with cluster removed as a fixed factor and as a random slope), the main effect of object type was only significant for trained (B = 2.30, SE = 0.15, $t(46.28) = 15.14$, $p = 0.009$) but not for untrained ones (B = 0.06, SE = 0.08, $t(45.42) = 0.73$, $p = 0.469$). Similarly, the two-way interaction of objet type and test phase was significant for trained (B = 0.32, SE = 0.10, $t(92.00) = 3.19$, $p = 0.002$) but not for untrained words (B = -0.10, SE = 0.13, $t(92.00) = -0.81$, $p = 0.420$): Again, there was only differential change for target/featured objects when words were included in training and participants received feedback on the correct mappings. In contrast, change in sensitivity across test phase was non-existent for untrained words.

In addition, for the post-training test only, d' values for featured competitors of trained words were entered into a one-sample t-test, testing whether they were still above 0. Thus, this analysis asks whether training was successful in eliminating incorrect associations with the featured competitor. It was found that d' values were still significantly above 0 (mean = 0.65, SD = 1.20, $t(46) = 4.07$, $p < 0.001$, *Cohen's d* = 0.593). This suggests that even though the strength

of the incorrect association between the word and the featured competitor decreased over training, it was never completely unlearned. As a control, we also completed the same analysis for featured competitors of untrained words (i.e., words that were included in pre-training but had not been in training). Non-surprisingly, they were significantly above 0 as well (mean = 1.71, SD = 1.12, $t(46) = 10.41$, $p < 0.001$, *Cohen's d* = 1.519).

### 5.4.3 *Question 2: Functionality of sub-threshold associations*

To examine the functionality of sub-threshold associations, we next analyzed the relearning phase of the experiment. We started by analyzing the learning rate during the relearning trials for the two word clusters (trained, untrained) and relearning conditions (featured, baseline). Next we examined the strength of correct and incorrect associations at the test after relearning.

*Relearning*

For words in the featured condition, the featured object now constituted the target; however, for the words in the baseline relearning condition, this was not the case. Thus, these conditions were not equal in the fact that the baseline could contain a high-association featured competitor as a foil while the featured condition could not. Thus, to create equivalency across the conditions, trials that included the old featured object were excluded from analysis for words in the baseline relearning condition. Trials that included the old target object were included for both the featured and baseline relearning condition, as they existed at the same frequency for both. Participants' relearning performance is presented in Figure 25. It shows that learning rate

appeared to be better for untrained than for trained words, and for words in the featured than in the baseline relearning condition.

Relearning data were analyzed similarly to training data by entering them into a binomial mixed effects model. The fixed effects in this model were cluster (trained/untrained), relearning condition (featured/baseline) and relearning block (1, 2) as well as all of their interactions. Block was centered, and cluster and relearning condition were contrast coded (+/- 0.5) as before. The model that best fit the data and still converged included random effects of subject and stimulus, but no random slopes ($\chi^2(1) = 14.72$, $p < 0.001$).

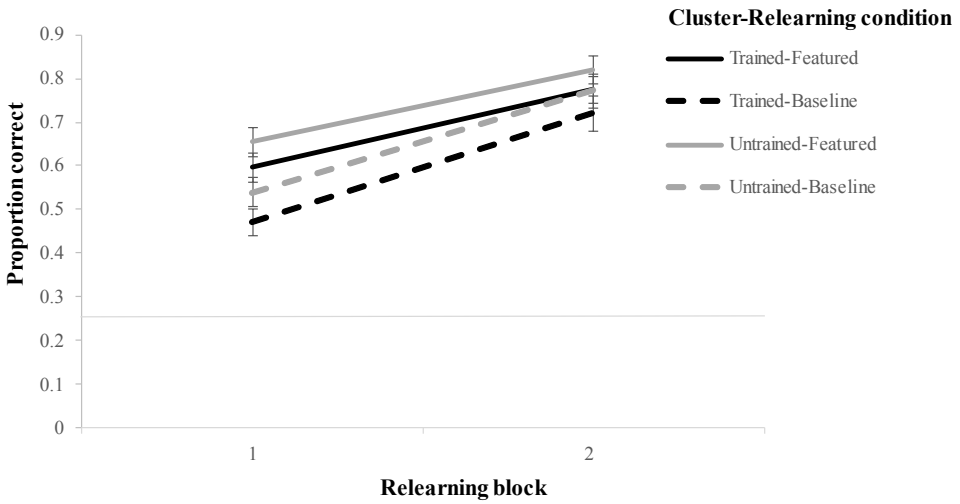


Figure 25: Participants' proportion correct during relearning in Experiment 6. Error bars indicate standard error of the mean.

Using this model, there was a significant effect of block (B = 1.16, SE = 0.06, z = 19.94, $p < 0.001$), as participants learned the new word-object-mappings and accuracy increased. In addition, both cluster and relearning condition reached significance (cluster: B = -0.33, SE = 0.06, z = -5.80, $p < 0.001$; relearning condition: B = 0.50, SE = 0.06, z = 8.62, $p < 0.001$): Participants' performance was better on untrained than trained words; in addition, their accuracy was higher for words in the featured than in the baseline relearning condition. The only interaction that reached significance was the one of relearning condition and block (B = -0.25, SE = 0.11, z = -2.20, $p = 0.028$; all others: $p > 0.8$).

To examine this interaction, we split the data by block, and ran the same model without block as a fixed factor. This revealed a significant effect of relearning condition in both blocks (relearning block 1: B = 0.57, SE = 0.07, z = 7.97, $p < 0.001$; relearning block 2: B = 0.40, SE = 0.09, z = 4.35, $p < 0.001$), suggesting that differences in relearning condition were too subtle to detect via such a coarse follow-up test. It should be noted, however, that the estimate of the effect was slightly higher in the beginning of the experiment than at the end.

As an exploratory follow-up analysis, we decided to replicate the previous analysis for trained words, but only including data from participants whose sensitivity to the featured competitor for trained words was below zero at the test administered after training. Thus, this analysis asked whether relearning in the featured condition was easier than the baseline one, even if incorrect associations to the featured competitor were completely unlearned.

There were 11 participants who met this criterion at the post-training test (M = -0.22, SD = 0.31). Their data was analyzed using a binomial mixed effects model with relearning condition (featured/baseline) and relearning block (1, 2) as the fixed effects. The model that best fit the data included random intercepts for subject, word and target object but no random slopes ($\chi^2(1) = 16.86$, $p < 0.001$). Using this model, there was unsurprisingly a significant effect of block (B = 1.07, SE = 0.16, z = 6.53, $p < 0.001$) as a result of an increase in performance over relearning. More interestingly, the main effect of condition reached significance as well (B = 0.39, SE = 0.19, z = 2.08, $p = 0.038$): Participants' accuracy for trained words was higher for words in the featured than for ones in the baseline condition. This suggests that better performance in the featured condition for trained words might not be just the result of remaining incorrect associations between the words and featured competitors at the end of training.

*Post-relearning test results*

To analyze data from the post-relearning test, d' values were calculated for the featured competitor and the foil from the same cluster for words that were in the featured as well as in the baseline relearning condition. Sensitivity was calculated for within-cluster foils, as they were the new targets in the baseline condition. Similarly, sensitivity was calculated for featured competitors, as they constituted the new target in the featured condition. To calculate d', trials were the old target was present were not included (i.e., they were neither counted as false alarms or hits). As before, hit/false alarm rates of 0 or 1 were adjusted with the previously introduced correction. This was particularly important, as some d' were expected to be very low, e.g. sensitivity for the within-cluster foil in the featured condition, where it did not carry any special value (i.e., it was always a foil).

As can be seen in Figure 26, sensitivity was higher for objects that were now the target in the relearning phase than for objects that were foils during relearning. Moreover, sensitivity was higher for objects that used to be the featured competitor; this was even true when the featured competitor was a foil during relearning.

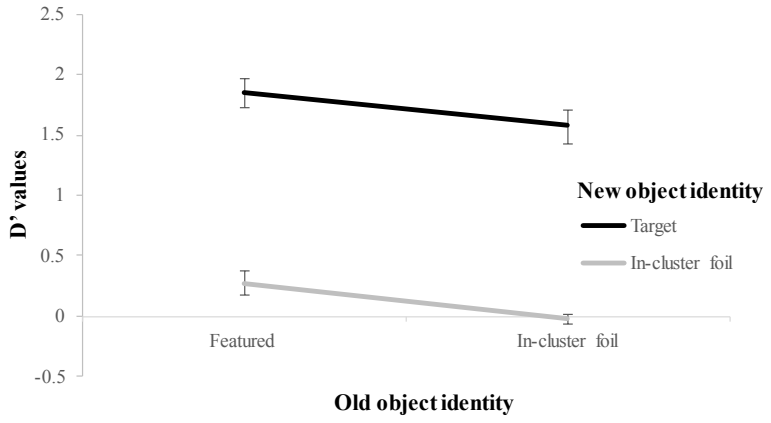


Figure 26: Participants' proportion correct during relearning. Error bars indicate standard error of the mean.

Subsequently, values were entered into a binominal mixed effects model. The following fixed effects were coded: the old identity of the object types of interest (called old object here; featured/in-cluster foil), their new identity (called new object; target/in-cluster foil) and cluster

(trained/untrained). The model that best fit the data is described in (2) ($\chi^2(2) = 88.06$, $p < 0.001$). Please note again that more complicated models were oversaturated.

d' ~ cluster * old object * new object + (new object| subject) (2)

Using this model, we found that there was a significant effect of old object (B = 0.29, SE = 0.06, $t(240) = 4.54$, $p < 0.001$) and new object (B = 1.59, SE = 0.14, $t(40) = 11.57$, $p < 0.001$; no other effects or interactions reached significance (all $p > 0.1$). D' values were higher for objects that used to be featured than ones that were in-cluster foils. In addition, participants were more likely to "accept" objects that had been reinforced during relearning (i.e., new targets) than objects than had not (foils; see Figure 26 for an overview). Surprisingly, there was no significant effect of cluster; that means d' values at the end of relearning were the same for referents of words that had been included in training and ones that participants had not been trained on.

## 5.5 Discussion

The goal of this study was to answer two questions: (1) How quickly are incorrect associations unlearned? (2) Do unlearned or suppressed associations have the potential to influence participants' behavior at a later time point? We will discuss these two questions in turn before considering limitations of this study and returning to a more general discussion.

### 5.5.1 *Question 1: Unlearning incorrect associations*

Participants were able to acquire the word-object-mappings, where each word was associated with two referents (the target and the featured object). In addition, participants'

acceptance of the featured object as a potential referent declined significantly, when feedback reinforced the target object only and punished the selection of the featured competitor. However, it never reached zero: Even as participants had learned the correct word-object-mappings, their sensitivity to the featured competitor remained significantly above zero. Importantly, this change was restricted to trained words, i.e. words that were included during training. In contrast, both target and featured objects remained equally acceptable referents for untrained words.

These findings are surprising, and suggest that incorrect associations are maintained even as they are actively suppressed by feedback. One of the reasons for this might be that unsupervised statistics still supported an association between the word and its featured competitor, as they repeatedly co-occurred during training. This is a form of cross-situational word learning (Yu & Smith, 2007), where the association between two words is strengthened as a result of their repeated co-occurrence.

In addition, it is possible that the pre-training where each word was paired with two meanings "loosened up" people's use of the mutual exclusivity assumption. Previous research has shown that the mutual exclusivity assumption and people's ability to fast-map novel words onto unknown objects is shaped by their linguistic experience (Halberda, 2003; McMurray et al., 2012): Bilingual children are less likely to use the mutual exclusivity bias, presumably because they more frequently encounter two terms for the same meaning (Houston-Price et al., 2010; Kalashnikova et al., 2015). Similarly, in this experiment, pre-training might have taught participants that one word can map onto more than one meaning, thus decreasing the need to unlearn the incorrect association—even as it received negative feedback during training. However, it is currently not known how quickly the extent of one's reliance on the mutual exclusivity bias can change: Bilingual people have extended experience with two languages and

thus more than one word for the same meaning. It is less clear whether participants' use of the mutual exclusivity bias could have been "relaxed" as the result of a pre-training that took approximately ten minutes.

To summarize, our data indicate that incorrect associations between words and objects are maintained at low levels, even if the correct referent for a word is selected repeatedly. Moreover, there appears little reduction in associative strength between the test conducted in the middle and at the end of training (see also Figure 24); this suggests that the majority of unlearning was completed in the first half of training. This was also true for the strengthening of the correct associations: Sensitivity to the target increased most between the preliminary test (the one conducted right after pre-training) and the mid-training test.

### 5.5.2 *Question 2: Functionality of sub-threshold associations*

During relearning, untrained words were more easily mapped onto a new referent. This makes sense when considering participants' sensitivity to different object types at the end of training (i.e., at the post-training test): Trained words were strongly associated with one object, the target, and less so the featured competitor. In contrast, untrained words were equally associated with the target and the featured competitor—but the *overall strength* of those connections were lower (as they had only been reinforced during pre-training).

However, new mappings were also easier to learn when they included the featured competitor as the new target than when the new target was a foil from the same cluster. This result is harder to explain in the context of differences in associative strength: Why should it be easier to associate a word with an object that was previously "unlearned" (or suppressed) than one that had never been "special" in respect to that particular word? One possibility is that this

was the case, as the connection between the word and the featured competitor was never fully suppressed or unlearned. Converging evidence for this explanation comes from the finding sensitivity to the featured competitor was significantly above zero even at the end of training.

However, our exploratory analysis with the subset of participants suggests that learning was easier in the featured condition for trained words than in the baseline one, even when only data were included from participants that no longer associated the words with their featured competitor. This provides first tentative evidence that sub-threshold associations can influence subsequent behavior even if they were completely unlearned beforehand.

Overall, these findings suggest that small associations are functional even after they undergo unlearning, and that they can be used to build positive associations. Thus, these results are similar to what was observed in Yurovsky et al. (2014), where sub-threshold associations existed as a result of some unsupervised exposure to word-object-mappings.

### 5.5.3 *Limitations*

One limitation of this design is the necessary inequality of the featured and baseline condition trials during relearning: For the featured condition, objects that were the old featured competitor were now the target. However, for the baseline condition, a previous foil became the target; the featured competitor was still used as a foil. To counteract this, we excluded trials of words in the baseline condition that used the old featured competitor as a foil from analyses. However, it is clear that learning is not isolated from those particular trials, and they might have negatively impacted participants' overall acquisition of the word-object-mappings in the baseline relearning condition.

