The Pennsylvania State University

The Graduate School

College of the Liberal Arts

STATISTICAL WORD LEARNING AND THE ROLE OF CONTEXTUAL CUES IN THE FORMATION OF MULTIPLE MAPPINGS

A Thesis in

Psychology

by

Timothy J. Poepsel

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2013

The thesis of Timothy J. Poepsel was reviewed and approved* by the following:

Daniel J. Weiss
Associate Professor of Psychology and Linguistics
Thesis Advisor

Reginald B. Adams
Associate Professor of Psychology

Judith F. Kroll
Distinguished Professor of Psychology, Linguistics, and Women's Studies

Kristin A. Buss
Associate Professor of Psychology

Mel Mark
Department Head
Professor of Psychology

*Signatures are on file in the Graduate School.

Abstract

A fundamental challenge for statistical learners is tracking statistics in environments that contain multiple inputs, as in bilingual language acquisition. To optimize learning, learners must be able to differentiate within-language variability from between-language variability, and accordingly decide when to assimilate input into an existing language representation or accommodate a new input by forming a new representation. Previous research from several language learning domains suggests that learners may be able to form separate representations for multiple inputs when provided with an adequate contextual cue (e.g., a shift from a male to female voice) that occurs at the point of transition from one input to another. Here this problem is examined in the domain of adult word learning. A series of four experiments investigates how several individual contextual cues modulate word-learners' application of the mutual exclusivity constraint (i.e., the preferential formation of one-to-one word-object mappings) to input containing many-to-one word-object mappings. In conditions lacking contextual cues, learners formed mutually exclusive one-to-one word-object mappings; by changing cues such as speaker voice and accent over successive presentations of training input, it was found that learners successfully formed many-to-one mappings. Additionally, contextual cues significantly reduced mapping ambiguity in conditions containing both one-to-one and many-to-one mappings, which in turn increased learners' confidence that they had correctly formed both types of mappings. These findings have a number of important implications for statistical learning and language acquisition: language learners appear to track statistics in a contextually sensitive manner, taking advantage of cues to optimize processing; and contextual cues, although not required for acquisition, serve as shortcuts for learners, reducing the overall amount of exposure necessary for learning when multiple inputs are available.

Table of Contents

List of Figures

Acknowledgments

There are many people who deserve acknowledgement for their role in helping me to complete this thesis, whether they are aware of their contribution or not. My parents, my advisors, and my cat, however, deserve explicit mention. Thank you.

1 Introduction

When experienced language learners listen to spoken language input, they hear hierarchically organized structure. There are phonemes that combine into syllables, which in turn combine to form words and phrases. However, across languages, there are no invariant acoustic cues that reliably mark boundaries between these structures. For example, although learners can use stress patterns to find boundaries between words within an utterance, stress patterns differ across languages. For a naive learner, this effectively precludes the use of stress in finding boundaries between words across languages (Polka and Sundara, 2003). Pauses are another salient feature of spoken input that learners can use to find word boundaries. Yet, due to articulatory processes, the distribution of pauses is inconsistent; they frequently occur word internally as well as between words, again preventing a naive learner from using this feature to reliably find word boundaries (Cole and Jakimik, 1980). Given the inconsistency of acoustic cues in marking word boundaries, it is natural to wonder how learners begin to hear spoken input as a series of words. Answering this question requires a critical examination of how much reliable information spoken input contains, and whether that information is sufficient to support the segmentation of speech into its component structures.

In approaching the question of how learners find structure in language input, there are two primary theoretical orientations. These orientations hold largely opposite views regarding the ability of learners to develop a complete representation of language from language input alone. An argument from the nativist perspective suggests that language input is impoverished relative to what adult speakers know about their language. To acquire language in the face of this "poverty of the stimulus", the learner must apply innate knowledge of the hierarchical structure of language (Chomsky, 1980; Pinker, 1989). A notable example of this is auxiliary fronting in English, which is the process of forming a question from a declarative sentence (Perfors, Tenenbaum and Regier, 2006). Given the sentence “The man in the restaurant is hungry”, a question can be formed by moving the auxiliary to the front of the sentence: “Is the man in the restaurant hungry?” From the perspective of a naive leaner, this movement can be accomplished in two ways (Chomsky, 1965). One option is to use rules that operate only on the linear sequence of words. With this “structure-independent” strategy, a learner would move the left-most auxiliary of the sentence to the front in order to form a question. Another option is to use rules that depend on the hierarchical structure of language. With this “structure-dependent” strategy, the learner would move the auxiliary of the main clause of the sentence (i.e., the “is” in “the man is hungry”) to the front of the sentence in order to form the question. In a sentence with only one auxiliary, using either of these strategies results in the correctly formed question “Is the man in the restaurant hungry?” In complex sentences containing multiple clauses and auxiliaries, such as “The man who is in the restaurant is hungry”, only the structure-dependent strategy results in the correct interrogative form. Employing the structure-independent strategy results in the incorrect sentence “Is the man who in the restaurant is hungry?”

The problem for a learner trying to determine the correct way to form interrogatives is that language input contains few examples that support the structure-dependent strategy (Lidz, Waxman, and Freedman, 2003; Legate and Yang, 2002). Because learners receive input that does not support one interpretation of the underlying syntax over another, they should be unable to distinguish correct from incorrect strategies. Yet children and adults alike do not produce errors of the sort mentioned above; rather, their productions reflect the correct structure-dependent analysis of syntax. This problem of insufficient input extends to many other grammatical "rules", but in all cases speakers overcome the limitations of the input and produce grammatically well-formed utterances (Regier and Gahl, 2004; Elman, 1993; Tomasello, 2000). Further compounding the problem of impoverished input is the presence of disfluencies in speech, false-starts, and unfinished utterances, which supply the learner with a constant source of noise (Fox Tree, 1995; Shriberg, 1996). This noise further distorts the structural relationships that a learner must find in order to "know" their language. With both the impoverishment and noisiness of the available input, a learner should be unable to rule out all of the possible hypotheses about how language is structured. Accordingly, speakers should produce errors that reveal their incorrect assumptions about language structure. As already pointed out, however, this is not the case.

The problems of impoverished and noisy input can be resolved by assuming that learners are equipped with innate, language-specific knowledge and learning biases. In one formulation of this theory, innate knowledge is operationalized as competing alternative strategies, or parameters, for analyzing aspects of syntax (such as auxiliary movement to form interrogatives) (Chomsky, 1981). Parameters that fit the available data are selected while the alternative parameters fade from competition. In this way, innate knowledge in the form of these parameters constrains the number of structural interpretations that are possible and facilitates the development of a complete representation of language. Modern nativist approaches even accommodate bidialectal or bilingual input by allowing the learner to maintain several parameters for a given syntactic feature, allowing them to activate those that are most consistent with the local linguistic evidence (Goldwater, Sharon, and Johnson, 2003). Thus, through exposure to language input and the application of innate knowledge to it, a learner can arrive at the correct interpretation of language structure.

In contrast to the Nativist position, numerous recent studies suggest that learners can infer the correct structural relationships in perceptual input through the use of domain-general learning mechanisms (i.e., mechanisms that are not tailored to input from any single domain) (Gomez and Gerken, 1999; Fiser and Aslin, 2001; Saffran et al, 1999). Thus, with mere exposure to a hierarchically structured visual, linguistic, or non-linguistic auditory signal, a learner can develop a complete representation of the relationships in that signal. One of the proposed mechanisms that underlies this ability to find structure is implicit learning, and specifically a sensitivity to how frequently perceptual events co-occur (Rebuschat and Williams, 2011; Yim and Rudoy, 2013). Through this sensitivity, events that often occur together in both time and place are assumed to be related as members of a higher-order structure. For instance, if the administration of certain drug and the curing of an illness often co-occur, it is reasonable to assume that the drug is related to the illness. Conversely, if the drug rarely or never cures the illness, it is reasonable to assume that there is no relationship between the drug and the illness.

This ability to track how frequently events co-occur has been termed statistical learning. Of relevance to this paper, statistical learning has been demonstrated to influence the processing and acquisition of linguistic structure (and as noted above, is thought to be an implicit learning mechanism, although the relatedness of implicit and statistical learning mechanisms is a topic of debate – see Perruchet and Pacton (2006) for a review). In a statistical learning account of language acquisition, the learner begins by inferring the most basic relationships within the signal (e.g., how often certain syllables co-occur, which is useful for finding where words begin and end). Once these relationships have been acquired, the learner is able to focus on larger chunks of the signal and use new sets of reliably co-occurring events to engender a more comprehensive set of strategies for inducing structures (e.g., if there are reliable patterns of stress placement within learned words, learners can then deploy a new strategy uing stress to identify word boundaries). Statistical learning has been posited to operate in the processes of phonetic learning (e.g. Maye, Werker, and Gerken, 2002; Maye, Weiss, and Aslin, 2008), speech segmentation (e.g., Saffran, Newport, and Aslin, 1996), word learning (e.g., Yu and Smith, 2007, Kachergis, Yu, and Shiffrin, 2010) and syntactic learning (Perfors, Tenenbaum, and Regier, 2006; Regier and Gahl, 2004). Importantly, a core assumption of statistical learning is that these reliable co-occurrence relationships exist in all languages. With this assumption, it follows that an innate sensitivity to statistical relationships would allow a human learner to acquire any language given sufficient exposure.

In a foundational study of statistical learning, Saffran, Newport, and Aslin (1996) tested the abilities of infants in an artificial language learning task. Learners were familiarized to a continuous stream of syllables. Unbeknownst to learners, the stream was composed of many "words", each 3 syllables in length. Furthermore, the probability with which certain syllables transitioned to others was not constant across the stream; these transitional probabilities were lower between words than within them. The authors hypothesized that if learners were sensitive to differences in transitional probabilities, they would be able to use the differences to find word boundaries, and consequently to learn the set of words that comprised the continuous stream. At test, learners reliably discriminated words of the language from "non-words" (i.e., syllable triplets that had never occurred) and "part-words" (i.e., syllable triplets that contained portions of two words). Importantly, the authors controlled all other acoustic and temporal features of the familiarization stream so that no cues to word boundaries, aside from the transitional probabilities, existed. The results of the study suggest that human learners track and store fine-grained features of linguistic input and may be able to accomplish the task of learning through sensitivity to reliable statistical relationships in the input. As statistical learners begin to extract and store more words from speech, the statistically reliable patterns in other features of these stored forms may emerge (Pelucchi and Saffran, 2009).

Learners can also utilize statistically reliable information from outside of the auditory domain to facilitate language acquisition. Thiessen (2010) synchronized the syllables of an artificial language (similar to that used in Saffran et al 1996) to video of a talking face. Compared to conditions that presented only auditory or visual input, learners in the multi-modal input condition were significantly more successful at test. These behavioral results show that humans are adept statistical learners who can take advantage of reliable patterns in the distributions of many linguistic and non-linguistic cues.

1.1 Variation in Learning Input

The vast majority of statistical learning studies present learners with a single underlying structure. This structure is statistically uniform, in that at any point in learning, the same relationships between elements are available to the learner. However, natural language contains several types of variation that may cause a learner's input to be statistically non-uniform. Within-language variation may take the form of noise (such as disfluencies, false-starts and unfinished utterances) or dialectical variation, which presents variation in phonological, lexical, or syntactic structure. Between-language variation presents systematic variation in phonological, lexical, and syntactic structure, although this variation may more extreme in nature (i.e., dialects may often be mutually intelligible, whereas separate languages may not be) (Tang and van Heuven, 2009).