Thus, as a follow-up experiment, we decided to replicate the previous design with a couple of important changes: During relearning, the old featured competitor was never a foil in the baseline condition. To mimic this for the featured condition, one baseline foil could no longer appear as a foil during relearning (thus keeping the overall number of potential foils the same across conditions). Moreover, the previous target was never included as a potential foil; this was true both for the featured and baseline condition. As a result, participants' essentially will acquire an additional referent for each word, instead of learning a new target object (as the old target will never be explicitly rejected). Thus, these design changes will address whether lower learning rates for the baseline relearning condition than then featured condition can be attributed to this difference across relearning conditions.

In addition, control words in the follow-up experiments will also be included during relearning, thus providing an additional baseline of how quick relearning proceeds for words without a featured competitor. We also decided to double the number of training blocks with the goal of maximizing unlearning of incorrect associations with the featured competitor. Data collection for the follow-up experiment is currently in progress.

Moreover, as previously discussed, the pre-training design might have biased participants to maintain more than one meaning for each word. This might have extended to other phases of the experiments. For example, it is possible an incorrect association would have been unlearned more quickly if participants had been only trained one meaning per word, which was subsequently unlearned. Thus, future research needs to clarify if secondary meanings or sub-threshold associations are also maintained under different circumstances.

### 5.5.4 *The importance of unlearning incorrect associations in vocabulary acquisition*

Words are often thought to be acquired either by fast-mapping or cross-situational word learning. In a typical fast-mapping experiment (Carey & Bartlett, 1978), young children are presented with two objects they know (e.g., BALL, SHOE) and one they do not (a novel object). They then are asked to pick out one of the items. When the experimenter asks the participants to pick out the TOMA, a novel, made-up word, children reliably select the novel object they do not know and that does not have a label attached to it. Thus, the novel word is fast-mapped onto the object, facilitated by the mutual exclusivity assumption that each object only has one label. Importantly, fast-mapping has often argued to be a form of one-shot, all-or-none learning. As a result, fast-mapping is highly susceptible to errors: If a child gets it wrong, an incorrect word-object-mapping has been encoded.

While cross-situational word learning is less likely to suffer from mistaking a word's meaning than fast-mapping, as it relies on the accumulation of statistics across time (not a single naming situation), it comes with its own potential need for unlearning. As discussed previously, there is evidence that spurious associations between a word and an object can be encoded, and that such associations can both slow down processing as well as the acquisition of the correct word-object-mappings (Roembke & McMurray, 2016). Thus, studying how incorrect associations are unlearned is an important aspect of research on vocabulary acquisition.

How to these sub-threshold associations relate to other language processes? One intriguing possibility is that incorrect associations remain part of the lexicon, the storage of a person's known words. Connections between lexical entries are supposed to reflect both semantic similarity between words as well as their co-occurrence (Elman, 2004; Hills, Maouene, Riordan, & Smith, 2010; Hills, Maouene, Maouene, Sheya, & Smith, 2009). This description

does not sound that different to how an incorrect association between a word and meaning may arise, suggesting a potential connection between lexical structure and the learning/unlearning of incorrect associations between words.

### 5.5.5 *Conclusions*

We found that incorrect associations between words and objects were not fully unlearned, even after participants underwent a training on the correct word-object-mappings and performed at or close to ceiling when selecting the target object. Moreover, these sub-threshold associations were found to be functional in influencing subsequent word learning. These results indicate that incorrect associations might be maintained for relatively long periods of times and that they can impact people's behavior.

Word learning does not only consist of building correct associations between a word and its meaning; one also needs to know what word and meaning do not map onto each other. Better understanding how incorrect associations are unlearned might provide important insights into how vocabulary development and in-the-moment language processing occurs.

# CHAPTER 6: ACTIVATION OF NEWLY ACQUIRED WORDS

## 6.1 Introduction

The process of learning of a novel word is not simply a matter of mapping a phonological word form onto a semantic representation (meaning). Instead, the ultimate goal of word learning is to build a functional lexicon of words. Thus, testing whether a word is learned or not should go beyond accuracy (can the participant pick out/recognize the correct referent?). Instead, the question becomes whether a word is integrated into a person's existing lexicon *in addition* to whether it is linked to its respective meaning. That is, we must ask whether novel word behave in similar ways as previously acquired ones.

This difference in how word learning could be conceptualized was captured by Leach and Samuel's (2007) distinction between lexical configuration (defined as the fast acquisition of a word via episodic memory) and lexical engagement (defined as the slower emergence of interactions between novel words and the rest of the lexicon). Lexical engagement is needed to allow for the swift activation of a word as part of real-time processing (e.g., recognizing a word or producing it; c.f., Elman, 2009; Gupta & Tisdale, 2009). In such a framework, word learning is more adequately described as a gradual process that is not all-or-nothing (see also McMurray, Horst, & Samuelson, 2012; Tamura, Castles, & Nation, 2017).

So what is the difference between merely knowing a word and embedding it in a network of words? A word can be activated by bottom-up phonological information that maps onto an abstract representation of the word form. This is what we think of as "knowing" the words' sound pattern and meaning. However, in addition, word form representations are also thought to inhibit each other (Dahan, Magnuson, Tanenhaus, et al., 2001; P A Luce & Pisoni, 1998). This process can facilitate and speed up word recognition, as a more active word (for example a word

with a better match with bottom-up information) can inhibit competitors. In addition, there may also be top-down connections that feed information back from the abstract word form representation to the phoneme level (Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McClelland, Mirman, & Holt, 2006; Norris, McQueen, & Cutler, 2003). For example, a word may be activated more quickly if it was presented twice in a row or if it was preceded by a semantically related word (Meyer & Schvaneveldt, 1971). Theoretically, either and/or all of these connections (bottom-up pathways, lateral connections between words, top-down connections) can change over learning (e.g., a word is activated more quickly as it is experienced more often) and development (e.g., bottom-up information may generally be accessed more quickly by adults than children). In general, inhibition between words and feedback/flow of top-down information can be seen as evidence for lexical engagement, whereas bottom-up pathways map onto lexical configuration of words.

The goal of Question 4 was to examine aspects of lexical integration in more detail, asking how quickly the word-object-mappings (that were previously acquired as part of the Experiments used to address Questions 1 and 2) are activated in comparison to known words. By doing so, we start to address whether differences in competition between novel and existing words as well as the rest of the lexicon are partly driven by differences in how rapidly newly acquired words are activated relative to familiar competitors. It should be highlighted that the experimental set-up was not optimal for this purpose, but may nevertheless provide interesting insights into this area of research. As a result, Question 4 should be treated as exploratory.

### 6.1.1 *Evidence that lexical integration requires sleep-based consolidation*

Gaskell and Dumay (2003) investigated lexical integration by teaching participants novel words that were highly similar to multisyllabic existing words (e.g. CATHEDRUKE, derived from CATHEDRAL) in a phoneme-monitoring task. Subsequently, participants' familiarity with the novel words was tested in a forced-choice recognition task. This tests something akin to the learning of the words' configuration (knowledge of the sound pattern). To assess whether the words were integrated into mental lexicon of existing words, Gaskell and Dumay's (2003) examined inhibition/competition. This was done by examining participants' reaction time in the pause detection paradigm (Mattys & Clark, 2002): Participants were asked to make speeded judgements on the presence of a 200 msec pause, which had been added near an existing word's point of disambiguation (e.g, CATHE_DRAL). Participants were tested on existing words that either had a newly acquired competitor (test words; e.g., CATHEDRAL) or not (matched control words; e.g., DOLPHIN). If a newly acquired competitor is integrated into the lexicon, participants should be slower at detecting the pause in test words than control words. However, if a novel word has not been integrated yet, there should be no difference in reaction time between test and control words.

Gaskell and Dumay (2003) found that participants' knowledge of the phonological patterns (as tested by a 2AFC recognition task) was always high. Despite this, lexical competition only emerged after one week (in the absence of additional exposure to the novel words). These results suggest that integrating a novel word into the existing mental lexicon is not immediate. In contrast, the process may take an extended period despite people's ability to acquire phonological information quickly (Gaskell & Dumay, 2003).

Similar findings indicating that lexical integration may rely on a prolonged process of sleep-based consolidation have been reported in subsequent studies. For instance, Dumay and

Gaskell (2007) controlled the amount of time between acquisition and test, while manipulating whether participants slept in the delay or not (using a classic AM/PM design). They found that the no-sleep group who had been trained in the morning and tested in the evening did not show interference effects, whereas the sleep group who had been trained in the evening and tested in the morning did. These findings suggest that a time delay by itself may not be enough for showing integration, but that sleep is necessary to allow for lexical engagement (as measured in the pause detection paradigm; Dumay & Gaskell, 2007). Similar results have been found when using a lexical decision task to measure lexical integration (Gaskell & Dumay, 2003; Lindsay & Gareth Gaskell, 2013).

Together, these data were seen as evidence for the complementary learning systems (CLS) account of word learning (Davis & Gaskell, 2009): In this account, novel words are initially acquired via hippocampal mediation between neocortical regions (similar to lexical configuration; Leach & Samuel, 2007; Tamura et al., 2017). Subsequently, sleep is required for hippocampal replay to support the integration of the new mappings with existing knowledge in neocortical long-term memory (comparable to lexical engagement; Davis & Gaskell, 2009; McClelland, McNaughton, & O'Reilly, 1995). A similar pattern (i.e., emergent lexical competition after sleep) has also been observed in children (Henderson, Weighall, Brown, & Gaskell, 2013), suggesting this model may be valid across development.

According to the CLS account, Dumay and Gaskell (2012) argue that learning new words may interfere with phonologically similar words already existing in the lexicon, a form of catastrophic interference. As a result, novel words may first be encoded using hippocampal systems that keep new representations separated from long-term lexical mappings. During this initial hippocampal encoding, representations are hypothesized to be sparser and thus slower to

activate. Through more gradual consolidation, direct cortical mappings are formed (Dumay & Gaskell, 2012); these allow for more rapid activation of a word. In the original account of the CLS, words encoded through hippocampal connections were completely separate from neocortical long-term knowledge. As a result, it was argued that no inhibitory connections could exist between newly acquired and well-known words (Davis & Gaskell, 2009). In more recent versions of the CLS account, it is proposed that weak competition between the two can arise: However, as hippocampal connections are slower, a non-consolidated word cannot be activated quickly enough to compete with existing words, which are activated via more direct, cortical connections. After consolidation, however, cortical mappings for newly acquired words are strengthened, thus allowing for lexical competition with the existing lexicon (see Davis, Di Betta, Macdonald, & Gaskell (2009) for evidence; McMurray, Kapnoula, & Gaskell, 2016).

### 6.1.2 *Immediate lexical integration in the absence of sleep*

More recently, an emerging body of evidence supports immediate lexical integration of novel words in the absence of sleep: For instance, Kapnoula, Packard, Gupta, and McMurray (2015) investigated novel word integration in an eye-tracking (visual world) paradigm (but see also Fernandes, Kolinsky, & Ventura, 2009; Lindsay & Gareth Gaskell, 2013 for further evidence of lexical integration without sleep-based consolidation). Participants first completed an exposure phase during which they heard novel words (modeled after Gaskell & Dumay, 2003). They were tested in a visual world paradigm using the sub-phonemic mismatch paradigm (developed by Dahan et al., 2001; Marslen-Wilson & Warren, 1994). To elicit lexical competition between words, auditory stimuli were constructed by splicing the end of one word (e.g., -CK from NECK) onto the initial portion of another word. The initial portion could be

taken from three possible word types: First, it could be from a different token of the same word (e.g., $NE_{neck}$; the matching splice condition). Second, the initial portion could come from a different word (e.g., $NE_{net}$; taken from NET; the other word splice condition). Third, it could be from a novel, made-up word (e.g., $NE_{nep}$; the nonword splice condition); we will refer to the novel word as a nonword to avoid confusion between item types in this study. If the beginning splice and end splice of a word did not match (e.g., $NE_{net}$CK as in the other word splice condition), co-articulatory information included in the initial splice would be misleading about the upcoming consonant. If this co-articulatory information was consistent with another known word ($NE_{net}$CK is briefly consistent with NET), this should engage inhibition. Here, one would predict that participants' ability to activate the correct word is slowed down if the initial splice comes from a real word, but not from an untrained nonword.

Kapnoula et al. (2015) leveraged this to examine lexical inhibition from newly learned words. Participants were either exposed to the nonword during training or not, and then immediately tested in this paradigm. If consolidation is not necessary for lexical engagement, this slowing down may even be observed for nonwords participants were trained on right before testing. It was found that participants were quickest at activating words from the matching splice condition and slowest at activating words from the other word splice condition (as predicted). Importantly, for trained nonwords, the nonword splice condition showed more of a delay than untrained. That is, these nonwords acted more like splices from existing words in the lexicon, inhibiting the activation of the target word (due to the misleading co-articulatory information). Thus, newly acquired words can be integrated immediately into the lexicon. These findings were replicated in the same paradigm with the exception that speaker voice was changed between training (initial exposure) and test (Kapnoula & McMurray, 2016). These results suggest that the

observed competition cannot be explained by exemplar-based episodic memory, but rather must be due to more abstract representations of the newly acquired words.

In summary, while sleep-based consolidation likely benefits lexical integration (as measured by interference/inhibition), it may not be necessary for integration. How lexical integration is measured may influence significantly whether a novel word is found to engage with the existing lexicon. Eye-tracking in the visual world paradigm may allow for a more sensitive and direct measure of inhibition between words. In contrast, performance in the pause detection paradigm may be more reliant on representations relevant to "offline" or more reflective processing, which in turn could depend more strongly on sleep-based consolidation (Kapnoula et al., 2015; McMurray et al., 2016; Weighall, Henderson, Barr, Cairney, & Gaskell, 2017).

### 6.1.3 *Activating novel words*

One major source of uncertainty in the studies discussed above is that it was not possible to compare how quickly newly acquired words are activated in comparison to known words. Instead, the indirect impact of newly acquired words on a known word's activation was studied as evidence of lexical engagement of the new word. This is because participants were only taught novel word forms without any meanings (Davis et al., 2009; Gaskell & Dumay, 2003; Kapnoula & McMurray, 2016a; Kapnoula et al., 2015). Word meanings are needed to use techniques that can test novel word activation (e.g., cross modal priming, the visual world paradigm).

How quickly a novel word can be activated might have significant implications for our understanding of lexical integration: Right now, we do not know whether differences in how a novel word competes with an existing ones is due to changes in bottom-up strengths (i.e., its

configuration) or its integration (i.e., lexical engagement). During word learning, the connections between a word form and its meaning are strengthened; it is possible that these changes in configuration allow for inhibition with existing words. In this account, no top-down connections or lateral inhibition between words are needed to account for changes in how a novel word competes with existing ones, all of the changes derive from differences in the degree to which the competitor is activated at all. The degree to which word A inhibits word B is directly related to how active it is. As a result, if word B is not very active, it cannot inhibit word A. Thus, changes in configuration can lead to apparent changes in inhibition even if the inhibitory connections are present the whole time. Or, there may not be any behavioral evidence for inhibition even if the inhibitory connections are present, because the competitor is not active enough.

To examine this possibility, McMurray et al. (2016) conducted a number of computational simulations using TRACE, an interactive model of speech perception (McClelland & Elman, 1986). They added a novel word (e.g., SUITABIT) to the standard lexicon that closely matched an existing word (e.g., SUITABLE) to test how changes in how well the novel word was known affected activation of the existing word. McMurray et al. (2016) found that the more frequent the novel competitor was, the slower was the activation of the familiar word. In addition, the final level of activation was lowered at the highest frequencies that were tested. However, both of these differences in interference were observed without changing the inhibitory strength of the newly learned word (SUITABIT). These simulations indicate that how much a well-known and a novel word compete with each other might depend on how "strong" the novel word is. That is, its lexical configuration might impact whether lexical engagement with other words can be observed. Thus, we might miss a crucial piece of the puzzle

when we try to investigate lexical engagement without clearly understanding how quickly a novel word is activated—relative to existing words—after acquisition.

Variants of the visual world paradigm can be used to study the activation of a word more directly, where participants' eye movements to objects are measured while hearing a word or sentence (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and the relevant newly learned words are given a referent to estimate their activation dynamics more directly.

Allopenna, Magnuson, and Tanenhaus (1998) showed how this works in the context of familiar words. They presented participants with four pictures and monitored their eye movements in the visual world paradigm. Participants were asked to move one of the present items (e.g., "Pick up the beaker; ..."); trials always included a picture of the target object (e.g., BEAKER), but could also include a cohort competitor (e.g., BEETLE) or a rhyme competitor (e.g., SPEAKER). An unrelated item (e.g., CARRIAGE) was always included as a measure of baseline looks. Allopenna et al. (1998) found clear cohort and rhyme activation (i.e., increased eye movements to phonologically related competitors). This incremental processing and in particular the activation of cohort competitors is very robust when investigating speech perception (Allopenna et al., 1998; Dahan, Magnuson, Tanenhaus, et al., 2001; James S Magnuson, Dixon, Tanenhaus, & Aslin, 2007; McMurray et al., 2010) and is reliable across days (Farris-Trimble & McMurray, 2013).

Magnuson, Tanenhaus, et al. (2003) extended this approach to newly learned words. They trained adult participants on novel word-object-mappings, which consisted of cohort pairs (e.g., PIBO and PIBU overlap at onset, and are therefore cohorts). Subsequently, they assessed if participants were equally likely to look at the target picture and image of the cohort competitor (if it was included) during the duration of the phonological overlap of the two stimuli. After

training, participants were eye tracked in a visual world paradigm task similar to the one used by Allopenna et al. (1998). There was evidence for competition between cohort competitors and thus incremental processing, just as has been observed with real words (e.g., Allopenna et al., 1998). However, importantly, Magnuson, Tanenhaus, et al. (2003) also tested whether real-word lexical neighbors of novel words impacted participants' fixations. They failed to observe such competition between novel words and real words, suggesting that even though novel words compete with each other, they are not integrated into the larger lexicon.

In a recent study by Weighall et al. (2017), both children and adults learned to map novel words (e.g., BISCAL) onto novel objects. Participants were taught one set of word-object-mappings on day 1 and one on day 2. On day 2, Weighall et al. (2017) asked to what extent the trained novel words would behave as cohort competitors to real words (e.g., BISCUIT) over and above of an unnamed object.[7] This pattern of activation then was compared to trials in which both target, cohort and unrelated competitors were real words.

Weighall et al. (2017) found that participants looked more to the cohort than the unrelated foil image of a novel object, even if the cohort was a newly acquired (trained) word and not a real word. This was true for both children and adults. For adults, there was not much difference between how strongly words were activated that were learned the day before and ones learned on the same day. Whereas, for children, words learned the day before were activated more than ones learned on the same day as testing.

Importantly, familiar word competitors were activated more strongly as well as suppressed more efficiently than novel ones, suggesting quantitative differences in lexical competition. As a result, Weighall et al. (2017) concluded that even though this is evidence for

[7] Unnamed objects had not been seen before, but were introduced in filler trials during testing, where they appeared with three real words, thus allowing for participants to fast-map the novel word to the referent.

immediate competition, these data may indicate that there are still important differences in how newly acquired words are processed *before* consolidation (that arguably takes longer than one night of sleep) and after. Novel words may be able to immediately inhibit/compete with existing words, but they still will not act like fully-fledged lexical representations until after a more prolonged, sleep-based consolidation has taken place.

There are a couple of limitations to this study. First, none of the trials that included a cohort competitor on the screen required participants to select the novel word. This did not only potentially bias participants against its selection, but also meant that Weighall et al. (2017) never tested how much activation of newly acquired words is influenced by existing words (only the other way around). In addition, exposure to novel words was relatively limited (less than 10 repetitions/word), leaving open the possibility that activation of novel words was less robust because they acted like existing low-frequency words, not because they could not compete with known words.

### 6.1.4 *Integrating written words*

Of course, lexical integration can not only be studied in auditory but also in written words. Indeed, learning of written words was subject of Experiment 4 of this dissertation. These issues of competition and integration have also been studied in this context.

Bowers, Davis, and Hanley (2005) first trained participants on novel written words, and subsequently measured response times to known words in a semantic decision task. Known words were selected to be orthographic neighbors of newly acquired words. Thus, Bowers et al. (2005) tested whether newly acquired words impacted and thus competed with known written

words. They found that a delay in response times only emerged after a 24 hours delay, similar to findings of Gaskell and colleagues (Davis & Gaskell, 2009; Gaskell & Dumay, 2003).

Moreover, data from a study by Bakker, Takashima, van Hell, Janzen, and McQueen (2014) indicated that the timecourse of lexical integration may vary for different types of mappings. In their study, participants were trained on either novel spoken or written words. Subsequently, they tested whether novel words competed with known words in either modality. They found that for novel words that were acquired auditorily, interference within both the spoken and written modality was not observable immediately after training, but after one night of sleep. Moreover, interference effects grew with increased opportunities for consolidation, indicating words' gradual, multi-day integration into the lexicon. Importantly, Bakker et al. (2014) found that competition between words learned via written presentation and real spoken words was not present until after a week (though there was some interference with written words after 24 hours). This result opens up the possibility that the lexical integration of written words may be accomplished with a different timecourse.

This is an important finding for at least two reasons: First, the full integration of a word form does not just require the mapping between a word form and a meaning, but the creation of multiple links (including orthography to phonology and orthography to semantics for reading; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Seidenberg, 2005). Thus, the observation that the timecourse of lexical integration may differ for different mappings is significant, and underscores that full integration of all possible linkages may be a prolonged process. Second, many words children acquire after they enter the school system are from reading (not from spoken language; Nagy, Herman, & Anderson, 1985). Thus, if integration is slower for written words, this may affect students' ability to recognize words automatically during reading:

Automaticity in word recognition has been hypothesized to be crucial for fluent reading (LaBerge & Samuels, 1974; Perfetti, 1985; Roembke, Hazeltine, et al., 2018). Reading fluency, in turn, may be required for reading comprehension: Through more automatic and thus more fluent reading, people are able to use cognitive resources, which would have otherwise been used for lower-level processes, for reading comprehension (Walczyk, 2000).

The lexical integration of novel written words acquired via reading was investigated in 9-to-11-year-old children in a recent study by Tamura et al. (2017). Children were exposed to novel written words that were embedded in stories. Lexical competition (engagement) was assessed testing the emergence of a prime-lexicality effect: In a masked priming paradigm, participants first briefly saw a prime and then a target word; their job was to indicate whether words were real or made-up. A nonword (e.g., ANPLE) that shares characteristics with a real word (APPLE) facilitates the latter's activation, resulting in shorter response times. A real word (e.g., AMPLE) as a prime, in contrast, does not facilitate another related word's activation (e.g., APPLE) and may even slow it down (Davis & Lupker, 2006; Forster & Veres, 1998). As a new word becomes more and more familiar, it is predicted to lose its facilitatory prime role, resulting in slower response times because it starts to inhibit the target word.

Tamura et al. (2017) found that novel words that children had been exposed to four times did not act like lexically engaged words, but novel words that participants had been exposed to twelve times did. Importantly, lexical engagement was always tested at least one day after exposure; this means that children always had access to sleep-based consolidation. Nevertheless, only children in the higher-exposure group showed lexical integration of newly acquired words. Currently, it is unclear whether this was due to the increased number of exposures to a word or

the longer time period for consolidation. Either way, these data suggest that the integration of novel written words may be a relatively slow process in children.

To summarize, there currently appears to be a lack of studies examining the integration of words that are learned via the written modality; this is important given how many words are learned this way (e.g., Nagy et al., 1985). Given Bakker et al.'s (2014) and Tamura et al.'s (2017) results, it is likely that words learned via print may be integrated at a different timecourse than words that are learned via the spoken modality, though additional evidence is needed.

### 6.1.5 *Question 4: Activating newly acquired words*

As described in the previous sections, one of the most critical questions in the lexical integration of novel words is whether sleep is required to allow for consolidation. At this point, research indicates that some interference is possible even right after acquisition, though observing integration of novel words may differ depending on the task paradigm (e.g., pause detection task vs. visual world paradigm; see McMurray, Kapnoula & Gaskell, 2016 for a review). This suggests that sleep might not be needed for lexical integration.

We do not currently have a good understanding of how quickly newly acquired words are activated relative to known words. As a result, we do not know if newly acquired words are less activated as cohorts of known words than existing word cohorts because they inhibit less or whether they are not active enough. To examine this, the visual world paradigm was used as a sensitive measure of participants' consideration of different words. In Experiments 2-4, participants completed a word recognition section after the no-correct testing phase.

For the real time word recognition phase, each novel word was selected to have two real word competitors (e.g., novel word: BEAMLER; cohort competitors: BEAKER, BEAVER). As

in Magnuson et al. (2003), participants completed a typical visual world paradigm task: Participants were presented with a novel or familiar word and were asked to click on the visual image that matched it; this was consistent with instructions in the rest of the experiment (no feedback was provided). Trials mixed real words and novel words, and trial configurations carefully counterbalanced which foils were present on each trial to avoid confounds.

Eye-tracking in the visual world paradigm allows us to measure activation of different words at the same time. Thus, this set-up was well-suited to investigate how quickly newly acquired words were activated relative to known words. More specifically, the following predictions were made: Novel words should be activated more slowly than known words. Moreover, if the target and cohort stimulus are both real words, cohorts should be activated more strongly than when the cohort is a novel word. If the target is a newly acquired word and the cohort a real word, participants may show increased cohort competition than when both target word and cohort are known words. That is, the known word cohort competitor might be favored, as connections to existing words are more automatic and/or consolidated than ones to newly acquired words. Moreover, this pattern may differ for spoken (Experiments 2-3) and written (Experiment 4) words, with less cohort activation observed in written words. There were no trials in which both the target and the cohort were a newly acquired word; this was a consequence of the design for the learning questions (the primary focus of this experiment series), requiring novel words to be phonologically dissimilar.

There were a couple of important differences between the current study and the one carried out by Weighall et al. (2017). First, participants' exposure to words was much more extensive in this study than the one by Weighall et al. (2017): Participants completed 9

trials/word in Weighall et al.'s (2017) study before testing, but 21 trials/word[8] in Experiments 2 - 4. Moreover, newly trained words were never used as the target in Weighall et al. (2017), thus not allowing for the test of whether novel words are slowed down by existing words. This might have also biased participants to consider novel words, as they could rule them out as referents in those trials. Thus, Question 4 may add to the ongoing debate of how lexical engagement arises in word learning by examining how quickly newly acquired words can be activated.

Due to the similarity of the design and statistical approach chosen for analysis, real-time word recognition in the three experiments are discussed together. In these analyses, we treat the word recognition portion of Experiment 3 as a replication of the one conducted as part of Experiment 2.

## 6.2 Methods

Participants were the same as in Experiments 2, 3 and 4. After completing the no-correct testing section, they participated in the real-time word recognition section, which was independent of the rest of the experiment and examined Question 4. The details of the training procedures are described as part of Chapters 3 and 4.

In Experiment 2, participants were trained on 16 auditory-word-object-mappings. They first completed a pre-training where each word was paired with two objects, one of which was the target and one the so-called featured competitor. In Experiment 3, participants were also trained on 16 mappings between auditory words and objects. However, in contrast to Experiment 2, words were paired more often with their featured competitor than the target during the pre-

[8] This number does not include testing trials during which no feedback was provided. However, given that unsupervised trials also include co-occurrence statistics, some learning likely also took part during the testing sections (c.f. Roembke & McMurray, 2016).

training (see Chapter 3). As a result, learning of the correct word-object-mappings was slower in Experiment 3 than in Experiment 2. Experiment 4 was identical to Experiment 2 with the exception that words were always presented in the written modality. Thus, participants learned mappings between 16 written words and objects. In the following sections, I will first describe the methods of all experiments before going to the results.

### 6.2.1 *Experiment 2 methods*

*Participants*

All participants (N = 40) that entered Experiment 2 also participated in this section of the experiment, with one exception. Due to technical issues with the eye-tracking set-up, one participant's time ran out before they reached the word recognition portion and could therefore not be included.

*Stimuli*

In Experiment 2, participants learned to associate 16 auditory words with objects. Newly learned words were always two-syllable pseudo words (see Table 6 for an overview); we will refer to these as nonwords to distinguish them from known (real) words. For each nonword, two real words were selected as cohorts. Cohorts overlapped with nonwords in the beginning consonant and vowel. For example, the two cohorts of the nonword BEAMLER were BEAKER and BEAVER. Real words were always easily imageable words. All stimuli were recorded by a female, native speaker of English. Per auditory stimuli, five exemplars were selected.

The same clipart images of novel objects were used as in the rest of Experiment 2. Moreover, existing clipart images of real words were edited, so that the style of novel objects and existing objects matched.