Both within and between-language variation present a challenge to acquisition.

Specifically, learners must determine whether the variation they encounter in language input represents normal within-language variation, dialectical variation, or between-language variation (Weiss, Gerfen, Mitchel, 2009; Gebhart, Aslin, Newport, 2009). Subsequently, learners must decide whether to assimilate that input into an existing representation or accommodate the input by creating a new representation (Piaget, 1962). That is, if the variation is within-language, it should be incorporated into the existing representation for that language; alternatively, in the case of between-language variation, the learner must determine where transitions between languages occur, and subsequently create and maintain separate representations for each language. However, transitions between language inputs are frequently unannounced, or implicit. For example, code-switching involves fluid alternation between multiple languages, which may occur within or between utterances (Joshi, 1982). Determining the point at which one language transitions to another may be accomplished in several ways (Qian, Jaeger and Aslin, 2012). A learner encountering input with statistically reliable patterns may create a representation or model of that input. Using this model, the learner may be able to predict upcoming input. When variation is low, the error in a learner's predictions of upcoming input is low. In this case, the model accurately reflects the structure of the input. When variation is high, however, the model is unable to accurately predict upcoming input. Thus, high levels of prediction error may be one way in which learners detect points of transition.

Additionally, learners may approach the challenge of language acquisition with different prior expectations for a change in the structure of language input. For example, a learner consistently exposed to multiple languages may expect to frequently encounter transitions from one language to another. A learner in a monolingual environment, however, may expect to encounter only one language in most or all situations. These differences in prior expectation for a change in the input may thus influence the ability of learners to detect changes in the input. Several studies, in fact, have demonstrated that bilinguals are not bound to the same restrictive assumptions about language structure compared to their monolingual counterparts (Byers-Heinlein and Werker, 2010; Kovacs and Mehler, 2009a). Furthermore, changes in the environment (i.e., contextual cues) that are correlated with structural changes in language input may serve to alert the learner to the change. For instance, Kovacs and Mehler (2009) presented infants with speech stimuli that followed two patterns. Each pattern predicted the appearance of a visual reward in a different location on a screen. When the patterns were presented in an identical voice, monolingual infants were only able to learn one of the associations between pattern and reward location. However, when the pitch of the voice in which each pattern was presented differed, monolinguals were able to learn both associations. In this case, the contextual cue of a pitch change seemed to increase learners' awareness of the presence of multiple reward contingencies, thereby facilitating the acquisition of both. Thus, while experience with multiple languages may adjust a learner's expectations regarding the structural consistency of language input, contextual cues are also salient signals of changes in structure. Sensitivity to contextual cues and how they correlate with underlying linguistic structure may be a crucial component of resolving the ambiguity that arises from variation in language input.

1.2 Multiple Inputs and the Role of Context

To acquire multiple inputs, a learner must first be able to reliably discriminate those inputs. In the case of language acquisition, there are many cues that may signal to the learner that multiple inputs are present in the learning environment. Many of these cues may be language internal, arising from properties of the languages themselves. For example, there may be differences in stress patterns, phonological patterns, or phonemic inventory that learners can use to detect a transition from one language to another. Additionally, the learning environment itself may provide cues that reliably signal transitions between languages (i.e., external cues). For example, a learner may encounter one language in the home environment and another at work or in public spaces. Even within a home, a learner may encounter different languages depending on who they talk to (i.e., older and younger generations may speak different languages). Thus, to acquire multiple inputs, a learner must determine the set of language-internal and external cues that best encapsulate a language.

A number of recent studies outside the realm of statistical learning have begun to investigate language acquisition in the context of multiple inputs; interestingly, these studies have also investigated the influence of contextual cues on learning. Kovacs and Mehler (2009) exposed infants to speech-like stimuli that followed two patterns, both of which were learnable in isolation. Each pattern cued the appearance of a visual reward in a different location (e.g., pattern 1 cued the appearance of the reward on the right side of a computer screen, while pattern 2 cued a reward on the left side of the screen). The authors found that monolingual infants were able to learn one of the associations between pattern and reward location, but not the other. However, when a contextual cue differentiated the associations, such that the auditory stimuli used to present each pattern differed in pitch (analogous to input from different speakers), monolinguals were able to learn both associations. In a related study, infants were again exposed to speech-like auditory stimuli that cued the appearance of a visual reward in a certain location (Kovacs and Mehler, 2009). In a pre-switch phase, the reward appeared consistently on one side of the visual display, simultaneously with the presentation of the auditory stimulus. In the post-switch phase, the reward consistently appeared on the opposite side of the display, simultaneously with the presentation of previously unencountered auditory stimuli. Monolingual infants were able to learn the association between stimuli and reward in the pre-switch trials, but failed to learn the new reward location post-switch (i.e., they perseverated in their looks to the previous reward location). The results of both studies suggest that acquiring a first representation may interfere in the acquisition of a second (i.e., a primacy effect). However, when transitions between inputs are adequately cued by contextual changes (such as a change in pitch), learners may be better able to detect transitions between representations and accordingly acquire both.

The challenge of multiple inputs, as well as the role of contextual cues on acquisition, has also been addressed in the domain of statistical learning. In a speech segmentation task, Weiss, Gerfen & Mitchel (2009) presented learners with multiple artificial language streams in conditions both with and without contextual cues. Four language streams, each composed of four distinct trisyllabic words, were tested in isolation (in twelve minutes streams) and found to be learnable. In an experimental condition, two of these streams were interleaved in twelve two-minute segments. When the statistics of both languages were congruent, such that word boundaries were well defined by dips in transitional probabilities, both streams were learnable regardless of whether the streams were differentiated by a contextual cue (i.e., a change in speaker voice). However, when the statistics of both languages were incongruent, such that combining their statistics led to increased statistical noise, the languages were only learnable when differentiated by a contextual cue. In a follow-up study, Mitchel and Weiss (2010) extended the finding of the use of contextual cues to acquire multiple inputs in the visual domain. As in the previous study, learners were familiarized to the two incongruent languages, interleaved in multiple two-minute segments. Learners simultaneously watched a visual display.

When the display showed a different talking face synchronized to each familiarization stream, both streams were learnable despite being presented in the same voice. This same effect was not found if the faces were static, nor if, in the place of faces, a different background color cue was used for each language.

In a related study, Gebhart, Aslin and Newport (2009) also presented learners with two artificial language streams. The languages overlapped by 50% in syllabic inventory, and each language stream was learnable in isolation (after five minutes of presentation). In several experimental conditions, these language streams were presented sequentially for five minutes each, with a single point of transition between them. When both streams were presented in the same voice (i.e., with no contextual cue), learners exhibited a primacy effect, acquiring only the first of the two streams. When the pitch of the voice presenting each stream changed, the authors again found a primacy effect, although learning of the second stream was marginally significant. When the two streams were separated by a thirty-second pause, and participants were explicitly instructed that they would be hearing two languages, both streams were learned at above chance levels. Finally, when there were no contextual cues aside from the 50% overlap in syllabic inventory, but the duration of the second stream was tripled relative to the first, both streams were again learnable.

The results of these studies offer a number of conclusions regarding the acquisition of multiple inputs in a SL paradigm, as well as the influence of contextual cues on learning. A common theme across all of these studies is that, in the absence of adequate contextual cues, statistical variation that approximates input from a bilingual environment is difficult for learners to accommodate. Specifically, learners either show interference or primacy effects. However, contextual cues (such as a change in pitch, speaker voice, or visual scene) that are associated with changes in structure may facilitate the acquisition of multiple inputs. Specifically, learners may be able to use contextual cues to segregate the statistics of non-overlapping inputs and thereby minimize interference between them, thus demonstrating that statistical learning may be a contextually sensitive mechanism. Despite this, not all cues seem to be useful to learners; that is, learners demonstrate some selectivity regarding the salience of contextual cues. Furthermore, a small number of studies on the acquisition of multiple inputs suggest processing differences between bilinguals and monolinguals, especially in environments without contextual cues (Byers-Heinlein and Werker, 2011; Bartolotti et al, 2012). Thus frequent exposure to multiple inputs may adjust a learner's prior expectation for a change in the structure of the input, such that a bilingual requires less contextual support to determine the number of distinct inputs present in their learning environment.

1.3 Word Learning

The studies reviewed thus far have focused exclusively on language acquisition in the context of speech segmentation, or finding words in language input. The next challenge in acquisition is mapping those words to their referents. One aspect of the mapping challenge is indeterminacy, or the idea that a novel word can potentially map to any object, feature, or event in an environment, and can even map to something entirely outside of the learning environment. A well-known example of indeterminacy involves an explorer who encounters a group of people whose language is unknown to the outside world (Quine, 1960). When one of these people points at a rabbit and says, "Gavagai", that person may simply be referring to the rabbit itself with its conventional label. However, "gavagai" might just as well refer to the rabbit's ears, the color of its fur, or the action it is performing (e.g., running, hopping, eating). An additional challenge in mapping involves multiple mappings, or the fact that both within and across languages, a single referent may have several names (i.e., within-language: sofa and couch; across-language: perro and dog). To successfully map words to objects, learners must therefore discover which of the infinite number of possible mappings for a word is correct, and accommodate the variability in mapping caused by the existence of many-to-one mappings both within and across languages.

There are a number of possible strategies for overcoming the challenge posed by indeterminacy. Since the number of potential mappings for a given word is nearly infinite, previous research has proposed the existence of heuristics that constrain the number of possible mappings. Among these are a preference for mapping words to whole objects rather than features or subparts of those objects (i.e., a whole-object preference) (Golinkoff et al., 1992), the use of morphology to infer word class (such as noun, adjective, verb, etc.) (Chen et al., 2009), as well as the application of syntactic knowledge, such as how referring and non-referring words are typically distributed (i.e., knowledge of where subjects, objects and verbs, as well as determiners, typically occur within an utterance) (Monaghan and Mattock, 2012). To accommodate the variability associated with multiple-mappings, a mutual exclusivity preference has been proposed to influence word-mapping. Specifically, learners may preferentially assume that words and objects enter into 1:1 relationships. This assumption, while limiting the number of potential mappings of an unknown word to referents in the world, may also confound learning in the case of multiple mappings. Specifically, learners in a bilingual environment who rely on mutual exclusivity to constrain the mapping space may do so at the expense of learning translation equivalents. However, there is some debate regarding the development of the mutual exclusivity preference in bilinguals, with a number of recent studies suggesting that bilinguals do not rely on this constraint to the same extent as monolinguals (Byers-Heinlein and Werker, 2009; Houston-Price, 2010).

Carey and Bartlett (1978) showed that learners can use their pre-existing knowledge of word-object mappings to overcome indeterminacy. In their "fast-mapping" study, infant learners were exposed to a simple scene containing one object with a known name and one novel object with an unknown name (e.g., a cup and a vise). Participants were then asked to point to one of the objects, by hearing the name of the object inserted into a carrier phrase such as "Point to the ____". When the name of the unknown object was inserted into the carrier phrase, infants reliably pointed at the unknown object and retained the novel mapping after only a single exposure. Importantly, this result suggests that learners may apply a mutual exclusivity assumption in solving the mapping problem. However, the fast-mapping paradigm omits several critical aspects of the challenge of mapping. Specifically, learners in the real world often encounter environments filled with many objects, many or all of which may have unknown names. Additionally, learners encounter objects multiple times, and may rarely or never be exposed to the same object in similar environments (e.g., a child may encounter hotdogs at home in the kitchen, at a barbecue in the park, or in a commercial on TV, in which case the set of non-hotdog objects in each scene is always changing). In the case that a learner encounters many unknown objects and the composition of the environment in which these objects appear is highly variable, the learner cannot rely on a fast-mapping.

Recent research suggests that a statistical learning mechanism may facilitate word learning in highly variable environments (Yu and Smith, 2007; Voulomanos et al., 2008). The cross-situational statistical learning (CSSL) paradigm assumes that a learner is sensitive to the co-occurrence of words and objects, and furthermore, can track co-occurrence statistics across many learning environments. Thus, despite the high variability of any single environment, statistical word learners can infer that words and objects that consistently co-occur across many environments form a pair. In the basic CSSL paradigm, learners are presented with a series of trials in which they view an array of 2-4 visual objects and hear each of the objects labeled. The visual array of objects and the auditory sequence of words are randomly ordered, making each trial locally ambiguous with respect to which words and objects are paired. However, there is perfect consistency in which words and objects appear together across trials, and each word-object pair appears in multiple trials. Both children and adults can successfully map words to objects in the CSSL paradigm, suggesting that human learners are capable of tracking the co-occurrence statistics of words and objects across many learning environments (Yu et al. 2009; Yurovsky et al. 2009).

To date, a majority of CSSL studies have presented learners exclusively with 1:1 word-object mappings, which may approximate input from a monolingual environment (not considering synonyms). Many learners also encounter bilingual input, which requires learners to remap previously learned objects (i.e., to acquire 2:1 mappings). For instance, a learner of English can map the word "dog" to the family pet; if that learner begins acquiring French, the French word "chien" should likewise be mapped to the family pet. However, learners typically demonstrate a bias towards forming 1:1 mappings, potentially because an assumption of mutual exclusivity considerably reduces the size of the mapping space for novel words and objects. Work by Ichinco, Frank, and Saxe (2008) has examined the mutual exclusivity preference in a cross-situational word learning paradigm. In their experiments, learners acquired a set of 1:1 word-object mappings in an initial learning phase. In a subsequent learning phase, a subset of items from the first learning phase were remapped to new items, thus providing evidence for 2:1

mappings to learners. The results of this experiment showed that learners failed to acquire the 2:1 mappings, suggesting that a mutual exclusivity preference is operant in cross-situational learning contexts. In particular, this finding poses a potential problem for bilinguals who frequently encounter 2:1 mappings; specifically, if learners follow mutual exclusivity in word learning, how do bilinguals overcome this preference (recent evidence suggests that learners can form 2:1 mappings in the absence of contextual cues in some paradigms – see Yurovsky and Yu (2008), discussed below, for a review)?

The speech segmentation literature reviewed above suggests that learners may exploit contextual cues in order to form multiple mappings. That is, changes in speaker voice or pauses between streams facilitated acquisition of both streams compared to conditions containing no such cues. Previous work in the word learning literature has also examined how changes in speaker voice or other acoustic properties of the input (such as affect or phonetic variation) influence learning. For example, several studies with infant learners have shown that high variability acoustic training improves the learning of lexical neighbors; this facilitation effect has been attributed specifically to variation in the identity of the speaker (i.e., the uniqueness of a voice) (Rost, McMurray, 2009; Rost, McMurray, 2010). Lively et al. (1993) employed a high variability training paradigm to train Japanese-English bilinguals on the /r/-/l/ phonemic contrast, which exists in English but not Japanese. Learning of the phonemic contrast was significantly greater when the training input originated from multiple native speakers of English, compared to a single speaker, a result that has been replicated a number of times (Lively et al., 1994; Bradlow et al., 1997). Similarly, Bradlow (2004) showed that comprehension of the speech of a novel L2 speaker is improved when listeners are trained on multiple rather than a single L2 speaker with the same accent. In the above studies, the advantages conferred by multi-speaker training input are thought to arise as a result of learners expanding category boundaries by storing speaker-specific variation rather than discarding it. That is, learners exploit variability in the signal in order to optimize processing, using context to create representations of structure that can accommodate both previously encountered and novel inputs. A recent study confirms this conclusion: Trude and Brown-Schmidt (2011) examined whether learners can use speaker-specific knowledge to determine the identity of words presented in isolation. Learners were familiarized to speakers with familiar or unfamiliar regional dialects, such that the intended target of a vowel sound depended on the dialect of the speaker producing the sound. The authors found that learners were able to exploit speaker-specific knowledge to identify words containing otherwise ambiguous vowels. Crucially, providing learners with a contextual cue to speaker identity prior to test strengthened this effect, demonstrating that learners are able to exploit context in order to optimize processing.

Collectively, these studies establish that learners are sensitive to changes in contextual cues in word learning paradigms, and further, that contextual cues may facilitate the acquisition of lexical or phonemic contrasts in a monolingual setting. The present work aims to investigate the role of contextual cues on word learning in bilingual environments. Ichinco, Frank and Saxe (2009), in a paradigm without changes in contextual cues, showed that word learners follow mutual exclusivity, forming 1:1 mappings even when evidence for 2:1 mappings is available. The question following from work in both the speech segmentation literature, as well as the previous work in word learning, is whether the addition of contextual cues to the learning environment facilitates the formation of 2:1 word mappings. One possible mechanism for this facilitation effect is the role of context in marking points of transitions between inputs; that is, learners may be more apt to accommodate systematic variation in their input given a contextual cue that signals the emergence of a new distribution.

An additional question of the present research concerns the time course of learning in SL tasks. Specifically, the studies reported to this point have assessed learning at a single point immediately after training. As a result, these studies are unable to provide data regarding the time course of statistical learning. A number of word-learning studies have investigated the time course of learning. Medina et al. (2010) presented learners with 1:1 word-object mappings in a cross-situational paradigm. After each trial, learners made a guess about the word-object mapping and also rated their confidence (on a scale from 1 to 5) in that guess. The authors reported that confidence ratings rose across a learning phase as learners gained more experience with each mapping. In another cross-situational paradigm containing both one-to-one and many-to-one mappings, Yurovsky and Yu (2008) found that the trajectory of confidence ratings was influenced by the types of mappings present in the learning phase., Although confidence ratings rose across the learning phase of this study, they were significantly lower when both one-to-one and many-to-one mappings were available in the same learning phase. The results of both studies suggest a positive correlation between experience with mappings and confidence in those mappings. At the same time, these results seem to indicate that the presence of many-to-one mappings increases mapping ambiguity for one-to-one and many-to-one mappings alike. However, to the best of our knowledge, no studies have investigated the influence of contextual cues on the time course of learning. Collecting data on the time course of learning in conditions both with and without contextual cues may provide important data about how a learner's prior expectation for a change in the input influences learning. Specifically, these data can further an understanding of how contextual cues modulate ambiguity when both one-to-one and many-to-one mappings are available, as well as the point in learning at which contextual cues begin to facilitate the acquisition of many-to-one mappings.

1.4 Proposed Experiments

The present research addresses the formation of many-to-one word-object mappings using the CSSL paradigm, and thereby examines whether contextual sensitivity is a crucial component of the statistical learning mechanism across several domains of linguistic structure. In two series of experiments, we investigated how contextual cues influence statistical word learning in environments containing multiple mappings, paralleling studies of the acquisition of multiple inputs in the speech segmentation literature. We collected both end-state measures of learning (post-training data), as well as more incremental measures of learning from several discrete points across the training phase. These measures thus provide data regarding the absolute state of learning after training, in addition to the rate of learning. Respectively, these measures address the questions of whether contextual cues facilitate remapping in the CSSL paradigm and whether contextual cues exert a detectable influence on the rate of remapping across a training phase. In each series of experiments, a baseline measure of performance is collected in conditions presenting multiple inputs without explicit contextual cues. Subsequently, the experimental conditions add contextual cues to the input. Several contextual cues are tested in each series of experiments. In each case, comparison of the baseline and experimental conditions provide results relevant to the influence of contextual cues in the acquisition of many-to-one word-object mappings.

In the first series of experiments, we adapt the CSSL paradigm to determine whether the addition of contextual cues to the training input might aid learners in the formation of many-to-one mappings. Arguably, this is a first step toward determining the factors that may allow learners to overcome the mutual exclusivity constraint when confronted with multiple statistically distinct inputs. Experiment 1a employs a methodology similar to that of the initial CSSL study performed by Yu and Smith (2007) in order to ensure that the paradigm is viable. In Experiment 1b, we borrowed many of the methods used in Ichinco, Frank, and Saxe (2009) to determine whether our participants demonstrate a preference for mutually exclusive mappings when the CSSL paradigm presents co-occurrence evidence that could support the formation of many-to-one mappings. Specifically, in Experiment 1b we present learners an initial set of word-object mappings and subsequently remap a subset of the learned words to new objects. As no contextual cues are present in Experiment 1b, we predict that a mutual exclusivity preference may prevent learners from acquiring the many-to-one mappings. This condition serves as the baseline for comparisons with conditions adding contextual cues. In Experiment 1c-e, we test how the addition of contextual cues to the same training input from Experiment 1b influences the preference for mutually exclusive mappings. As noted, findings from the speech segmentation literature have shown that learners can better accommodate statistical variability when given a salient contextual cue; that is, learners can make contextually sensitive hypotheses about the structures in their input. Thus, if learners are also able to track statistics in a contextually manner in a statistical word learning task, we predict that the addition of contextual cues to the training input will facilitate the formation of many-to-one mappings compared to conditions containing no contextual cues. In the case that the statistical learning mechanism is insensitive to context in the domain of word mapping, we predict the addition of contextual cues to the training input to have no effect on the formation of many-to-one mappings.

In the second series of experiments, we further adapt the CSSL paradigm to examine how contextual cues influence the rate of learning in environments that support the formation of many-to-one mappings. The findings of these experiments may further inform an understanding of how contextual cues modulate the speed with which a learner postulates a change in the underlying statistics of their input, specifically by demonstrating at what point in the training phase the effects of contextual cues appear. Using the methodology of Yurovsky and Yu (2008), we familiarized learners to word-object pairs in multiple learning phases and collected confidence ratings for each word-object pair immediately after every trial presented during training. The confidence rating data thus provided a more online measure of how knowledge of word-object mappings changed across training, while the post-training tests still provided a measure of end-state knowledge of word-object mappings.

Experiment 2a served as a baseline condition, in which we collected data on the rate of learning both 1:1 and many-to-one mappings in an environment containing no contextual cues. Experiments 2b-c tested the influence of a subset of the contextual cues from Experiment 1 on the rate of learning both one to one and many-to-one mappings. In the case that contextual cues quickly signal the emergence of a new input distribution, we predict that the addition of contextual cues to Experiments 2b-c will have an early and positive influence on learner's confidence in many-to-one mappings. Additionally, we predict the rate of increase in confidence ratings for many-to-one mappings across a familiarization to be significantly greater in conditions with contextual cues compared to baseline.