*Design*

Two word cohort competitors for each nonword were selected to allow for the test of cohort activation when both the target as well as the cohort were real words. Each trio of stimuli (the nonword and its two associated cohorts) could appear in all possible pairs, where one of the two was the target and the other the cohort. An example of this can be seen in the first half of Table 10. As a result, there were three trial types: word-word trials, where both the target and the cohort were a real word; word-nonword trials, where the target was a word and the cohort a nonword; nonword-word trials, where the target was a nonword and the cohort a word. To create all possible trial types, other trios of stimuli (nonword and its two associated cohorts) were matched to each set. The two nonwords of the two trios (e.g., BEAMLER and SAUBLE; see rest of Table 10) were put together to form nonword pairs, where no cohort was included. These trial types were presented twice as often as the others, to guarantee that nonwords were the targets to

Table 10: Overview of trial types used in word recognition portion of Experiments 2, 3 and 4. Trial types were created by combining two nonwords and their associated cohorts. Number refers to the relative number of trials of each type. Nonword-nonword trials were presented twice as often because there were fewer combinations of them.

| Trial type | Target | Cohort or Unrelated | Number |
|---|---|---|---|
| Word-word | BEAVER | BEAKER | 1 |
| Word-word | BEAKER | BEAVER | 1 |
| Word-nonword | BEAVER | BEAMLER | 1 |
| Word-nonword | BEAKER | BEAMLER | 1 |
| Nonword-word | BEAMLER | BEAVER | 1 |
| Nonword-word | BEAMLER | BEAKER | 1 |
| Word-word | SAUSAGE | SAUCER | 1 |
| Word-word | SAUCER | SAUSAGE | 1 |
| Word-nonword | SAUSAGE | SAUBLE | 1 |
| Word-nonword | SAUCER | SAUBLE | 1 |
| Nonword-word | SAUBLE | SAUSAGE | 1 |
| Nonword-word | SAUBLE | SAUCER | 1 |
| Nonword-nonword | BEAMLER | SAUBLE | 2 |
| Nonword-nonword | SAUBLE | BEAMLER | 2 |

a comparable extent as words. Nonword-nonword trials were filler trials and were never analyzed.

This construction process was repeated for all 16 nonwords and their associated cohorts, and resulted in eight sets of trial pairs (e.g., Table 10 is one example of those eight). To create two unrelated objects for each trial type, two sets of trial pairs were combined. More specifically, on each trial, one of the 16 trial pairs from the other set was randomly selected without replacement. For example, if BEAMLER and SAUBLE created a set of trials pairs as well as RAIMMER and WAMMOCK, unrelated objects for all trials associated with BEAMLER and SAUBLE (again see Table 10) could come from the trial pairs associated with RAIMMER and WAMMOCK. Each trial pair was repeated twice, resulting in an overall number of 256 trials (16 trials × 8 word sets × 2 repetitions). Trials were separated into two blocks of 128 trials and were randomized within block.

After running 13 participants, a confound in the design was noticed, which may have biased participants to look more to known words than nonwords. Thus, trials that used nonwords as unrelated objects were added. We only analyzed data from participants that completed the redesigned word recognition section (N = 27). The design section here describes how trials were created after the change to the word recognition section was made.

*Procedure*

After the completion of the no-correct testing session, participants entered the word recognition phase. First, participants were familiarized with the real-world words used in the study. In this short task participants saw one object with its written label provided underneath (e.g., an edited clipart picture of a beaker and the word BEAKER). The familiarization was included to guarantee participants' activation of the correct semantic concept based on the image

presented. This was not included for nonwords, as participants had been exposed to them repeatedly (with their referents) during the preceding experiment phases. Participants were instructed to press the space bar to move on to the next trial.

Subsequently, trials followed the same structure as during word learning: On each trial, participants saw four objects (300 x 300 pixels) in the four corners of the screen. In addition, a blue dot was presented in the center of the screen. The blue dot turned red after 1050 msec, cuing participants to click on it. This triggered the presentation of the auditory stimulus (the target) word over headphones. A trial ended when a participant selected the object that matched the word they had heard. No feedback was provided, and trials were never time-limited.

Participants' eye movements were tracked throughout the experiments. The same eye-tracking recording apparatus and method of analysis were used as in previous experiment sections. Participants underwent drift corrects approximately every 30 trials to guarantee the eye-tracker's adequate calibration throughout recording.

### 6.2.2 *Experiment 3 methods*

All of the participants took part in the updated version of the word recognition session. The same exclusion criteria were used as for the rest of Experiment 3. For two additional participants, very little eye movement data were recorded, likely due to a technical issue with the eye-tracker. As a result, it was not possible to analyze their data and they were excluded from all other analyses. This led to the overall exclusion of five participants (see Chapter CHAPTER 3: for details). In addition, one participant did not complete the word recognition section as they ran out of time; this left 36 participants for analysis. Experiment 3's word recognition was identical to the one used as part of Experiment 2.

### 6.2.3 *Experiment 4 methods*

21 participants took part in Experiment 4 after the word recognition section was updated. Only data from these subjects were analyzed. One participant ran out of time and therefore did not complete the word recognition section. One additional participant was excluded due to highly unusual eye movement patterns (high looks in the beginning of trial but not at end; overall very low number of looks), suggesting an issue with eye-tracking. This left 19 participants for analysis.

The word recognition section was identical to the updated one used in Experiment 2, with the exception that words were always presented in written form to match the rest of Experiment 4's design. During familiarization, words were written below the images and never covered with a mask. After familiarization, participants saw again four objects in the four corners of the screen as well as the blue dot. After 1050 msec, the blue dot turn red, signaling participants that they could click on it. Clicking on the red dot made it disappear, and the written word was presented in the center of the screen. The word was masked after a 75 msec delay and the mask ('#########') was presented for 100 msec in the center of the screen before disappearing as well.

## 6.3 Results and discussions

I discuss the results of each experiment in turn before a more general discussion.

### 6.3.1 *Experiment 2 results*

*Overview*

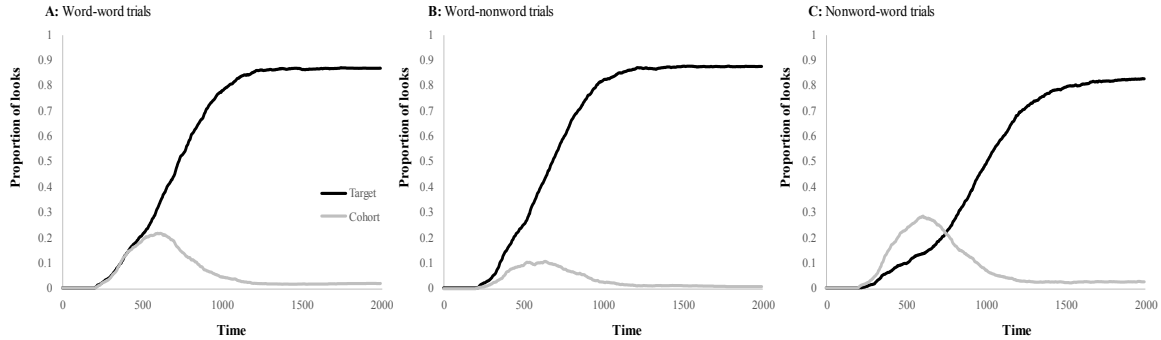


Figure 27: Looks to the target and cohort objects in word-word (Panel A), word-nonword (Panel B) and nonword-word trials (Panel C) in Experiment 2.

Accuracy during the word recognition section was high with experimental trial accuracy above 99% correct. Figure 27 depicts proportion of looks to all target and cohorts in trials where both targets and cohorts were words (Panel A). Participants' looks to the cohort competitor were equal to the target in the beginning of the word. This is evidence of incremental processing, a hallmark of speech perception (Allopenna et al., 1998). In panel B, looks during word-nonword trials are depicted, where the target was a word and the cohort a nonword. Looks to the cohort are relatively (peak height < 0.15) and are never equal to looks to the word; this suggests that words were activated more easily than nonwords. Finally, Panel C depicts nonword-word trials, where the target was a nonword and the cohort a word. At the beginning of the trial, participants only consider the word object, even as bottom-up information starts to favor the nonword target.

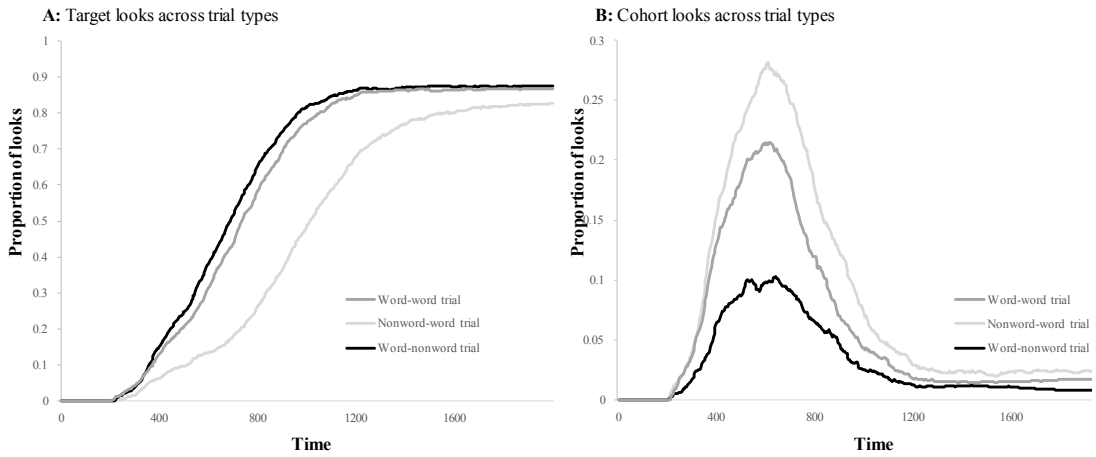


Figure 28: Looks to the target (Panel A) and cohort (Panel B) objects separated by trial type in Experiment 2.

To compare activation of target and cohorts across trial types, we next plotted looks to targets and cohorts separately (see Figure 28). Fixations converged on the target most rapidly when the target was a word and its cohort was a nonword (word-nonword trial). If both the target and its cohort were real words (word-word trial), activation was a little slower. Finally, target activation was the slowest if the target was a nonword and its cohort competitor a word (nonword-word trial). Moreover, cohorts were considered the most when the target was a nonword. This was followed in rank order by cohort activation in word-word trials and last in word-nonword trials. Importantly, activation of cohorts was similarly quick when they were words, but appeared to be slowed in word-nonword trials, where cohorts were nonwords.

*Statistical analysis*

To capture this pattern statistically, we fit nonlinear curves to the data and analyzed the parameters. For this purpose, a previously developed curve fitting methodology was used (Farris-Trimble & McMurray, 2013; McMurray, Farris-Trimble, & Rigler, 2017; McMurray et al., 2010), during which the shape of the timecourse of looking to each competitor (e.g., target, cohort) is estimated for each participant and each trial type separately. We then used the

parameters describing these curves as the basis of analyses intended to characterize the timecourse of target activation and cohort suppression across conditions.

Target fixations

Target fixations between 0 and 2000 msec were fit to a four parameter logistic (see Panel A of Figure 29; McMurray et al., 2017). The cross-over point (in msec)—the midpoint of the function—reflects the overall delay of looks. The slope of the function at the cross-over describes the speed at which fixations builds. The upper asymptote represents the degree of final fixations, indicating commitment to the selection made relative to other competitors; the lower asymptote is typically close to 0, and was not analyzed here (McMurray et al., 2017). For targets, all curve fits were found to be excellent (mean of $R^2 = 0.996$). An overview of the estimated parameters is presented in Table 11.

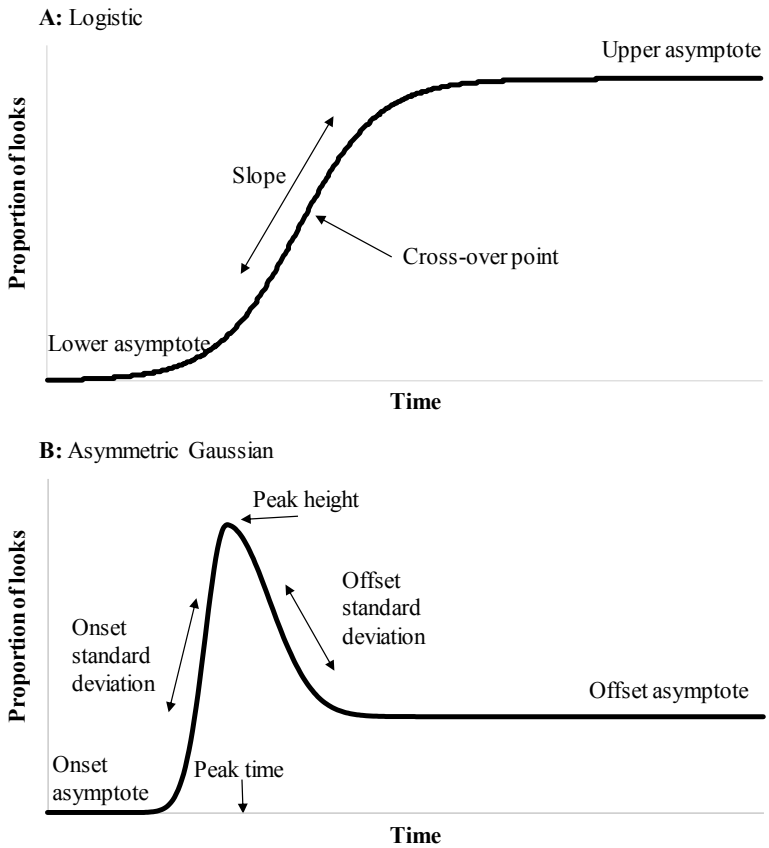


Figure 29: Overview of used parameters for curve fitting targets (logistic function, Panel A) and cohorts (asymmetric Gaussian function, Panel B).

Table 11: Mean and standard deviations for logistic fits of target fixations in Experiment 2.

| | M (SD) | | |
|---|---|---|---|
| | Word-word | Word-nonword | Nonword-word |
| Cross-over | 687.41 (67.63) | 641.97 (82.34) | 942.64 (93.51) |
| Slope | 0.0017 (0.0008) | 0.0017 (0.0004) | 0.0015 (0.0019) |
| Upper asymptote | 0.87 (0.12) | 0.88 (0.12) | 0.83 (0.12) |

To investigate differences in parameter due to trial type, two planned paired-samples t-tests were conducted for each parameter. Contrast 1 investigated differences in target fixations due to auditory stimulus identity (word or nonword), while holding cohort item identity constant (always word). Contrast 1 thus compared whether fixations to the auditory stimulus differed depending on whether it was a known word or a newly acquired nonword. Contrast 2 investigated differences in target fixations due to cohort item identity (word or nonword), while holding target item identity constant (always word). Contrast 2 compared whether fixations to the target differed depending on whether the cohort was a word or a nonword.