2 Experiment 1

In Experiment 1, we attempt to replicate the findings from Yu and Smith (2007) using an original set of auditory and visual stimuli. The general methods conformed to the original experiments, with the exception of the constraints for repetition of stimuli within familiarization trials. In our experiment, unlike the original study, no object could be viewed in successive frames during familiarization. Thus, if learners relied on this type of local information, we predicted there would be a decrement in learning.

2.1 Experiment 1 Subjects

Twenty Introductory Psychology students (14 female, 6 male; 18-23 years old) at Penn State University were given course credit for their participation in this experiment. None had prior experience in a statistical learning experiment. Two additional students participated in this experiment but were excluded from the analysis as outliers based on their OSpan performance (see below).

2.2 Experiment 1 Stimuli and Procedure

The stimuli for Experiment 1 consisted of fifty-four unique word-object pairs created by randomly pairing novel objects with nonce words. The objects consisted of black and white complex line drawings. Eight objects appeared in the stimuli used by Creel, Aslin, Tanenhaus (2011) and served as a template for creating the remaining 46 objects (using MS Paint ©). All objects were converted to a .jpeg file format with a size of 150x150 pixels. Nonce words bore English phonological patterns and consisted of an equal distribution of monosyllabic, disyllabic, and trisyllabic items chosen from the English Lexicon Project non-word database (http://elexicon.wustl.edu). The words were rendered in a female American English voice (*Crystal*) using the AT&T Natural Voices text-to-speech synthesizer (http://www.naturalvoices.att.com), and converted into WAV files sampled at 22050 Hz using Praat (Boersma, 2001). The fifty-four word-object pairs were split into three sets of 18 for presentation in the three experimental conditions (with word length distributed equally across conditions). No word-object pair appeared in more than one condition.

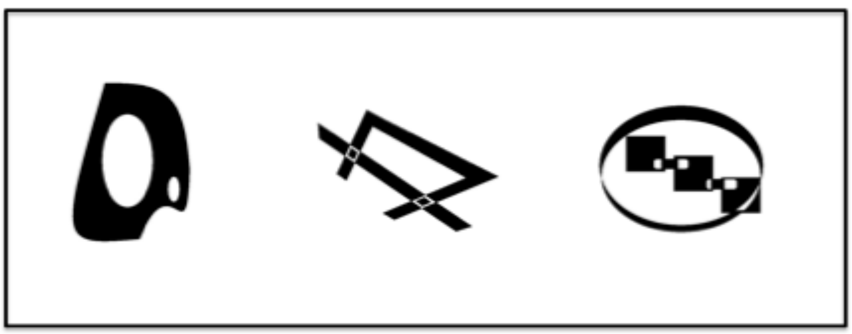

Figure 2.1. Three examples of the novel objects created for Experiment 1.

All experiments were conducted in a sound-attenuated chamber and were programmed using E-Prime 2.0. Participants in Experiment 1 completed a cross-situational learning task followed by an Operation Span (OSpan) task, which provided a measure of working memory capacity. The cross-situational task consisted of three conditions, each of which was differentiated by the number of word-object pairs presented in familiarization trials. Accordingly, participants completed a 2x2 (two word-object pairs), a 3x3, and a 4x4 condition. Each of these conditions was composed of a familiarization phase followed by a test phase. The order of presentation of these conditions was counterbalanced across participants.

During familiarization, participants observed objects appearing on a computer screen while listening to words presented over speakers. Preceding each trial, a fixation cross appeared for 750ms. Within the trial, all objects appeared simultaneously while the corresponding nonce words were played serially at 3s intervals. The onset of the visual presentation of objects was synchronized with the presentation of the first word of the trial.

There was no systematic relationship within a trial between the placement of an object in the visual array and the location of its corresponding word in the auditory stream; object locations and word orders were randomly assigned. Progression through the trials was automatic; the end of one trial cued the presentation of the next. The order of the trials was pseudo-randomized so that no word-object pair appeared in two consecutive trials. Within each condition, every word-object pair was presented 6 times. Accordingly, the total number of trials varied by condition: there were fifty-four trials in the 2x2, 36 in the 3x3, and 27 in the 4x4. Total time of familiarization within a condition remained constant at 320 seconds.

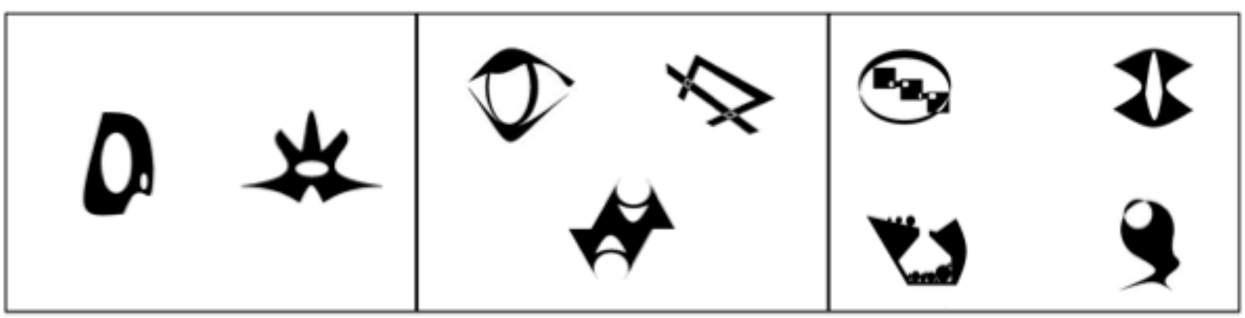

Figure 2.2. From left to right, examples of training trials presented in the 2x2, 3x3 and 4x4 conditions of Experiment 1.

Following each familiarization phase, participants completed a 4 alternative forced-choice (4AFC) test. On each test trial, participants viewed four objects (in randomized positions) while hearing a single word. Three of these objects were distractors randomly selected from the set of objects presented within a given condition. The remaining object was the correct referent for the presented word. All objects within a test trial were presented simultaneously, with one object located in each corner of the screen. Additionally, each object was labeled with a number (1-4), located in the corners of the screen. Participants were asked to press the number key corresponding to the correct object. No time limit was imposed. Test sessions consisted of 18 trials, one for every word-object pair presented in the preceding familiarization phase.

Before the experiment began, participants were informed that they would see many new objects and hear many new words, and that their task was to figure out which words and objects went together. No further information was given regarding the nature of the mappings available in the familiarization phases, and participants were not informed that each word would map to only one object and vice versa.

Following the cross-situational statistical learning task, participants completed an Operation Span (OSpan) task as a measure of working memory capacity. The task consisted of multiple trials in which participants were given 3.5 seconds to indicate whether an arithmetic equation (e.g., (10/2) + 3 = 6) was true or false via button press, and subsequently 2 seconds to memorize a single English word (see Christoffels, De Groot, & Kroll 2006). Following a set of such trials (i.e., between 2 and 6), participants were asked to recall the words from that set by typing them into a response box. Participant's scores on the OSpan task were utilized in two ways. First, they provided a covariate of performance for the main experimental task. Secondly, they were used as a filter: any score more than 2.5 standard deviations below the mean score on the task was established as a threshold for acceptable effort on the experimental tasks. Consequently, the data of participants scoring below this threshold were removed from all subsequent analyses.

2.3 Experiment 1 Results

All statistical analyses reported in Experiments 1-3 use a two-tailed distribution. In Experiment 1, participants learned, on average, 89% of word-object pairs in the 2x2 condition; 69% of pairs in the 3x3 condition, and 53% of pairs in the 4x4 condition (See Figure 1). Performance in each of the three conditions was significantly above chance (25%; one-sample *t*-test: 2x2: $t(20)=18.0$, $p<.001$; 3x3: $t(20)=7.8$, $p<.001$; 4x4: $t(20)=6.5$, $p<.001$). An ANOVA revealed significant differences in test performance between the three experimental conditions ($F(2,57)=15.7$, $p<.001$), and a linear contrast analysis indicated that test performance declined significantly as the number of word-object pairs within a scene increased from 2 to 4 ($F(1,57)=31.2, p<.001$).

There was a significant positive correlation ($R = .541$, $p < .05$) between OSpan score and test performance in the 2x2 condition ($R = .66$, $p < .01$) and the 3x3 condition ($R = .47$, $p < .05$). This correlation did not reach significance in the 4x4 condition ($R = .27$, $p = .26$).

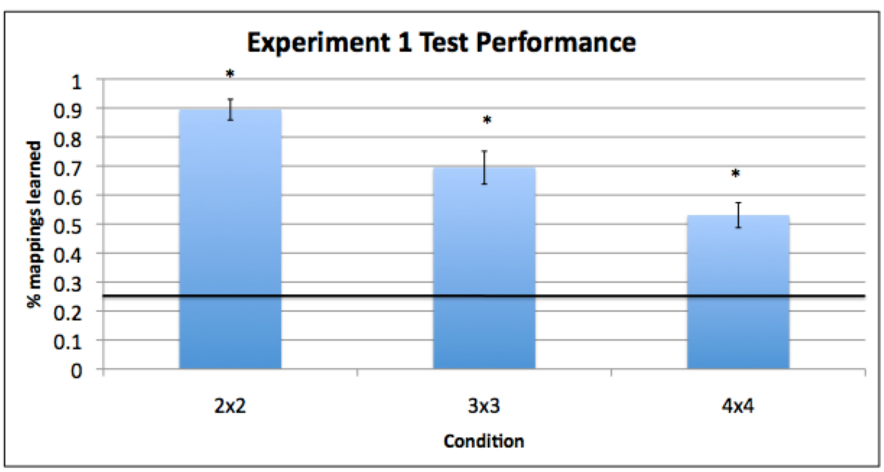

Figure 2.3. Mean accuracy at test in the three conditions of Experiment 1. Chance performance is set at 25%.

2.4 Experiment 1 Discussion

The results of Experiment 1 are consistent with those reported by Yu and Smith (2007). Despite the local ambiguity of word-object pairings within any given trial, learners were able to correctly identify word-object pairings across the course of a session, performing well above the level of chance even in the condition of highest ambiguity (4x4). The successful replication was evidenced despite the stricter sequencing constraints for familiarization trials (i.e., no word-object pair appeared in two consecutive trials). This type of local information was not controlled in the original experiment (Yu & Smith, 2007) and has subsequently been posited to exert a significant influence on the ability to perform in cross-situational statistical learning experiments (Onnis, Edelman, & Waterfall, 2010). Here we have demonstrated that learning can proceed relatively unhindered (comparing our results to those collected by Yu & Smith's original study) even in the absence of such information.

As in the original study, performance on the task was negatively correlated with the number of word-object pairs presented within a trial. This result suggests that an increase in the visual and auditory complexity of a trial made the co-occurrence statistics of words and objects more difficult to track across trials. Interestingly, the correlation between OSpan scores and performance on the CSSL task failed to reach statistical significance as within-trial complexity increased to its maximum in the 4x4 condition. As the correlation between performance and Ospan scores held for both the 2x2 and 3x3 condition, this lack of a relationship in the 4x4 condition may be attributable to a power issue. More testing is needed to determine the nature of the relationship between within-scene complexity and statistical learning.