We started by examining contrast 1. Targets in word-word trials had an earlier cross-over point ($t(26) = -16.93$, $p < 0.001$, *Cohen's d* = 3.259) and higher upper asymptote than targets in nonword-word trials ($t(26) = -2.23$, $p = 0.035$, *Cohen's d* = 0.429). However, there was no difference in slope ($t(26) = -0.39$, $p = 0.702$). Target words were activated earlier and with more confidence than nonword targets.

Turning to contrast 2, targets in word-nonword trials had an earlier cross-over point than targets in word-word trials ($t(26) = 4.99$, $p < 0.001$, *Cohen's d* = 0.959). Neither slope ($t(26) = 0.43$, $p = 0.669$) nor upper asymptote level ($t(26) = -0.99$, $p = 0.331$) differed for targets in word-nonword and word-word trials. Thus, while the presence of a well-known lexical competitor impeded recognition (relative to a nonword cohort), this did not affect confidence in the target selection when the latter was a real word.

Cohort fixations

To analyze the cohorts, an asymmetric Gaussian function was fit to the cohort data. This function is the combination of two Gaussians with the same peak height and center position. However, they differ in their standard deviations and asymptotes at onset and offset (see Panel B of Figure 29; McMurray et al., 2017). The parameters investigated here are as follows: Peak height reflects the maximum amount of competition, and peak time indicates when this occurs. The onset standard deviation can be interpreted as the speed at which fixations reach maximum activation, whereas the offset standard deviation is the speed at which fixations diminish. The offset asymptote is the degree to which looks to the cohort are fully suppressed at the end of the trial. The onset asymptote was not investigated, as it is uniformly low (close to 0) across participants and trial types.

Looks to cohorts, particularly for nonword-word trials, were noisy and did not always conform to the expected function for some individuals (though they did as a whole). As a result, it was not possible to obtain accurate fits for individual subjects. In addition, it would have not been accurate to exclude participants based on number of cohort looks, as that could have biased data towards the inclusion of participants with high cohort looks. Thus, the jackknife method of Miller, Patterson, and Ulrich (1998) was used (also see Apfelbaum, Blumstein, & McMurray, 2011; McMurray, Clayards, Tanenhaus, & Aslin, 2008 for applications in eye-tracking) to compute more reliable estimates for each subject. In this technique, data are averaged across every participant minus one, then the curves are fit to that data. Next, we repeat this, cycling through the whole data set excluding each participant in turn. To analyze curve fit parameters, error terms were adjusted to reflect previous jackknifing, resulting in a conservative analysis (Apfelbaum et al., 2011; McMurray, Clayards, et al., 2008; Miller et al., 1998).

Curve fits for cohorts were excellent after jackknifing (mean of $R^2 = 0.997$). Averages and standard deviations after jackknifing for all parameters of interest are presented in Table 12. Note that standard deviations will be small because of previous jackknifing. As for target looks, two paired-samples t-tests (contrast 1: word-word trials vs. nonword-trials; contrast 2: word-word trials vs. word-nonword trials) were conducted for each parameter.

Table 12: Mean and standard deviations for Gaussian fits of cohort fixations in Experiment 2.

| | M (SD) | | |
|---|---|---|---|
| | Word-word | Word-nonword | Nonword-word |
| Peak height | 0.21 (< 0.01) | 0.10 (< 0.01) | 0.28 (< 0.01) |
| Peak time | 557.71 (2.94) | 551.17 (6.21) | 578.40 (3.76) |
| Onset standard deviation | 145.89 (1.79) | 138.84 (3.36) | 153.06 (2.00) |
| Offset standard deviation | 216.65 (2.28) | 249.16 (4.16) | 231.43 (2.73) |
| Offset asymptote | 0.02 (< 0.01) | 0.01 (< 0.01) | 0.02 (< 0.01) |

When comparing word-word and nonword-word trials, there was a significant difference in peak height ($t(26) = 4.42$, $p < 0.001$), indicating increased looks to word cohorts when the target was a nonword. Differences in offset asymptote between the two trial types were marginal ($t(26) = 2.01$, $p = 0.055$), and no other parameter was significantly different from each other (*ps* > 0.2). This indicates that participants were more likely to consider the word cohort than the nonword target, and might have had a harder time suppressing it even at the end of the trial.

For word-word and word-nonword trials, peak height differed as well ($t(26) = 7.06$, $p < 0.001$), suggesting decreased cohort competition in word-nonword trials relative to word-word trials. Differences in offset asymptote were marginal ($t(26) = 1.95$, $p = 0.062$), indicating that nonword cohorts were easier to suppress than word ones. No other parameter comparison across word-word and word-nonword trials were significantly different from each other (*ps* > 0.2). Nonword cohort competitors were activated less and tended to be suppressed more easily than known words.

Table 13: Overview of exploratory word recognition results as part of Experiment 2. "X" indicates a significant difference in parameter size, "(X)" stands for a marginal difference and "—" stands for no difference.

| Object | Parameter | Contrast 1: word-word vs. nonword-word | Contrast 2: word-word vs. word-nonword |
|---|---|---|---|
| Target | Cross-over | X | X |
| | Slope | — | — |
| | Upper asymptote | X | — |
| Cohort | Peak time | — | — |
| | Peak height | X | X |
| | Onset standard deviation | — | — |
| | Offset standard deviation | — | — |
| | Offset asymptote | (X) | (X) |

### 6.3.2 *Discussion Experiment 2 word recognition*

In Experiment 2, participants' looks to the target object were always delayed if the target was a nonword and sped up if the cohort was a nonword. In addition, final confidence in target selection was reduced only if the target item was a nonword. At the same time, activation built similarly across trial types. Cohort competition was lower when the cohort was a nonword and the target a word. However, it was higher when the cohort was a word but the target a nonword. Moreover, cohort suppression might have been reduced in nonword-word trials, just as cohort suppression might have been particularly strong in word-nonword trials (see summary in Table 13). One important limitation of the design was that no nonword-nonword trials were included to maximize learning of novel words. This leaves unclear whether nonword cohorts elicited comparably sized competition to word cohorts when the target was a nonword, too.

These findings match Weighall et al.'s (2017) of immediate competition in combination with differences in competition size (see also Magnuson, Tanenhaus, et al., 2003) and how long

competition lasted (i.e., how easily it was suppressed). Moreover, this investigation included trials where the nonword was the target. Despite excellent accuracy in those trials, participants' final confidence level in their target selection was still lower (reflected in the level of the upper asymptote of the target fixations), and nonword target trials showed higher cohort effects than word-word trials. Together, this suggests that even though participants had been exposed to the nonwords repeatedly for approximately an hour, they did not behave like known/existing words.

There are several possible reasons for why that may be the case: Consistent with the complementary learning systems account of word learning, participants might have formed links between words and meanings via hippocampal mediation between neocortical regions, where competition between words can be observed, but with a relatively slow timecourse (Dumay & Gaskell, 2007; Weighall et al., 2017). This stands in contrast to real words, which are integrated with existing knowledge in neocortical long-term memory as a result of sleep-based consolidation, thus allowing for swift competition with other lexical entries (i.e., lexical engagement).

Alternatively, trained nonwords might not behave as existing words due to other differences between them. More specifically, a newly acquired word may simply not possess the same strength of connections between its word form and meaning. If such differences in configuration were present, one would also expect trained nonwords to be activated more slowly and to compete less with existing words as a result. Under this account, the role of sleep may be to build bottom-up connections, not inhibition between words. If this were the case, sleep's role in lexical integration may be less "special" than previously assumed.

According to Weighall et al. (2017), subtle differences in activation patterns speak against the differences in configuration explanation: Again, low-frequency words can be used as

a proxy for newly acquired words. Low-frequency words are typically activated less strongly, but not necessarily at a very slow, extended timecourse (Dahan, Magnuson, & Tanenhaus, 2001)—as was observed here with newly acquired words. However, these effects are hard to disentangle in an eye-tracking paradigm, where slowed down looks to the target are easily caused by increased looks to the cohort competition and thus higher levels of competition. At this point, it is unclear whether a complementary learning systems account is necessary to explain the differences in activation patterns between spoken words and trained nonwords.

### 6.3.3 *Experiment 3 results*

*Overview*

Accuracy was high on experimental trials (98%). Given these high accuracy levels, analyses of the word recognition section of Experiment 3 might be considered a replication of Experiment 2's word recognition section. (Note the only difference between the two experiments was that pre-training in Experiment 3 paired words with their featured competitor more than with their target object—in contrast to Experiment 2.)

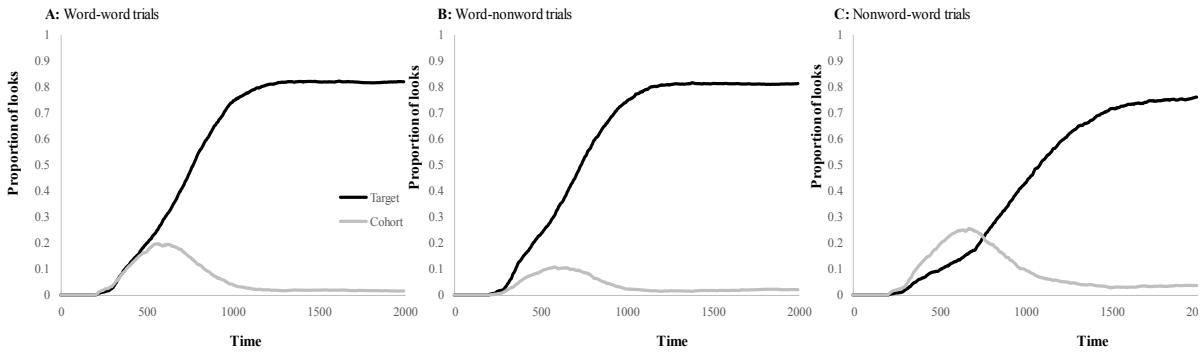


Figure 30: Looks to the target and cohort objects in word-word (Panel A), word-nonword (Panel B) and nonword-word trials (Panel C) in Experiment 3.

*Statistical analysis*

Target fixations

As for Experiment 2, a four parameter logistic was fit to target fixations. As before, fits were excellent (mean of $R^2 = 0.995$). Results are presented in Table 14. As can be seen in Figure 30, looks to targets and cohorts were similar to Experiment 2, with looks to the word target reaching high levels of looks at the end of the trial. In addition, looks to the nonword target were delayed in nonword-word trials (Panel C) with increased looks to the word cohort in the beginning of the trial.

Table 14: Mean and standard deviations for logistic fits of target fixations in Experiment 3.

| | M (SD) | | |
|---|---|---|---|
| | Word-word | Word-nonword | Nonword-word |
| Cross-over | 701.01 (60.54) | 666.88 (78.66) | 955.76 (104.14) |
| Slope | 0.0015 (0.0004) | 0.0015 (0.0004) | 0.0010 (0.0003) |
| Upper asymptote | 0.83 (0.16) | 0.81 (0.17) | 0.76 (0.16) |

For word-word and nonword-word trials, there were significant differences in cross-over point ($t(35) = -14.36$, $p < 0.001$, *Cohen's d* = 2.393), slope ($t(35) = 10.11$, $p < 0.001$, *Cohen's d* = 1.768) as well as upper asymptote levels ($t(35) = 4.39$, $p < 0.001$, *Cohen's d* = 0.732). This indicates that nonword targets were activated more slowly, that activation spread more slowly and that participants' final confidence in their response was never as high as for target words.

For the comparison of word-word and word-nonword trials, there was only a significant difference in cross-over point ($t(35) = 3.77$, $p = 0.001$, *Cohen's d* = 0.395) but not in slope ($t(35) = 0.35$, $p = 0.729$) or upper asymptote levels ($t(35) = 1.56$, $p = 0.121$). Together, these findings mostly replicate what was found in Experiment 2: Participants' looks to the target were delayed

if it was a nonword and if a nonword cohort was present. In addition, final confidence was never as high as when the target was an existing word.

Table 15: Mean and standard deviations for Gaussian fits of cohort fixations in Experiment 3.

| | M (SD) | | |
|---|---|---|---|
| | Word-word | Word-nonword | Nonword-word |
| Peak height | 0.20 (< 0.01) | 0.10 (< 0.01) | 0.25 (< 0.01) |
| Peak time | 574.61 (2.76) | 591.86 (4.11) | 621.73 (2.36) |
| Onset standard deviation | 155.60 (1.52) | 159.77 (2.17) | 174.47 (1.31) |
| Offset standard deviation | 210.74 (2.23) | 194.12 (3.36) | 240.66 (3.22) |
| Offset asymptote | 0.02 (< 0.01) | 0.02 (< 0.01) | 0.03 (< 0.01) |

Cohort fixations

Cohort fixations appear to be highest in word-word and nonword-word trials (Panel A and C of Figure 30), and lowest in word-nonword trials, where the cohort was a nonword (Panel B). Again we fit an asymmetric Gaussian to the jackknifed looks to cohorts. This resulted in excellent fits (mean of $R^2 = 0.996$). Averages and standard deviations are presented in Table 15. The same statistical analyses (two planned paired t-tests) were conducted and corrected for jackknifing (Apfelbaum et al., 2011; Miller et al., 1998).

We found a significant difference in cohort looks' peak time between word-word and nonword-word trials ($t(35) = 2.38$, $p = 0.023$), as a result of later peaks in the latter. In addition, peak height was significantly different between the two trial types ($t(35) = 5.49$, $p < 0.001$), with more looks to the cohort competitor in nonword-word than in word-word trials. Onset standard deviation marginally differed between word-word and nonword-word trials ($t(35) = 1.81$, $p = 0.079$) but not for the offset standard deviation ($p > 0.15$). Competition with cohorts build more quickly, was stronger and lasted for longer when the cohort was a word than when it was a nonword.

For the comparison of word-word and word-nonword trials, we found no difference in peak time ($p > 0.5$), but there was a difference in peak height ($t(35) = 8.36$, $p < 0.001$), with higher looks to the word cohort. Onset and offset standard deviations did not differ across trial types (all $p$s > 0.15). Finally, there was no difference in offset asymptote between word-word and word-nonword trials ($t(35) = 0.04$, $p = 0.967$). This indicates that overall competition dynamics were mostly comparable across cohort types, with the exception that word cohorts were activated more strongly than nonword ones.