3 Experiment 2

Having replicated the original findings of Yu and Smith (2007) with new stimuli and stricter sequencing constraints, Experiment 2 was designed to investigate mutual exclusivity in the context of statistical word-learning. Specifically, our goal was to provide additional evidence that learners prefer to acquire one-to-one word-object mappings, even when two-to-one mappings may be present in the learning environment. In Experiment 2, participants received two familiarizations, in which a set of six words that were successfully mapped to objects in the first familiarization was remapped to new objects in the second familiarization. Thus participants in Experiment 2 were exposed to both one-to-one and two-to-one mappings in subsequent familiarizations.

3.1 Subjects

Twenty-eight Introductory Psychology students at Penn State University who had not participated in Experiment 1 nor any other statistical learning experiment were given course credit for their participation. One participant was removed from subsequent analyses based on below threshold OSpan performance. Five additional participants failed to reach the criterion score on the test following the first familiarization (see below) and were excused from the experiment. Twenty-two participants (16 female and 6 male; 18-29 years old) were included in the final analysis.

3.2 Stimuli and Procedure

Experiment 2 consisted of two familiarization phases, each of which was followed by a test. The first familiarization and test were identical to those presented in the 3x3 condition of Experiment 1 (described above). In order to proceed to the second familiarization phase, participants were required to produce a minimum of 10 correct responses (out of 18 total) on the first test. Failure to achieve this criterion ended the experiment prematurely. Participants who scored 10 or more correct on the first test continued to a second familiarization phase with a new set of eighteen word-object pairs followed by another test phase. The second familiarization phase consisted of a unique set of 18 word-object pairs. Additionally, six learned words from the first familiarization (i.e., words that were correctly mapped to their referents in the test following the first familiarization) were transferred to the second familiarization and paired with new objects. Consequently, all trials in the second familiarization of Experiment 2 consisted of three visual objects paired with 4 auditory labels. The transferred words could be mapped to two possible objects (a first familiarization object or a second familiarization object). Similarly, six objects from the second familiarization could be mapped to both a novel word from the second familiarization and the transferred first familiarization word. All other characteristics of the second familiarization phase were identical to the first familiarization.

Following the second familiarization phase of Experiment 2, there was a test phase consisting of 54 trials. An initial set of 18 4AFC trials tested the pairings between unique second familiarization word-object pairs. Next, six 4AFC trials tested the many-to-one mappings resulting from the set of words that appeared in both the first and second familiarization phases. In these trials, a word transferred from the first familiarization was presented along with four objects from the second familiarization; one was its referent from the second familiarization and the other three were distractors. Following these test items, participants completed two sets of six preference test trials (hereafter *object preference* and *word preference* trials). Object preference trials presented a transferred word auditorily and a visual array containing that object's first and second familiarization object mappings, along with two distractor objects. Word preference trials presented a second familiarization object along with both its second familiarization and transferred first familiarization word mappings. A final set of 18 4AFC items retested the word-object mappings learned in the first familiarization phase.

3.3 Experiment 2 Results

Performance on the test following the first familiarization phase was significantly above the level of chance (77% correct, ($t(22)=23.7$, $p<.001$). Performance on the test following the second familiarization phase is reported by test trial type. Participants performed at above chance levels on trials testing 1:1 mappings between words and objects unique to the second familiarization (64% correct, ($t(22)=12.8$, $p<.001$). Performance on trials testing the learning of many-to-one mappings available in the second familiarization was not significantly above the level of chance (2:1 Mappings: 23% correct; one-sample *t*-test: $t(22)=-.51$, $p = .62$). Participants performed above the level of chance when asked to recall the set of 1:1 mappings presented in the first familiarization (71% correct, ($t(22)=16.4$, $p<.001$).

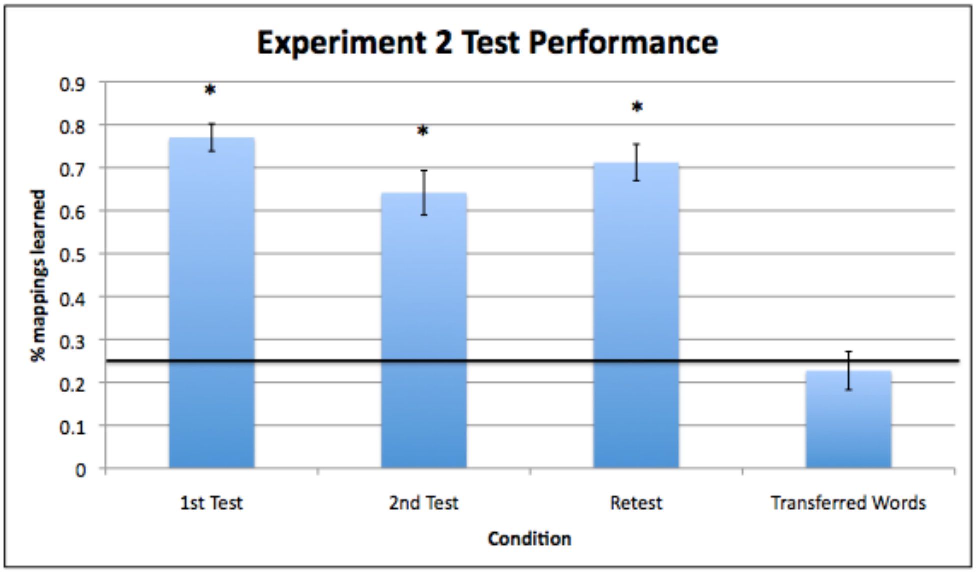

Figure 3.1. Mean accuracy at test on all trial types in Experiment 2. Chance performance is set at 25%.

The average OSpan score for participants in Experiment 2 was 44.1 out of 60, with a standard deviation of 7.5. Participants' OSpan scores did not correlate significantly with performance on the test following the first familiarization ($R = .07$, $p=.77$), the test of 1:1 mappings between words and objects unique to the second familiarization ($R = .28, p=.21$), the test of many-to-one mappings ($R = -.3$, $p=.16$) or the retest of 1:1 mappings from the first familiarization ($R = .3$, $p=.17$).

Object preference trial types presented a transferred word and four objects. One of these four objects was the first familiarization object (primacy) mapping to the transferred word; another was the second familiarization object (recency) mapping to the transferred word; the remaining two objects were distractors. Accordingly, chance selection of both primacy and recency mappings was set at 25%, while chance was set at 50% for distractor objects. In Experiment 2, participants selected the primacy mapping on 67% of trials and the recency mapping on 11% of trials. Participants incorrectly selected a distractor object on 22% of trials.

The rate of selection of the primacy mapping significantly exceeded the level of chance ($t(22)=6.6, p<.001$). The rate of selection of recency mappings was significantly below the level of chance ($t(22)=-3.4, p<.01$), as was the rate of selection of distractor objects (($t(22)=-7, p<.001$). A direct comparison of the rate of selection of the primacy and recency mapping revealed a significant difference ($t(22)=-5.6, p<.001$) which favored selection of the primacy mapping

On word preference trial types, participants saw a single object on the screen and heard two words played sequentially. The presented object was one of six second familiarization objects to which both a second familiarization word and a transferred first familiarization word were mapped. Accordingly, these two words were played on each trial. On word preference trials in Experiment 2, participants selected the transferred first familiarization word mapping in 24% of trials, and selected the second familiarization word mapping in the remaining 76% of trials. This pattern of responses differed significantly from chance ($t(22)=6.3, p<.001$) in the direction of a bias for second familiarization word mappings.

3.4 Discussion

In Experiment 2, a set of words previously mapped to objects in the first familiarization was remapped to a new set of objects in the second familiarization. While learners were able to acquire the one-to-one mappings from both the first and second familiarizations at above chance levels, they did not successfully remap the transferred words to acquire many-to-one mappings. These results are consistent with the findings reported in Ichinco, Frank, & Saxe (2009) and indicate that learners formed mutually exclusive word-object mappings. Specifically, regarding the many-to-one mappings, learners exhibited a strong preference for the first mapping of a word to an object. Subsequent mappings of previously learned words to new objects were not learned. In Experiment 3, we explore whether adding contextual cues that differentiate the first and second familiarization phases attenuates the mutual exclusivity preference and facilitates the formation of many-to-one word-object mappings. In separate conditions, we tested three contextual cues: 1) a voice cue (similar to that used in Weiss, Gerfen, & Mitchel, 2009); 2) a voice cue combined with an accent cue; and 3) and an explicit cue (similar to that used in Gebhart, Aslin, & Newport, 2009).

4 Experiment 3

4.1 Experiment 3 Subjects

Eighty-four Introductory Psychology students (46 female and 20 male; 18-25 years) participated in Experiment 3 for course credit. None had participated in Experiments 1 or 2 nor any other statistical learning experiment. Across all conditions of Experiment 3, four participants were removed from the final analysis based on their OSpan performance. Fourteen additional participants failed to reach the criterion score in the test following the first familiarization phase and were excused from the experiment. Sixty-six participants (46 female and 20 male; 18-25 years) were included in the final analysis.

4.2 Experiment 3 Stimuli and Procedure

Experiment 3 was very similar to Experiment 2 in terms of stimuli and procedure. The primary difference was the addition of a contextual cue to the second familiarization phase of Experiment 3. Three different contextual cues were used in three separate conditions. In all three of these conditions, words in the first familiarization were presented in the female American English accented voice used in the previous experiments (see above). In Experiment 3a (Gender Cue), words in the second familiarization were produced in Voice 2, a novel male American English voice whose fundamental frequency was, on average, 70 Hz lower than that of Voice 1. In Experiment 3b (Voice and Accent Cue), words in the second familiarization were produced in Voice 3, a male French accented voice (*Alain*). Specific changes from Voice 1 to Voice 3 included a drop in the fundamental frequency of approximately 30 Hz as well as changes in vowel quality (e.g., / ɪ / -> /i/) and consonant identity (e.g., /tS/ -> /S/, /θ/ -> /t/) consistent with French phonotactics. In Experiment 3c (Explicit Cue), there was no voice change, but rather a set of instructions intervening between the first and second familiarizations which informed participants that, in the second familiarization, they would be able to learn new object mappings for a set of previously learned words. Participants were randomly assigned to one of the three experimental conditions.

In the second test of Experiment 3a and 3b, the voice used to present test items varied according to test trial type. All Retest auditory stimuli were presented in Voice 1; all other auditory test stimuli were presented in the voice used in the second familiarization (Voice 2 for Experiment 3a and Voice 3 for Experiment 3b). In all other respects, the tests of Experiment 3 were identical to those of Experiment 2. The second test of Experiment 3c, as there was no change in speaker voice in this condition, was identical in all ways to the test following the second familiarization of Experiment 2.

4.3 Experiment 3a (Gender Cue) Results

A series of one-sample t-tests found that performance on the test following the first familiarization phase was significantly above the level of chance (80.5% correct, $t(22)=20.8$, $p<.001$), as was performance on trials testing the 1:1 mappings between words and objects unique to the second familiarization (50% correct, $t(22)=4.7$, $p<.001$). In contrast to Experiment 2, performance on the trials testing many-to-one mappings in Experiment 3 was significantly above the level of chance (one-sample *t*-test: 35% correct $t(22)=2.2$, $p<.05$), indicating that the contextual cue of a change in speaker voice from the first to the second familiarization supported the acquisition of many-to-one mappings (see Figure 1). Performance was also above the level of chance on trials retesting the 1:1 mappings formed in the first familiarization phase (Retest: one-sample *t*-test: 69% correct, $t(22)=9.7$, $p<.001$).