Table 16: Overview of exploratory word recognition results as part of Experiment 3. "X" indicates a significant difference in parameter size, "(X)" stands for a marginal difference and "—" stands for no difference.

| Object | Parameter | Contrast 1: word-word vs. nonword-word | Contrast 2: word-word vs. word-nonword |
|---|---|---|---|
| Target | Cross-over | X | X |
| | Slope | X | — |
| | Upper asymptote | X | — |
| Cohort | Peak time | X | — |
| | Peak height | X | X |
| | Onset standard deviation | (X) | — |
| | Offset standard deviation | — | — |
| | Offset asymptote | X | — |

### 6.3.4 *Discussion Experiment 3 word recognition*

Overall, Experiment 3 replicated results from Experiment 2, particularly for contrast 2, which compared word-word and word-nonword trials (see Table 16 for an overview). Differences in target slope and cohort peak time only emerged in Experiment 3 when comparing word-word and nonword-word trials. There are a couple of different reasons why this might have been the case: The two experiments included two different samples of participants; therefore,

differences might be the consequence of random fluctuations in how people processed words. In addition, learning in Experiment 3 was overall lower than in Experiment 2, likely due to the differences in pre-training design. As a result, the system is less committed to the nonword target over its competitors, despite overall high levels of accuracy at the end of training. Nevertheless, these results are consistent with the idea that newly acquired nonwords can compete with existing words even in the absence of sleep-based consolidation—but that such competition is weaker and more extended than for familiar words.

### 6.3.5 *Experiment 4 results*

Finally, we turn to the results on written word recognition.

*Overview*

Participants' accuracy was high on experimental trials (mean = 97%). To investigate the integration of written words into the lexicon, we first explored target and cohort activation visually. In Figure 31, looks to the target and cohort competitor are split by trial type. Whereas the pattern of looks appears to be similar across word-word and word-nonword trials, looks to

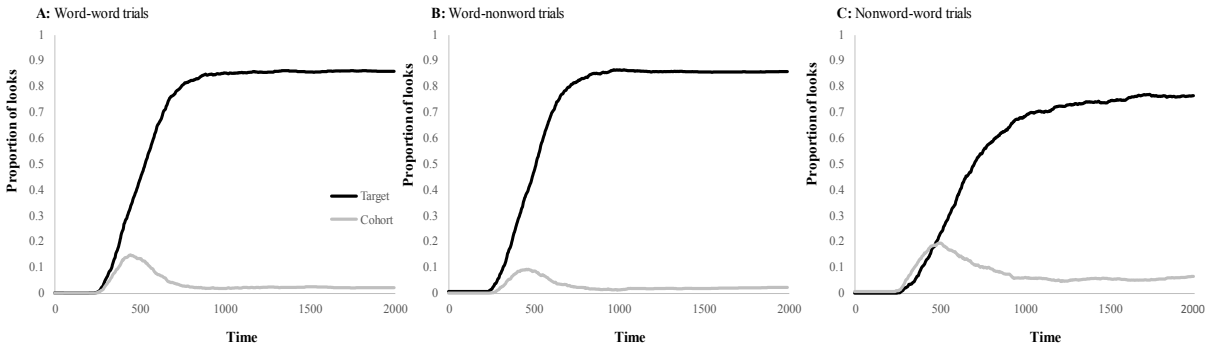


Figure 31: Looks to the target and cohort objects in word-word (Panel A), word-nonword (Panel B) and nonword-word trials (Panel C) in Experiment 4.

the target reach a lower asymptote in nonword-word trials. In addition, in the latter, cohort looks are more pronounced than in word-word and word-nonword trials.

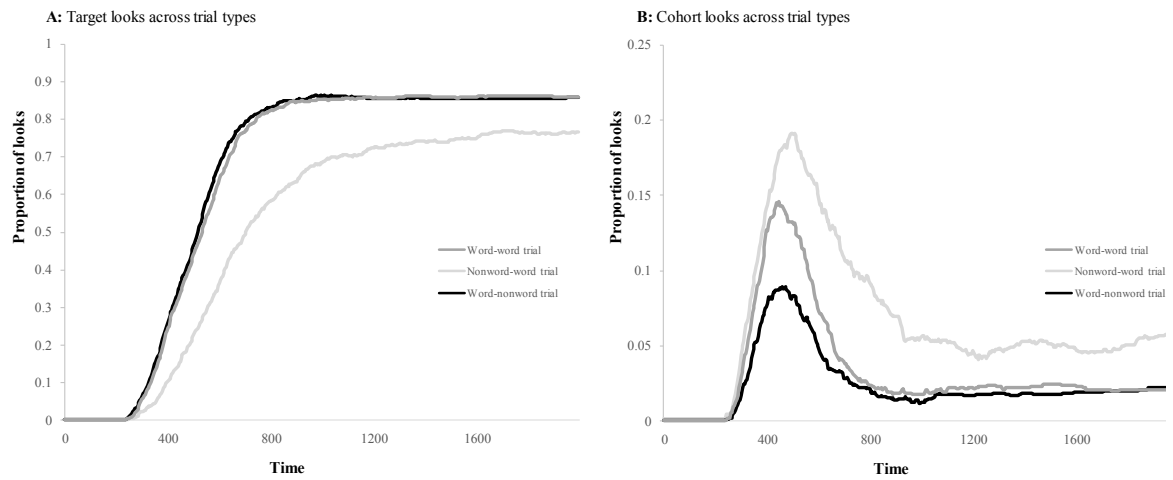


Figure 32: Looks to the target (Panel A) and cohort (Panel B) competitors across trial types in Experiment 4.

Another way to look at this is by plotting target and cohort looks separately. Panel A of Figure 32 indicates that target fixations for trials in which the target was a real word are almost identical (word-word and word-nonword trials), though words might have been activated more quickly when the cohort was a trained nonword but not a word. In contrast, nonword target fixations (Panel B) were slower and did not reach the same asymptotic level by the end of the trial. This pattern indicates that participants' activation of written words was swift, even in the presence of real-word cohort pictures.

Similarly, Panel B of Figure 32 shows the highest cohort activation in nonword-word trials, followed by word-word and word-nonword trials. Whereas nonword and word cohorts were suppressed quickly when the target was a word, cohort suppression was slower in nonword target trials.

There are two main differences with looks to target and cohort objects in the auditory version of this experiment: Target looks in nonword-word trials of written words never reached

the same upper asymptote levels as word targets; a similar pattern was present in Experiment 2 (the auditory version), but it was less pronounced. In addition, word cohorts were never fully suppressed in nonword-word trials—again, this can be observed in Experiment 2 (and the difference with word-word trials was marginal), but to a smaller extent than in written words.

*Statistical analysis*

Target fixations

To investigate this statistically, a four parameter logistic was fit target fixations; these results are presented in Table 17. Overall fit accuracy was high (mean of $R^2 = 0.996$). Differences in parameters due to trial type were investigated with two planned paired-samples t-tests (contrast 1: word-word and nonword-word trials; contrast 2: word-word and word-nonword trials).

Table 17: Mean and standard deviations for logistic fits of target fixations in Experiment 4.

| | M (SD) | | |
|---|---|---|---|
| | Word-word | Word-nonword | Nonword-word |
| Cross-over | 502.92 (57.35) | 486.50 (62.99) | 636.14 (94.22) |
| Slope | 0.0029 (0.0009) | 0.0037 (0.0023) | 0.0020 (0.0020) |
| Upper asymptote | 0.85 (0.10) | 0.85 (0.12) | 0.75 (0.12) |

When comparing word-word and nonword-word trials (contrast 1), there was a significant difference in cross-over point ($t(18) = -6.79$, $p < 0.001$, *Cohen's d* = 1.558), slope ($t(18) = 2.10$, $p = 0.050$, *Cohen's d* = 0.479) and upper asymptote level ($t(18) = 5.89$, $p < 0.001$, *Cohen's d* = 1.352). This indicates that participants' looks to the target were delayed when the target was a nonword, that activation built more slowly and that the system was less committed to the nonword over its competitors at the end of the trial.

In contrast, none of the parameters differed when comparing word-word and word-nonword trials (all *ps* > 0.1). This suggests that the nonword and word cohorts similarly influenced looks to the known target word.

Table 18: Mean and standard deviations for Gaussian fits of cohort fixations in Experiment 4.

| | M (SD) | | |
|---|---|---|---|
| | Word-word | Word-nonword | Nonword-word |
| Peak height | 0.14 (< 0.01) | 0.09 (< 0.01) | 0.18 (< 0.01) |
| Peak time | 440.24 (3.11) | 439.51 (3.97) | 458.98 (4.69) |
| Onset standard deviation | 76.95 (1.60) | 79.95 (1.49) | 90.38 (2.07) |
| Offset standard deviation | 130.19 (2.65) | 129.21 (4.38) | 208.23 (7.41) |
| Offset asymptote | 0.02 (< 0.01) | 0.02 (< 0.01) | 0.05 (< 0.01) |

Cohort fixations

An asymmetric Gaussian were fit to cohort looks after jackknifing. Curve fits for cohorts were excellent (mean of $R^2 = 0.996$). Means and standard deviations after jackknifing are presented in Table 18. As before, t-statistics were corrected for jackknifing, resulting in an overall conservative analysis (Apfelbaum et al., 2011; Miller et al., 1998).

Again, we start by comparing the word-word and nonword-word trials (contrast 1). Non-surprisingly, there was no significant difference in peak time between word-word and nonword-word trials ($t(18) = 1.09$, $p = 0.290$). There was a marginally significant effect of trial type when comparing cohort peak height between word-word and nonword-word trials ($t(18) = 2.01$, $p = 0.060$), but no differences in onset standard deviation ($p > 0.100$). Moreover, there were differences in offset standard deviation ($t(18) = 2.62$, $p = 0.018$) and offset asymptote ($t(18) = 4.16$, $p = 0.001$). In general, when the target was a nonword, cohorts were suppressed more slowly and less completely. Together, this suggests that cohorts were activated more strongly and less well suppressed when the target was a nonword instead of a word.

When comparing word-word and word-nonword trials, there was also no difference in peak time ($t(18) = 0.04$, $p = 0.968$; see Figure 32). However, differences in peak height emerged ($t(18) = 3.21$, $p = 0.005$). This indicates that target/cohort item identity influenced the amount of competition between words, even as the target was presented in written form. Moreover, neither onset standard deviation, offset asypmtote nor offset asymptote differed between the two trial types ($p$ always > 0.100).

Table 19: Overview of exploratory word recognition results as part of Experiment 4. "X" indicates a significant difference in parameter size, "(X)" stands for a marginal difference and "—" stands for no difference.

| Object | Parameter | Contrast 1: word-word vs. nonword-word | Contrast 2: word-word vs. word-nonword |
|---|---|---|---|
| Target | Cross-over | X | — |
| | Slope | X | — |
| | Upper asymptote | X | — |
| Cohort | Peak time | — | — |
| | Peak height | (X) | X |
| | Onset standard deviation | — | — |
| | Offset standard deviation | X | — |
| | Offset asymptote | X | — |

### 6.3.6 *Discussion Experiment 4 word recognition*

For written words, nonword cohorts exhibited little competition with existing words: This was reflected in the absence of any significant differences in target fixations between word-word and word-nonword trials. In addition, maximum cohort competition was significantly lower in word-nonword than in word-word trials. Moreover, nonword targets were activated later, more slowly and with lower confidence levels than real words. When the target was a nonword, competition with word cohorts was more difficult to suppress (see summary in Table 19).

Together, these findings indicate that newly acquired written nonwords—despite high accuracy levels during this section of the experiment—did not behave like fully fledged lexical representations. One glaring difference between Experiment 2 and Experiment 4 is that written nonwords never activated strongly enough to fully suppress word cohort competitors, as indicated by the significantly higher offset asymptote of cohort fixations in nonword-word trials. This potentially slower integration of written words is consistent with the protracted timecourse reported by Bakker et al. (2014).

As for spoken words, however, it is not straightforward to interpret differences in fixations between existing, familiar words and trained nonwords. Competition between words was observed when the knowledge/presence of existing words impacted the activation of nonwords. However, nonword cohorts did not slow down activation of words. The pattern observed here might then indicate that competition was not the result of lexical engagement, but rather was caused by differential activation of words (McMurray et al., 2016): Words' configuration might be more predictive of their activation than lexical engagement.

## 6.4 General discussion of real-time word recognition results

### 6.4.1 *Summary and discussion*

The goal Question 4 was to address how novel words are activated relative to known words. More specifically, we asked whether novel words were activated to the same extent as existing words, and whether they competed with real words. Across experiments, nonword targets were activated more slowly than existing words. In addition, final commitment to newly acquired words was lower than for known words. At the same time, looks to word competitors were higher when the target was not an existing word, and they were less easily suppressed.

Together, these findings suggest that activation of novel words was slowed, to some extent, likely as a result of increased consideration of existing words. One might consider this surprising, given that participants exclusively heard the novel words for approximately one hour before they completed the real-time word recognition section of the experiments. Even as people had more recent (and extensive) experience with the novel words, this did not outweigh their longstanding knowledge of real words.

Moreover, activation of nonword cohort competitors was lower (decreased peak height) than word cohort competitors. However, it is notable that the peak of competition occurred at the same time for both. This indicates that activation of nonwords happened as quickly as real words, but simply was not as strong. Even though these findings are theoretically consistent with the complementary learning account (Davis & Gaskell, 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003), they are also explainable by differences in activation functions due to differences in how well stimuli are known/configured (McMurray et al., 2016). Even as the newly acquired word is processed just as a quickly as a known one, each piece of information might provide less evidence for a novel word than for a known word due to weaker bottom-up connections, thus leading to less observed competition.

At this point, it is clear that sleep-based consolidation is not needed for a newly acquired word to behave as a previously known one (Kapnoula & McMurray, 2016; Kapnoula et al., 2015; Weighall et al., 2017); this is a key prediction of the complementary learning systems account of lexical integration. However, it is less clear where that leaves the rest of the CLS account. Moving forward, it is important to consider less complex explanations of competition effects—such as differences in activation across words—as possible alternative explanations.

Overall, real-time word recognition findings were similar for auditory and written words. However, a couple of differences appear: Written nonword cohorts competed less with word targets than auditory ones (based on visual inspection). Moreover, written nonwords never fully suppressed word cohort activations; this was not observed for auditory nonwords where complete suppression was observed. Together, this suggests that written words are more slowly integrated into the existing lexicon, consistent with Bakker et al.'s (2014) findings.

### 6.4.2 *Limitations*

One important limitation of the design employed is that novel words were not primarily selected to test differences in lexical configuration and engagement. As a result, no nonword cohorts were included to avoid nuisance variables during the learning portions of the experiment.

Additionally, the configuration of trials in this phase might have not been optimal for studying lexical integration of novel words. As real word stimuli had not been presented before, we were conscious of including a mix of foils on each trial to counteract potential novelty biases. It is a standard eye-tracking finding that participants make less eye movements as an experiment proceeds; as participants had been exposed repeatedly to nonwords before, this might have nonetheless influenced how much they looked to the real words and the nonwords. In contrast to the sub-phonemic mismatch paradigm used for previous investigations of lexical integration of novel words (Dahan, Magnuson, Tanenhaus, et al., 2001; Kapnoula & McMurray, 2016; Kapnoula et al., 2015), the paradigm used here does not include a real baseline: There is no way to test how easily words would have been activated if the nonword included as foils were completely novel.

Moreover, it should be noted that the real-time word recognition section of the experiments was conducted after the no-correct testing phase. Completing trials without the target right before the word recognition section might have lowered participants' overall levels of confidence into their selection, thus potentially leading to a reduction in looks to nonword targets. However, as accuracy remained high during the real-time word recognition section, this concern might be negligible. Nevertheless, all analyses should be treated as exploratory.

# CHAPTER 7: WORD RECOGNITION EFFICIENCY AND PRUNING

## 7.1 Overview

One of the reasons why pruning incorrect associations might be an important part of learning word-object-mappings is that the absence of incorrect connection might facilitate activation of the correct meaning. This is a prediction of the connectionist model of word learning by McMurray et al. (2012): The speed of word recognition was predicted by the strength of incorrect connections among words, not correct ones. That is, small, unpruned connections between an object and alternative words compete with the correct word. As a result, it is not possible to activate the correct word as quickly as if all incorrect connections with all alternative words were pruned.

To test this prediction, we decided to conduct additional analyses of the real-time word recognition phase of Experiments 2 and 3. It should be highlighted that none of the conducted experiments were designed to address whether pruning incorrect associations helped word recognition efficiency. Thus, all following analyzes should be treated as exploratory.

To investigate whether novel words were activated more easily when incorrect associations were pruned with the featured competitor, we took advantage of a design feature of the word recognition section. Pairs of four nonwords were created to construct the different trial types. To do so, a word and its featured competitor were always assigned to the same pair. Thus, in the nonword-nonword trials, the featured competitor was always one of the possible foils.[9] As a result, it was not possible to compare how quickly participants activated the correct meaning within subject or experiment (that would require a comparison to a nonword/nonword set in

[9] Please note that these were control trials in the rest of the word recognition analyses, and were never analyzed as part of Chapter 6.

which both words were foils). However, previous analyses (see Chapter 3) indicate that whereas pruning of incorrect associations was completed by the end of no-correct testing in Experiment 2, this may not have been the case in Experiment 3, where participants had been exposed frequently to associations between featured competitors and objects during pre-training. Thus, how quickly nonword targets are activated in nonword only trials in Experiments 2 and 3 might differ as the result of the presence of incorrect associations with the featured competitor in the latter. The goal of this Chapter was to explore this possibility.

## 7.2 Results

Accuracy during the word recognition phase was high in both Experiment 2 (99%) as well as in Experiment 3 (98%). As it was not originally planned to analyze nonword-nonword trials, there were only approximately 17 nonword-nonword trials per participant (these were originally included to be control trials; the exact number of trials that only included nonwords differed across participants due to how unrelated objects were selected [see Chapter 6]). Consequently, eye movement curves for individual participants were often unreliable. Data from one participant from Experiment 2 and two participants from Experiment 3 had to be excluded because their number of eye movements was two standard deviations below the average (e.g., total proportion of looks never crossed 0.5) in addition to the participants that were excluded in previous analyses. This left 26 participants for analysis in Experiment 2 and 34 participants for analysis in Experiment 3.

To conduct a comparison of word recognition parameters across experiments, participants' looks to the target object were first jackknifed (Apfelbaum et al., 2011; Miller et al.,

1998). Subsequently, a four parameter logistic function was fit to target looks (see Chapter 6 for a more detailed description of the analysis techniques).

The logistic curve fits resulted in three parameters of interest: cross-over, slope and upper asymptote. As described before, the cross-over point indicates overall delay of looks. In addition, the slope of target looks at the cross-over point is reflective of the speed at which activation builds. The upper asymptote indicates commitment to the nonword relative to other competitors at the end of the trial (McMurray et al., 2017). We predicted that target looks would be delayed (later cross-over point) and slower (lower slope) in Experiment 3, where pruning of incorrect associations with the featured competitor might not have been resolved by the end of training/testing, than in Experiment 2. Moreover, end-of-trial commitment to nonwords would be lower (lower upper asymptote) in Experiment 3 than 2.

Table 20: Mean and standard deviations for logistic fits of target fixations in nonword-nonword trials of Experiments 2 and 3.

| | M (SD) | |
|---|---|---|
| | Experiment 2 | Experiment 3 |
| Cross-over | 859.53 (5.67) | 882.52 (4.41) |
| Slope | 0.0010 (< 0.0001) | 0.0009 (< 0.0001) |
| Upper asymptote | 0.813 (0.01) | 0.798 (0.01) |

Overall, reliability of curve fits was very high (mean of $R^2 = 0.999$). Means for cross-over, slope and upper asymptote of target looks during nonword-nonword trials are presented in Table 20. As predicted, means indicate a later cross-over point and lower upper asymptote in Experiment 3 than in Experiment 2, though differences in slope appear to be minimal. It should be noted that due to jackknifing the data, standard deviations are very small.

To investigate differences across experiments statistically, parameters were separately entered into an independent t-test with experiment (Experiment 2 vs. 3) as the between-subject

variable. Error terms were adjusted to reflect previous jackknifing (Apfelbaum et al., 2011; McMurray, Clayards, et al., 2008; Miller et al., 1998).

There was a significant difference in cross-over point between the experiments ($t(58) = 25.61$, $p < 0.001$): Novel words were more slowly activated in Experiment 3 than in Experiment 2. Moreover, there was also a significant difference in slope ($t(58) = 17.11$, $p < 0.001$), with activation building more slowly in Experiment 3 than 2. Finally, there was a significant difference in upper asymptote level ($t(58) = 9.08$, $p < 0.001$), indicating less commitment to nonword targets at the end of the trial in Experiment 3 than 2.

## 7.3 Discussion

Together, these results indicate that word recognition efficiency was lower in Experiment 3 than in Experiment 2. This provides some preliminary evidence that efficiency in word recognition depends on the absence of incorrect associations, as predicted by the computational model by McMurray et al. (2012). Importantly, these conclusions should be considered with caution given the exploratory nature of these analyses.

Given that target activation of novel words is still relatively high at the end of a trial (upper asymptote levels were at above 0.8), one might wonder why these differences in the dynamics of recognition as a function of the degree of pruning matter. If a word is activated more slowly or with less overall confidence, this might have negative effects on downstream processing. Such effects might become clearer when considering real language processing, where words are not presented in isolation but are rather heard in a stream of sentences: If each word is processed a little bit more slowly than it should be, this could add up to substantive delays by the end of a sentence. In fact, higher levels of incorrect spurious associations might

account for difficulties in lexical processing in children with Developmental Language Disorder (Nation, 2014).

Future research should explicitly address whether efficiency in word recognition is better if words do not possess incorrect associations with other meanings or not, testing the prediction made by McMurray et al.'s model (2012) more directly. For this purpose, presence of a featured competitor should be manipulated across trials but within-subject.

# CHAPTER 8: OVERALL DISCUSSION

## 8.1 Summary

The goal of this dissertation was to determine whether the pruning and/or unlearning of incorrect associations is operative during human word learning. For this purpose, I conducted six experiments, five of which used eye-tracking to measure associative strength between items across learning. In the last experiment, a yes/no task was used to measure associative strength across experiment phases. It should be noted that I will use the words "pruning" and "unlearning" interchangeably to refer to the elimination of incorrect associations.

To summarize, participants were able to unlearn incorrect associations, though it is unclear whether pruning of incorrect associations between words and featured competitors was complete, or if small amounts remained. Experiment 6's results suggest that small traces exist even after extensive unlearning and that these associations can impact future learning, even as participants are near-ceiling when selecting the target object.

Moreover, Experiments 2, 4 and 5 indicated that mappings between symbols that are not auditory words (e.g., written words) and objects are more difficult to acquire than auditory-word-object mappings (c.f., Roembke, Wiggs, & McMurray, 2018). In addition, the unlearning of incorrect associations might be less robust when to-be-acquired mappings are not between auditory words and objects but between other symbols. This could indicate a symbiotic relationship between learning rate and pruning: Unlearning incorrect associations might facilitate the selection of the correct object and/or selecting the target object might make it easier to eliminate incorrect associations with other referents.

In the second half of the dissertation, I addressed an exploratory question: How quickly are newly acquired words activated relative to known words? For this purpose, participants in

Experiments 2-4 completed an additional word recognition section. In this phase, trials included either newly acquired words (the nonwords participants had been trained on) or existing words (e.g., CAMEL, BEAVER). Participants were slower at processing nonwords than real words, and nonwords competed less with real words than other existing words. Moreover, competition between existing words and nonwords was lower if they were presented orthographically instead of auditorily.

Together, these data are consistent with previous results that competition between existing and novel words (that were learned the same day) differs from the competition between known words. However, from our data, it is not clear that differences in competition are—as previous accounts have suggested (Dumay & Gaskell, 2012; Weighall et al., 2017)—qualitatively or whether they are simply quantitatively different.

## 8.2 Future directions

One conclusion became clear while working on this dissertation: Negative associative learning—pruning and unlearning—is a relatively unexplored area of research in vocabulary acquisition. Even though the conducted research in this dissertation offers some insight, findings were often not conclusive. Thus, there are plenty of unresolved questions.

First, future research should further examine the relationship between supervised and unsupervised learning. While this dissertation suggested that unsupervised statistics—at least in the context of human word learning—might interact and under some circumstances even outweigh supervised ones, one should directly test this explanation. Moreover, it is not clear whether this conclusion generalizes to other areas of learning and/or populations (e.g., children). More specifically, one should ask: Under what circumstances are associations strengthened, and

under what circumstances are associations pruned? Is one or the other more likely in the presence or absence of feedback?

Second, results from Experiment 6 suggest that latent associations between words and incorrect meanings are maintained even as participants select the target object at near-ceiling levels. This opens up the question: What type of information or experience is needed to eliminate incorrect associations completely (if possible)? One possibility is that training blocks simply were not long enough to allow for complete unlearning of incorrect associations. Alternatively, it is possible that sleep is needed to unlearn incorrect associations permanently (e.g., it could be that only associations above a certain strength are consolidated by sleep into long-term memory). Future experiments are needed to address these questions; these experiments should span across more than one day and control time between unlearning and testing the strength of incorrect associations.

Relatedly, in recent debates of word learning, one argument that has been made is that children wait for "diamonds"—situations in which it is very clear which object was named, thus facilitating the formation of a word-object-mapping and overcoming referential ambiguity (Medina et al., 2011; Trueswell et al., 2013). A similar argument could be made in the context of unlearning of incorrect associations, proposing that specific situations where an incorrect association is explicitly rejected are needed for unlearning. In all experiments conducted as part of this dissertation, accuracy tended to be high. As a result, participants were much more likely to receive positive feedback after selecting the correct object (the target) than receiving negative feedback after selecting the incorrect one. It is possible that feedback after correct selections was not a strong enough error signal to prune incorrect associations. Thus, future research could

address how information across trials/within one trial influences unlearning of incorrect associations.

Third, another open question based on Experiment 6 is how the maintenance of latent incorrect associations relates to other aspects of language processing. For example, it is well-established that the lexicon—the collection of words one knows—is structured based on words' semantic similarity (e.g., high overlap of features) as well as their co-occurrence in language input (Hills et al., 2010, 2009). Thus, it has been proposed that lexical structure reflects statistics in the learning environment that people are exposed to. Are latent associations reflected in lexical structure? This is an intriguing possibility, and could be explored further in future research.

## 8.3 Conclusions

The overall goal of this dissertation was to study how many-to-many mappings are acquired, and whether negative associative learning—the pruning or unlearning of incorrect associations—plays a role in this process. Across six experiments, we found evidence that this is the case in the context of word learning: Participants pruned incorrect associations as they were trained on the correct word-object-mappings. Moreover, incorrect associations between words and objects slowed down acquisition (as seen by the comparison of Experiments 2 and 3) and may have negatively impacted real-time word recognition efficiency. This underlines negative associative learning's importance in people's acquisition of many-to-many mappings, and specifically word learning.