The average OSpan score for participants in Experiment 3a was 42.5 out of 60, with a standard deviation of 10.2. There was a significant positive correlation between OSpan scores and performance on the test of 1:1 mappings following the 1st Familiarization phase ($R = .429$, $p < .05$), 1:1 mappings between words and objects unique to the second familiarization ($R = .519$, $p < .05$), and the retested 1:1 mappings from the 1st Familiarization ($R = .704$, $p < .01$). There was no correlation between OSpan scores and performance on trials testing 2:1 mappings ($R = .05$, $p=.83$).

On object preference trial types in Experiment 3a, participants endorsed the primacy mapping in 66% of cases, the recency mapping in 18% of cases, and a distractor object in the remaining 16% of cases. Endorsement of primacy mappings was significantly above the level of chance ($t(22)=7.3$, $p<.001$), and also significantly exceeded the rate of endorsement of recency mappings (*t*(22)=-5.4,p<.001). The rate of endorsement of recency mappings did not significantly exceed chance (*t*(22)=1.8,p=.08) The rate of endorsement of distractor objects was again significantly below the level of chance (*t*(22)=-9.6,p<.001).

On word preference trials in Experiment 3a, participants selected the transferred first familiarization word mapping in 30% of cases, and selected the second familiarization word mapping in the remaining 70% of cases. This pattern of responses differed significantly from chance (*t*(22)=4.5,p<.001) in the direction of a bias for second familiarization word mappings.

4.4 Experiment 3b (Accent Cue) Results

Performance on the test following the first familiarization phase was significantly above the level of chance (one-sample *t*-test: 83% correct, *t*(22)=16.7, *p*<.001), as was performance on trials testing the 1:1 mappings between words and objects unique to the second familiarization (one-sample *t*-test: 54% correct, *t*(22)=6.7, *p*<.001). Performance on the trials testing many-to-one mappings was significantly above the level of chance (2:1 Mappings: one sample *t*-test: 36% correct, *t*(22)=2.3, *p*<.05). Participants also performed significantly above the level of chance on trials retesting 1:1 mappings learned in the first familiarization phase (Retest: one-sample *t*-test: 79% correct, *t*(22)=12.3, *p*<.001).

The average OSpan score for participants in Experiment 3b was 46.4 out of 60, with a standard deviation of 9.4. There was no significant correlation between OSpan scores and performance on the test of 1:1 mappings following the 1st Familiarization phase (R = .369, p = .1), 1:1 mappings between words and objects unique to the second familiarization (R = -.11, p = .63), and 2:1 mappings (R = -.08, p=.74). This correlation did reach significance for the retested

1:1 mappings from the 1st Familiarization ($R = .434$, $p < .05$).

On object preference trials in Experiment 3b, participants endorsed the primacy mapping in 32% of cases, the recency mapping in 30% of cases, and a distractor object in the remaining 39% of cases. Endorsement of primacy mappings was not significantly above the level of chance, and did not significantly exceed the rate of endorsement of recency mappings. The rate of endorsement of distractor objects, in Experiment 3b, was not significantly different from chance, either.

On word preference trials in Experiment 3b, participants selected the transferred first familiarization word mapping in 44% of cases, and selected the second familiarization word mapping in the remaining 56% of cases. This pattern of responses did not differ significantly from chance ($t(22)=1.2, p=.26$), indicating no reliable bias for either transferred first familiarization or second familiarization word mappings to multiply mapped second familiarization objects.

4.5 Experiment 3c (Explicit Cue) Results

A series of one-sample *t*-tests indicated that performance on the test following the first familiarization was significantly above chance (79% correct, $t(22)=17.0$, $p<.001$), as was performance on trials testing the 1:1 mappings between words and objects unique to the second familiarization (59% correct, $t(22)=5.5$, $p<.001$). In contrast to Experiment 2, performance on the trials testing many-to-one mappings was significantly above the level of chance (2:1 Mappings: one-sample *t*-test: 37% correct, $t(22)=2.1$, $p<.05$), indicating that the explicit mention of the availability of many-to-one mappings supported the acquisition of these mappings (see Figure 1).

Trials retesting the 1:1 mappings presented in the first familiarization phase after the second familiarization were also significantly above chance (Retest: 70% correct, $t(22)=10.8$, $p<.001$).

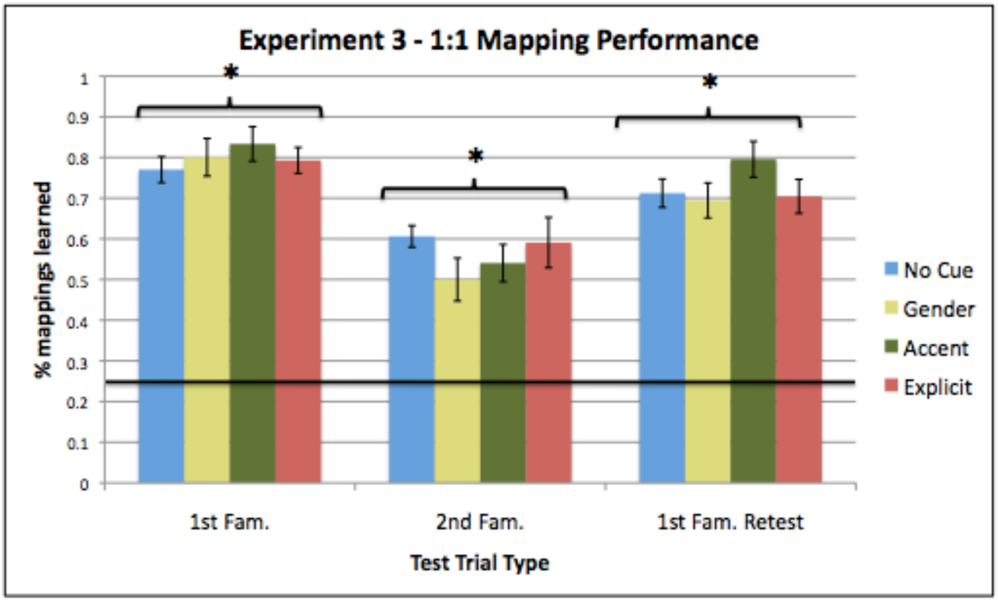

Figure 4.1. Performance on trial types testing one-to-one mappings in all conditions of Experiment 3. There were no significant differences in performance within any trial type. Performance in cases exceeded chance (25%).

The average OSpan score for participants in Experiment 3c was 44.9 out of 60, with a standard deviation of 7.9. There was no correlation between OSpan scores and performance on the test of 1:1 mappings following the 1st Familiarization phase (R = .39, p = .07), 1:1 mappings between words and objects unique to the second familiarization (R = .27, p = .23), 2:1 mappings (R = .36, p=.1), or retested 1:1 mappings from the 1st Familiarization (R = .39, p = .07).

On object preference trial in Experiment 3c, participants endorsed the primacy mapping in 61% of cases, the recency mapping in 21% of cases, and a distractor object in the remaining 18% of cases. Endorsement of primacy mappings was significantly above the level of chance ($t(22)=5.6, p<.001$), and also significantly exceeded the rate of endorsement of recency mappings ($t(22)=-3.7, p<.001$), revealing a preference for primacy mappings to transferred words. The rate of endorsement of distractor objects was again significantly below the level of chance($t(22)=-6.8, p<.001$).

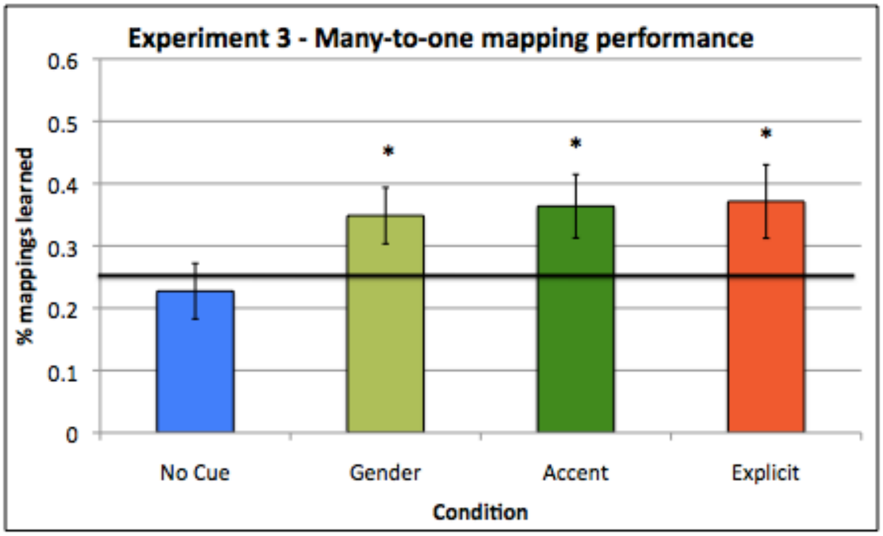

Figure 4.2. The remapping of previously learned words in conditions containing a contextual cue, compared to a condition with no contextual cue, significantly exceeds chance performance (25%).

On word preference trials in Experiment 3c, participants selected the transferred first familiarization word mapping in 31% of cases, and selected the second familiarization word mapping in the remaining 69% of cases. This pattern of responses differed significantly from chance ($t(22)=3.36, p<.01$) in the direction of a bias for second familiarization word mappings.

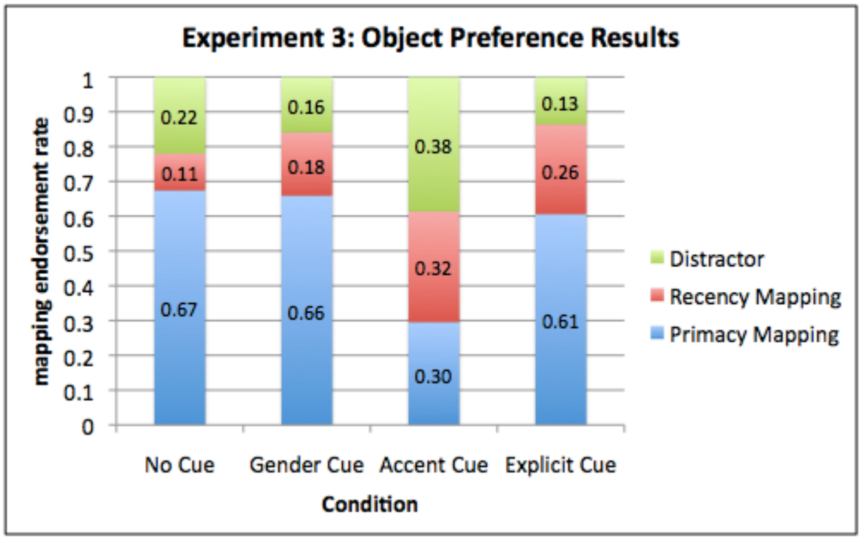

Figure 4.3. When presented with a transferred word, participants overwhelmingly endorsed the first object mapping acquired for that word. This pattern largely holds for the contextual cue conditions, indicating that despite remapping, participants retain a preference for primacy mappings.

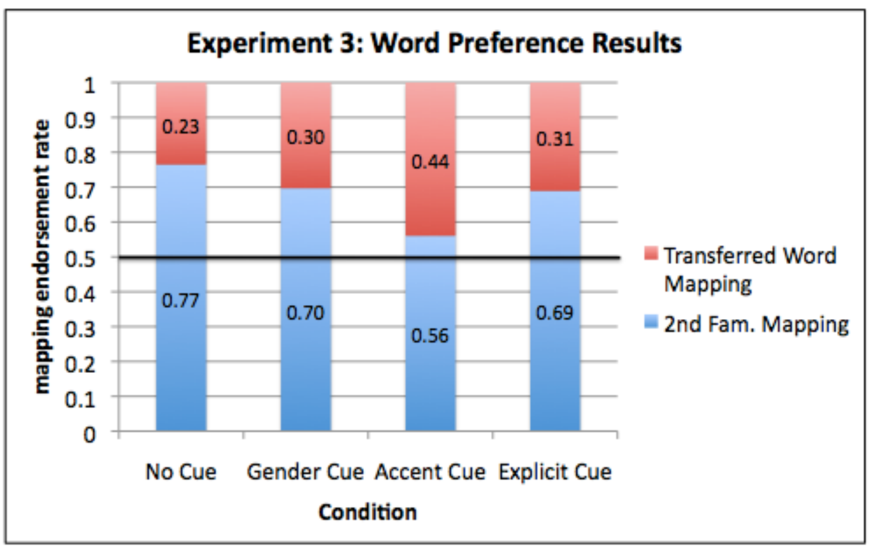

Figure 4.4 When presented with a doubly mapped 2nd Familiarization object, participants overwhelmingly endorsed its 2nd Familiarization word mapping, compared to a transferred 1st Familiarization word that could also be mapped to that object.

4.6 Analysis of Experiments 2 and 3

Across all conditions in Experiments 2 and 3, participants learned 80% of 1:1 mappings presented in the first familiarization, 57% of 1:1 mappings between words and objects unique to the second familiarization, and correctly recalled 73% of the retested 1:1 mappings from the first familiarization. Comparing across all conditions of Experiments 2 and 3, there were no significant differences in performance on trials testing 1:1 mappings following the first familiarization ($F(3,83)=1.2$, $p = .33$) 1:1 mappings between words and objects unique to the second familiarization ($F(3,83)=1.2$, $p=.3$, or trials retesting the 1:1 mappings presented in the first familiarization ($F(3,83)=1.6$, $p= .21$).

In Experiment 2, participants learned 77% of 1:1 mappings presented in the first familiarization, 64% of 1:1 mappings between words and objects unique to the second familiarization, and correctly recalled 71% of the retested 1:1 mappings from the first familiarization. Across all conditions of Experiment 3, participants learned 82% of 1:1 mappings presented in the first familiarization, 55% of 1:1 mappings between words and objects unique to the second familiarization, and correctly recalled 74% of the retested 1:1 mappings from the first familiarization. Comparing performance in Experiment 2 to Experiment 3, there were no significant differences in performance on trials testing 1:1 mappings following the first familiarization ($F(1,85)=1.2$, $p = .28$) 1:1 mappings between words and objects unique to the second familiarization ($F(1,85)=2.2$,$p=.15$), or trials retesting the 1:1 mappings presented in the first familiarization ($F(1,85)=.26$, $p= .61$).

To determine if the presence of contextual cues in Experiment 3 significantly altered the frequency of remapping transferred words, we compared remapping performance in Experiment 2 (which contained no contextual cues) to remapping performance across all conditions of

Experiment 3. In Experiment 2, participants remapped 23% of the transferred words. In Experiment 3 across all conditions, participants remapped 36% of transferred words. This difference was significant ($F(1,87)=5.4$, $p < .05$), indicating that the presence of contextual cues significantly increased the tendency of participants to remap transferred words. A series of post-hoc t-tests compared the rate of remapping in the individual conditions of Experiment 3 against the rate of remapping in Experiment 2. These comparisons all reached marginal significance (Experiment 3a: $t(44)=1.9$, $p=.063$; Experiment 3b: $t(44) = 2.0$, $p = .051$); Experiment 3c: $t(44) = 1.9$, $p = .059$).

Collapsing across all conditions of Experiment 3, there was a significant positive correlation between OSpan scores and performance on the test of 1:1 mappings following the 1st Familiarization phase ($R = .36$, $p < .01$). This correlation disappeared for 1:1 mappings between words and objects unique to the second familiarization ($R = .22$, $p = .08$) and 2:1 mappings ($R = .11$, $p=.38$). The correlation returned to significance for retested 1:1 mappings from the 1st Familiarization ($R = .48$, $p < .001$).

On object preference trials across Experiments 2 and 3, participants selected the recency mapping on 20% of trials, the primacy mapping on 57% of trials, and a distractor object on the remaining 23% of trials. The bias for selection of primacy mappings was significant ($t(88)=6.9, p<.001$), and selection of each object type was significantly different from chance (recency: $t(88)=6.4, p<.001$); primacy: $t(88)=16.3, p<.001$); distractor: $t(88)=6.7, p<.001$). To evaluate the effect of contextual cues on responses to object preference trials, we compared the rates of selection of primacy, recency, and distractor object mappings in Experiment 2 to those in Experiment 3. In Experiment 2, which contained no contextual cues, participants selected primacy object mappings on 67% of trials, recency mappings on 10% of trials, and distractor objects on 22% of trials. Across all conditions of Experiment 3, participants selected primacy mappings on 54% of trials, recency mappings on 23% of trials, and distractor objects on 23% of trials. An ANOVA revealed a significant increase from Experiment 2 to Experiment 3 in the rate of selection of recency mappings ($F(1,87)=5.1, p<.05$), and a marginally significant decrease in the rate of selection of primacy mappings ($F(1,87)=3.5$, $p=.06$). There was no significant change in the rate of selection of distractor objects ($F(1,87)=.24, p=.62$). These results suggest that the addition of contextual cues in Experiment 3 weakened participants' preferences for mutually exclusive word-object mappings.

On word preference trials across Experiments 2 and 3, participants selected the transferred word mapping in 32% of cases and the second familiarization word mapping in the remaining 68% of cases. This bias for the second familiarization mapping across Experiments 2 and 3 was significant ($t(88)=7.1$, $p < .01$). An ANOVA comparing preferences between Experiment 2 and Experiment 3 revealed a significant difference ($F(1,87) = 4.1, p < .05$), indicating that the presence of contextual cues significantly altered the pattern of responses. Specifically, when no contextual cues were present, participants favored the second familiarization mapping in 76% of cases; when contextual cues were present, this preference dropped to 65%.

We separately analyzed performance on 2nd Familiarization test trials which retested 1:1 mappings learned in the 1st Familiarization. Specifically, we compared performance on 1st Familiarization words that had not been remapped in the 2nd Familiarization (“untransferred items”) to performance on those which had been remapped (“transferred items”). Collapsing across Experiments 2 and 3, participants correctly recalled 69% of untransferred items and 79% of transferred items. Performance was significantly above chance for both untransferred (*t*(88)=17.1, $p < .001$) and transferred items (*t*(88)=25.1, $p < .001$), and the difference in performance on untransferred and transferred item was also significant (*t*(88)=3.9,$p < .001$).

Participants in Experiment 2 correctly recalled 69% of untransferred items and 74% of transferred items. Participants in Experiment 3 correctly recalled 69% of untransferred items and 80% of transferred items. An ANOVA comparing performance between Experiment 2 and Experiment 3 revealed no significant differences for either untransferred ($F(1,87)=.001$, $p = .98$) or transferred items ($F(1,87)=1.7$, $p = .19$).

4.7 Experiment 3 Discussion

Experiment 3 was identical to Experiment 2 except for the addition of contextual cues to the second familiarization. Unlike in Experiment 2, participants in Experiment 3 learned many-to-one mappings at above chance levels. Furthermore, results from the preference trials showed that participants' tendency to endorse mappings consistent with the mutual exclusivity principle was significantly weakened by the presence of contextual cues. These differences in performance regarding the formation of many-to-one word-object mappings from Experiment 2 to Experiment 3 suggest that contextual cues facilitate the formation of multiple representations over distinct statistical inputs. Our findings did not indicate any significant differences in remapping performance across the three contextual cue conditions. Thus while contextual cues seem to highlight the availability of many-to-one mappings in the input, our results suggest that the qualitative differences between the contextual cues used in Experiment 3 do not differentially influence remapping performance.

5 Experiment 4

The experiments presented to this point have employed offline tasks to measure learning. This method of evaluation offers a static description of learning immediately after familiarization, but offers no information regarding the time course of learning during familiarization. Experiment 4 adopts the methodology of Yurovsky and Yu (2008) in order to evaluate the time course of learning, and specifically how changes in the availability of contextual cues may modulate the formation of many-to-one mappings. Interestingly, this study found that the acquisition of many-to-one mappings increases significantly when the number of words and objects presented in each learning trial of the second familiarization is balanced (i.e., 3 words and 3 objects, compared to 4 words and 3 objects in Experiments 2 and 3), a paradigm which essentially forces the learner to remap, since unambiguous evidence for the new mapping, and no evidence for the previous mapping, is available. In Experiment 4, we take advantage of this methodology to ensure that many-to-one mappings are acquired, and accordingly, to determine how contextual cues modulate the time course of acquisition. Exp. 4 was thus composed of three conditions; a baseline condition, in which all stimuli across the first and second familiarizations were presented in Voice 1 (American English female); and two contextual cue conditions, in which stimuli in the first familiarization were presented in Voice 1, and stimuli in the second familiarization were presented either in Voice 2 (American English male) or following an explicit cue (i.e., stating that many-to-one mappings would be available in the second familiarization). To evaluate the time course of learning, we required participants to rate their confidence in each mapping from a training trial immediately after that trial. These ratings provided a measure of how confidence in mappings changed across a familiarization, as well as how confidence was influenced by the availability of contextual cues. As such, both of these measures allowed for an investigation of how contextual cues influence the time course of mapping and remapping in the cross-situational paradigm.

5.1 Experiment 4a Subjects

Twenty introductory Psychology students (15 female and 5 male; 18-23 years) participated in Exp. 4a for course credit. None had participated in Exp. 1, 2, 3 or 4.

5.2 Experiment 4a Stimuli and Procedure

The first familiarization of Experiment 4a was similar to that of Experiments 2 and 3; participants encountered 6 presentations of 18 unique word object pairs across 36 trials containing 3 objects and 3 words each. Additionally, participants were asked to rate their confidence in word-object mappings following each familiarization trial. Thus, following each trial, participants saw a series of three displays, each of which presented one of the objects from that trial. The order of presentation of the three objects was randomized. On each screen, participants were asked to "rate the confidence with which they thought they knew the object's name" on a scale from 1 to 9, where a rating of 9 indicated the highest level of confidence. Each slide remained visible until the participant made a response, at which time the next object of the series was immediately presented. After participants rated all three objects, a fixation cross appeared for 250 ms, followed by the presentation of the next training trial. The test following the first familiarization was identical to that used in Experiments 2 and 3; participants completed

18 trials, in which each word-object pair learned in the first familiarization was tested once.