REFERENCES


Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Apfelbaum, K. S., Blumstein, S. E., & McMurray, B. (2011). Semantic priming is affected by real-time phonological competition: evidence for continuous cascading systems. *Psychonomic Bulletin & Review*, *18*(1), 141–149. https://doi.org/10.3758/s13423-010-0039-8

Apfelbaum, K. S., Hazeltine, E., & McMurray, B. (2013). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology*, *49*(7), 1348–1365. https://doi.org/10.1037/a0029839

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: associative mechanisms in early word learning. *Cognitive Science*, *35*(6), 1105–1138. https://doi.org/10.1111/j.1551-6709.2011.01181.x

Apfelbaum, K. S., & McMurray, B. (2016). Learning during processing: Word learning doesn't wait for word recognition to finish. *Cognitive Science*. https://doi.org/10.1111/cogs.12401

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*(1), 116–130. https://doi.org/10.1016/j.jml.2014.03.002

Bauer, P. J. (2005). Developments in declarative memory. *Psychological Science*, *16*(1), 41–47. https://doi.org/10.1111/j.0956-7976.2005.00778.x

Berens, S. C., Horst, J. S., & Bird, C. M. (2018). Cross-situational learning is supported by propose-but-verify hypothesis testing. *Current Biology*, *28*(7), 1132–1136.e5. https://doi.org/10.1016/j.cub.2018.02.042

Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53. https://doi.org/10.1016/j.cognition.2012.08.008

Blumberg, M. S., Coleman, C. M., Gerth, A. I., & McMurray, B. (2013). Spatiotemporal structure of REM sleep twitching reveals developmental origins of motor synergies. *Current Biology*, *23*, 2100–2109. https://doi.org/10.1016/j.cub.2013.08.055

Borquez, M., Born, J., Navarro, V., Betancourt, R., & Inostroza, M. (2014). Sleep enhances inhibitory behavioral control in discrimination learning in rats. *Experimental Brain Research*, *232*(5), 1469–1477. https://doi.org/10.1007/s00221-013-3797-5

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*, B45-54. https://doi.org/10.1016/j.cognition.2005.02.002

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*(August), 17–29. Retrieved from http://www.mendeley.com/research/acquiring-single-new-word-1/

Changeux, J. P., & Danchin, A. (1976). Selective stabilisation of developing synapses as a mechanism for the specification of neuronal networks. *Nature*, *264*, 705–712. https://doi.org/10.1038/264705a0

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367. https://doi.org/10.1006/cogp.2001.0750

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5–6), 507–534. https://doi.org/10.1080/01690960143000074

Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 892–903. https://doi.org/10.1037/a0035657

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in english: evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 668–687. https://doi.org/10.1037/0096-1523.32.3.668

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*(4), 803–820. https://doi.org/10.1162/jocn.2009.21059

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111

Dumay, N., & Gaskell, G. M. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition*, *123*(1), 119–132. https://doi.org/10.1016/j.cognition.2011.12.009

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York, NY: Teachers College Press.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301–306. https://doi.org/10.1016/j.tics.2004.05.003

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

Farris-Trimble, A., & McMurray, B. (2013). Test – retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *56*(August), 1328–1346. https://doi.org/10.1044/1092-4388(2012/12-0145)group-level

Ferguson, B., Havy, M., & Waxman, S. R. (2015). The precision of 12-month-old infants' link between language and categorization predicts vocabulary size at 12 and 18 months. *Frontiers in Psychology*, *6*(August), 1319. https://doi.org/10.3389/fpsyg.2015.01319

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16420121

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fernandes, T., Kolinsky, R., & Ventura, P. (2009). The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. *Cognition*, *112*(3), 349–366. https://doi.org/10.1016/j.cognition.2009.05.002

Fitneva, S. A., & Christiansen, M. H. (2015). Developmental changes in cross-situational word learning: The inverse effect of initial accuracy. *Cognitive Science*, 1–21. https://doi.org/10.1111/cogs.12322

Forster, K. I., & Veres, C. (1998). The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(2), 498–514. https://doi.org/10.1037/0278-7393.24.2.498

Fukkink, R. G., Blok, H., & De Glopper, K. (2001). Deriving word meaning from written context: A multicomponential skill. *Language Learning*, *51*(3), 477–498. https://doi.org/10.1111/0023-8333.00162

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Gershkoff-Stowe, L. (2001). The course of children's naming errors in early word learning. *Journal of Cognition and Development*, *2*(2), 131–155. https://doi.org/10.1207/S15327647JCD0202_2

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*(1), 99–108. https://doi.org/10.1037/0012-1649.28.1.99

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gonzalez-Gomez, N., Poltrock, S., & Nazzi, T. (2013). A “bat” Is easier to learn than a “iab”: Effects of relative phonotactic frequency on infant word learning. *PLoS ONE*, *8*(3), e59601. https://doi.org/10.1371/journal.pone.0059601

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, *11*(4), 341–363. https://doi.org/10.1093/applin/11.4.341

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). Experience and brain development. *Child Development*, *58*(3), 539–559. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3038480

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, *61*(4), 481–502. https://doi.org/10.1016/j.jml.2009.08.001

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–B34. https://doi.org/10.1016/S0010-0277(02)00186-5

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491–528. https://doi.org/10.1037/0033-295X.106.3.491

Hearst, E., Besley, S., & Farthing, G. W. (1970). Inhibition and the stimulus control of operant behavior. *Journal of the Experimental Analysis of Behavior*, *14*(3S2), 373–409. https://doi.org/10.1901/jeab.1970.14-s373

Henderson, L., Powell, A., Gaskell, M. G., & Norbury, C. (2014). Learning and consolidation of new spoken words in autism spectrum disorder. *Developmental Science*, *17*(6), 858–871. https://doi.org/10.1111/desc.12169

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online lexical competition during spoken word recognition and word learning in children and adults. *Child Development*, *84*(5), 1668–1685. https://doi.org/10.1111/cdev.12067

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*, 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks. *Psychological Science*, *20*(6), 729–739. https://doi.org/10.1111/j.1467-9280.2009.02365.x

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention in 24-month-old infants. *Infancy*, *13*(2), 128–157. Retrieved from http://dx.doi.org/10.1080/15250000701795598

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, *15*(2), 125–150. https://doi.org/10.1111/j.1532-7078.2009.00009.x

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*, B23–B32. https://doi.org/10.1016/j.cognition.2004.10.003

Huttenlocher, J. (1974). The origin of language comprehension. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 331–368). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex - developmental changes and effects of aging. *Brain Research*, *163*(2), 195–205. https://doi.org/10.1016/0006-8993(79)90349-4

Huttenlocher, P. R., & Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *The Journal of Comparative Neurology*, *387*(2), 167–178. https://doi.org/10.1002/(SICI)1096-9861(19971020)387:2<167::AID-CNE1>3.0.CO;2-Z

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*, 190–211. https://doi.org/10.1016/j.jecp.2017.08.010

Kalashnikova, M., Mattock, K., & Monaghan, P. (2015). The effects of linguistic experience on the flexible use of mutual exclusivity in word learning. *Bilingualism*, *18*(4), 626–638. https://doi.org/10.1017/S1366728914000364

Kapnoula, E. C., & McMurray, B. (2016). Newly learned word forms are abstract and integrated immediately after acquisition. *Psychonomic Bulletin & Review*, *23*(2), 491–499. https://doi.org/10.3758/s13423-015-0897-1

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99. https://doi.org/10.1016/j.cognition.2014.09.007

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

Lindsay, S., & Gareth Gaskell, M. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(2), 608–622. https://doi.org/10.1037/a0029243

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3 Pt 1), 1242–1255.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing sopken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/MPG.0b013e3181a15ae8.Screening

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Mackintosh, N. J. (1974). *The psychology of animal learning*. London: Academic Press.

Magnuson, J. S., Dixon, J. a, Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133–156. https://doi.org/10.1080/03640210709336987

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: A tale of two systems? *Cognitive Science*. https://doi.org/10.1016/S0364-0213(03)00067-3

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Marascuilo, L. A. (1970). Extensions of the significance test for one-parameter signal detection hypotheses. *Psychometrika*, *35*(2), 237–243. https://doi.org/10.1007/BF02291265

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77. https://doi.org/10.1016/0364-0213(90)90026-S

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*(4), 653–675. https://doi.org/10.1037/0033-295X.101.4.653

Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language*, *47*(3), 343–359. https://doi.org/10.1016/S0749-596X(02)00037-2

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*(2013), 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–369. https://doi.org/10.1016/j.tics.2006.06.007

McGregor, K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1845–1856. https://doi.org/10.1044/1092-4388(2013/12-0233)

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838), 631. https://doi.org/10.1126/science.1144073

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3011988&tool=pmcentrez&rendertype=abstract

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071.

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition*, 147–164. https://doi.org/10.1016/j.cognition.2017.08.013

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

McMurray, B., Kapnoula, E. C., & Gaskell, M. G. (2016). Learning and integration of new word-forms: Consolidation, pruning and the emergence of automaticity. In M. G. Gaskell & J. Mirković (Eds.), *Speech Perception and Spoken Word Recognition*. London, UK: Taylor and Francis.

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*, 99–115. Retrieved from http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=70919

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*(2), 233–253. https://doi.org/10.2307/747758

Namy, L. L., Campbell, A. L., & Tomasello, M. (2004). The changing role of iconicity in non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures. *Journal of Cognition and Development*, *5*(1), 37–57. Retrieved from papers3://publication/uuid/7B33814C-779F-4D48-BD1D-3593A2352DD4

Nation, K. (2014). Lexical learning and lexical processing in children with developmental language impairments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 1–10. https://doi.org/10.1098/rstb.2012.0387

Newport, C., Wallis, G., Temple, S. E., & Siebeck, U. E. (2013). Complex, context-dependent decision strategies of archerfish, Toxotes chatareus. *Animal Behaviour*, *86*(6), 1265–1274. https://doi.org/10.1016/j.anbehav.2013.09.031

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*. https://doi.org/10.1016/S0010-0285(03)00006-9

Overman, W. H., Pate, B. J., Moore, K., & Peuster, A. (1996). Ontogeny of place learning in children as measured in the Radial Arm Maze, Morris Search Task, and Open Field Task. *Behavioral Neuroscience*, *110*(6), 1205–1228. https://doi.org/10.1037/0735-7044.110.6.1205

Pascalis, O., De Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*, 1321–1323. https://doi.org/10.1126/science.1070223

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56

Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. Cambridge, MA: The MIT Press.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rakic, P., Bourgeois, J., Eckenhoff, M., Zecevic, N., & Goldman-Rakic, P. (1986). Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science*, *232*(4747), 232–235. https://doi.org/10.1126/science.3952506

Rescorla, R. A. (2001). Retraining of extinguished Pavlovian stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(2), 115–124. https://doi.org/10.1037/0097-7403.27.2.115

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II Current Research and Theory* (Vol. 20, pp. 64–99). Appleton-Century-Crofts. Retrieved from http://homepage.mac.com/sanagnos/rescorlawagner1972.pdf

Roembke, T. C., Hazeltine, E., Reed, D., & McMurray, B. (2018). Automaticity of word recognition is a unique predictor of reading fluency in middle-school students. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000279

Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language*, *87*, 105–127. https://doi.org/10.1016/j.jml.2015.09.005

Roembke, T. C., Wasserman, E. A., & Mcmurray, B. (2016). Learning in rich networks involves both positive and negative associations. *Journal of Experimental Psychology: General*, *145*(8), 1062–1074. https://doi.org/10.1037/xge0000187

Roembke, T. C., Wiggs, K. K., & McMurray, B. (2018). Symbolic flexibility during unsupervised word learning in children and adults. *Journal of Experimental Child Psychology*, *175*, 17–36. https://doi.org/10.1016/j.jecp.2018.05.016

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3011987&tool=pmcentrez&rendertype=abstract

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for Parallel Distributed Processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 45–76.

Samuelson, L. K., & McMurray, B. (2017). What does it take to learn a word? *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1–2). https://doi.org/10.1002/WCS.1421

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*, *14*(5), 238–242. https://doi.org/10.1111/j.0963-7214.2005.00372.x

Share, D. L. (2011). On the role of phonology in reading acquisition: The self teaching hypothesis. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining individual differences in reading: Theory and evidence* (pp. 45–68). New York, NY: Psychology Press.

Sharon, T., Moscovitch, M., & Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(3), 1146–1151. https://doi.org/10.1073/pnas.1005238108

Shatz, C. (2002). Emergence of order in visual system development. In M. Johnson, Y. Munakata, & R. O. Gilmore (Eds.), *Brain Development and Cognition: A Reader*. Oxford, UK: Blackwell Publishing.

Shiu, L. P., & Chan, T. C. (2006). Unlearning a stimulus - response association. *Psychological Research*, *70*, 193–199. https://doi.org/10.1007/s00426-004-0201-x

Siegler, R. S. (1989). Mechanisms of cognitive development. *Annual Reviews Psychology*, *40*, 353–379. https://doi.org/10.1146/annurev.psych.40.1.353

Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Sciences*, *3*, 430–435. https://doi.org/10.1016/S1364-6613(99)01372-8

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243. https://doi.org/10.1111/1467-9280.02438

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1–2), 39–91. https://doi.org/10.1016/S0010-0277(96)00728-7

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1–2), 39–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8990968

Smith, C. N., Urgolites, Z. J., Hopkins, R. O., & Squire, L. R. (2014). Comparison of explicit and incidental learning strategies in memory-impaired patients. *Proceedings of the National Academy of Sciences*, *111*, 475–479. https://doi.org/10.1073/pnas.1322263111

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17692305

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, *44*(5), 430–444. https://doi.org/10.1037/h0062885

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*(1), 32–71. https://doi.org/10.2307/747348

Steele, S. C. (2012). Oral definitions of newly learned words: An error analysis. *Communication Disorders Quarterly*, *33*(3), 157–168. https://doi.org/10.1177/1525740110387129

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278. https://doi.org/10.1038/nature04286

Storkel, H. L. (2001). Learning new words: phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, *44*, 1321–1337. https://doi.org/1092-4388/01/4406-1321

Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, 1312–1323. https://doi.org/10.1044/1092-4388(2003/102

Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics and Phonetics*, 407–425.

Stout, S. C., & Miller, R. R. (2007). Sometimes-Competing Retrieval (SOCR): A formalization of the Comparator Hypothesis. *Psychological Review*, *114*(3), 759–783. https://doi.org/10.1037/0033-295X.114.3.759

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition*, *163*, 93–102. https://doi.org/10.1016/j.cognition.2017.03.002

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/DOI 10.1126/science.7777863

Thorndike, E. L. (1898). *Animal intelligence: An experimental study of the associative processes in animals.* New York: NY: Macmillan.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science (New York, N.Y.)*, *277*(5324), 376–380. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9219696

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology*, *3*(46), 1–8. https://doi.org/10.3389/fpsyg.2012.00046

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729–742. https://doi.org/10.1016/j.cognition.2007.08.007

Walczyk, J. J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly*, *35*(4), 554–566. https://doi.org/10.1598/rrq.35.4.7

Warren, D. E., & Duff, M. C. (2014). Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus*, *24*(8), 920–933. https://doi.org/10.1002/hipo.22279

Wasserman, E. A., Brooks, D. I., & McMurray, B. (2015). Pigeons acquire multiple categories in parallel via associative learning: a parallel to human word learning? *Cognition*, *136*, 99–122. https://doi.org/10.1016/j.cognition.2014.11.020

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2017). Eye-tracking the time- course of novel word learning and lexical competition in adults and children. *Brain and Language*, *167*, 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

Weliky, M., & Katz, L. C. (1997). Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity. *Nature*, *386*, 680–685. https://doi.org/10.1038/386680a0

Werker, J. F., & Tees, R. C. (1984). Cross-language speech-perception - evidence for perceptual reorganization during the 1st year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/Doi 10.1016/S0163-6383(84)80022-3

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development*, *70*(1), 65–77.

Yee, E., Blumstein, S. E., & Sedivy, J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *Journal of Cognitive Neuroscience*, *20*(4), 592–612. https://doi.org/10.1162/jocn.2008.20056

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, *145*, 53–62. https://doi.org/10.1016/j.cognition.2015.07.013

Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, *21*(1), 1–22. https://doi.org/10.3758/s13423-013-0443-y