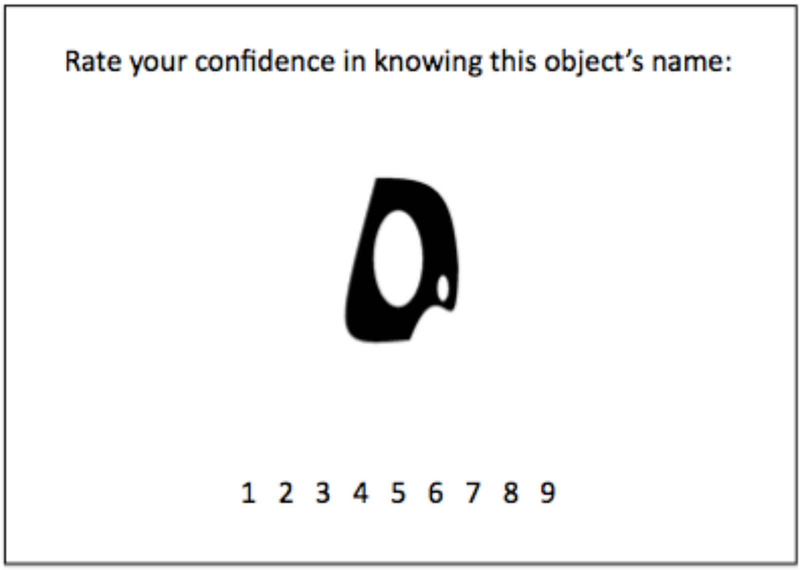

Figure 5.1 A confidence rating was collected for each mapping presented in a training trial, immediately after presentation of that training trial. Participants rated their confidence on a 9 point scale, on which a rating of 1 indicates the lowest level of confidence.

Following this test, the organization of the second familiarization diverged significantly from Experiments 2 and 3. A set of six learned words was again transferred to the second familiarization. There were 36 familiarization trials, each of which presented 3 objects and 3 words. These objects and words formed 3 new word-object mappings; two of these mappings were composed of novel objects and words that had not been encountered in the first familiarization; the third mapping was composed of one novel object and one of the transferred first familiarization words. Thus, many-to-one mappings were available to participants in the second familiarization. Each transferred first familiarization word co-occurred perfectly with its novel second familiarization object mapping (i.e., a transferred word always co-occurred with its second familiarization object mapping). All second familiarization mappings were presented 6 times across the 36 trials. As in the first familiarization, participants were required to rate their confidence in word-object mappings following each trial of the second familiarization.

The test following the second familiarization in Experiment 4a also differed from that of Experiments 2 & 3. Specifically, this test contained two blocks of trials testing the many-to-one mappings available in the second familiarization. A first set of these test trials was embedded and interspersed among the twelve trials testing the 1:1 mappings available in the second familiarization. The second block of trials testing many-to-one mappings immediately followed the Retest items. The purpose of this second set of many-to-one test trials was both to increase statistical power (i.e., previous experiments contained only six of these many-to-one test trials), and also to provide a measure of retention of these mappings following reactivation of primacy mappings by the Retest trials.

5.3 Experiment 4a Results

Participants acquired 70% of mappings in the first familiarization, 48% of the 1:1 mappings presented in the second familiarization, and recalled 64% of the first familiarization mappings at Retest. Performance on these test trial types (1st Test, 2nd Test, Retest) was significantly above chance levels (25%) (single-sample t-test: 1st Test: $t(20) = 10.5$, $p < .001$; 2nd Test: $t(20) = 7.5$, $p < .001$; Retest: $t(20) = 7.4$, $p < .001$). There was no significant difference between performance on the 1st Test and Retest (paired samples t-test: $t(20) = -1.7$, $p = .11$); however, the difference in performance between the 1st Test and 2nd Test items was significant (paired samples t-test: $t(20) = 4.6$, $p < .001$).

Participants completed two sets of trials testing multiple mappings; one embedded within the 2nd Test (Trans Item 1) and one immediately following the Retest trials (Trans Item 2).

Participants accuracy on Trans Item 1 and Trans Item 2 trials was 70% and 61% respectively. Performance on both of these sets of test trials was significantly above chance (single-sample t-test: Trans Item 1: $t(20) = 7.7$, $p < .001$; Trans Item 2: $t(20) = 7.1$, $p < .001$), indicating that the recency mappings for transferred words were successfully acquired. This replicates the finding of Yurovsky and Yu (2008), demonstrating that learners can successfully form multiple mappings when provided with explicit evidence of their existence. The 9% decrement in performance from Trans Item 1 to Trans Item 2 trials reached marginal significance (paired samples t-test: $t(20) = 1.8$, $p = .08$), and may indicate a decay in the retention of recency mappings, or, alternatively, interference from the immediately preceding Retest items (which reactive the primacy mappings of the transferred words).

5.4 Experiment 4a Confidence Ratings

Participants provided six confidence ratings for each object seen during familiarization (one rating following each presentation of an object). Confidence ratings across the first familiarization of Experiment 4a rose from an average of 4.7 out of 9 at the first presentation to 6.5 at the sixth presentation. This difference was significant (paired samples t-test: $t(21) = 3.4$, $p < .01$). The average slope by participant across the first familiarization was calculated using the average confidence ratings from the first and sixth presentation of each object; this value was .34. Confidence ratings across the second familiarization, in which recency mappings to previously encountered items were available, rose monotonically as well, from 3.1 to 6.2 from the first to the sixth presentation of each object. This difference was also significant (paired samples t-test: $t(21) = 5.2$, $p < .001$). The average slope of the confidence function across the second familiarization was .63. The difference in slopes between the first and second familiarization was significant (paired samples t-test: $t(21) = 3.3$, $p < .01$), indicating that the rise in confidence ratings across the second familiarization was greater than that of the first. Examining confidence ratings across the familiarizations by presentation number (i.e., 1-6), we found that ratings across the first three presentations were significantly lower in the second familiarization compared to the first (paired samples t-test: Presentation 1: $t(21) = 5.1$, $p < .001$; Presentation 2: $t(21) = 4.6$, $p < .001$; Presentation 3: $t(21) = 2.8$, $p < .05$), while ratings across the last 3 presentations in each familiarization did not differ statistically. This may indicate an initial period of confusion caused by the availability of many-to-one mappings, after which confidence recovered and ratings reached the levels of the first familiarization.

5.5 Experiment 4a Discussion

Experiment 4a replicates Yurovsky and Yu (2008), indicating that learners can acquire many-to-one mappings given explicit evidence of their existence, and furthermore that the formation of many-to-one mappings incurs a significant cost in terms of learner confidence early in the acquisition process. Specifically, confidence ratings in the second familiarization, where many-to-one mappings were available, were lower compared to those of the first familiarization, which presented only 1:1 mappings. This decrease in confidence ratings was most apparent over the first three presentations of an object in the second familiarization, after which confidence ratings matched those from the first familiarization. This decrease in confidence ratings may be caused by the presence of recency mappings for transferred words as well as learners' initial preference for mutually exclusive mappings, a preference which erodes as more evidence supporting many-to-one mappings is encountered. In Experiments 4b and 4c, we add contextual cues to the second familiarization. We predict that the addition of these cues may mitigate the initial period of confusion in the second familiarization, specifically by increasing learners' prior expectation for a change in the input. Accordingly, we predict that contextual cues will lead to higher confidence for both 1:1 and 2:1 mapping types across the second familiarization.

5.6 Experiment 4b (Gender Cue) Subjects

Twenty introductory Psychology students (11 female and 9 male; 18-25 years) participated in Exp. 4b for course credit. None had participated in Exp. 1, 2, 3 or 4.

5.7 Experiment 4b Stimuli and Procedure

Experiment 4b was identical to Experiment 4a, save for the addition of a contextual cue to the second familiarization. Thus, all words presented in the second familiarization were produced in Voice 2 (male American English).

5.8 Experiment 4b Results

Participants acquired 78% of mappings in the first familiarization, 45% of the 1:1 mappings presented in the second familiarization, and recalled 74% of the first familiarization mappings at Retest. Performance on these test trial types (1st Test, 2nd Test, Retest) was significantly above chance levels (25%) (single-sample t-test: 1st Test: $t(20) = 15.9$, $p < .001$; 2nd Test: $t(20) = 5.2$, p < .001; Retest: *t*(20) = 11.1, p < .001). There was no significant difference between performance on the 1st Test and Retest (paired samples t-test: t(20) = -1.3, p = .21); however, the difference in performance between the 1st Test and 2nd Test items was significant (paired samples t-test: *t*(20) = -18.8, p < .001).

Participants accuracy on Transferred Items 1 and Transferred Items 2 trials was 65% and 61% respectively. Performance on both of these sets of test trials was significantly above chance (single-sample t-test: Trans Item 1: *t*(20) = 5.9, p < .001; Trans Item 2: *t*(20) = 7.1, p < .001), indicating that the recency mappings for transferred words were successfully acquired. The 4% decrement in performance from Transferred Items 1 to Transferred Items 2 trials was not significant (paired samples t-test: *t*(20) = .74, p = .47).

We compared learning performance in Experiment 4b to performance in Experiment 4a. Every test trial type was compared: there were no significant differences in learning performance within any of the test trial types between experiments (First Test: *t*(39) = -1.4, p = .14; Second Test: *t*(39) = .62, p = .3; Retest: *t*(39) = -1.5, p = .465; Transferred Items 1: *t*(39) = .55, p = .38; Transferred Items 2: *t*(39) = -.04, p = .7).

5.9 Experiment 4b Confidence Ratings

Confidence ratings across the first familiarization of Experiment 4b rose from an average of 5.0 out of 9 at the first presentation to 6.9 at the sixth presentation. This difference was significant (paired samples t-test: t(20) = 3.5, p < .01). The average slope by participant across the first familiarization was calculated using the average confidence ratings from the sixth and first presentation of each object; this value was .39. Confidence ratings across the second familiarization were grouped by mapping type. Ratings for 1:1 mappings rose from an average of 2.9 to 6.89; this difference was significant ($t(20) = 6.6$, $p < .001$). Ratings for 2:1 mappings rose from an average of 2.6 to 6.8; this difference was significant ($t(20) = 6.7$, $p < .001$). The average slope of the confidence function for 1:1 mappings across the second familiarization was .78. The difference in the slope of the confidence function for 1:1 mappings between the first and second familiarization was significant (paired samples t-test: $t(20) = 3.65$, $p < .01$). Examining confidence ratings for 1:1 word-object pairs across the familiarizations by presentation number (i.e., 1-6), we found a significant difference for the first and second presentations (paired-samples t-test: 1st Presentation: $t(20) = 4.8$, $p < .001$; Second Presentation: $t(20) = 3.4$, $p < .01$), such that ratings from the second familiarization were lower than those from the first.

5.10 Experiment 4c (Explicit Cue) Participants

Twenty-one introductory Psychology students (13 female and 8 male; 18-22 years) participated in Exp. 4c for course credit. None had participated in Exp. 1, 2, 3 or 4.

5.11 Experiment 4c Stimuli and Procedure

Experiment 4c was identical to Experiment 4a, save for the addition of an explicit contextual cue to the second familiarization. Thus, after the test following the first familiarization and before the onset of the second familiarization, participants were told that a subset of words they had previously encountered would be mapped to new objects.

5.12 Experiment 4c (Explicit Cue) Results

Participants acquired 71% of mappings in the first familiarization, 50% of the 1:1 mappings presented in the second familiarization, and recalled 68% of the first familiarization mappings at Retest. Performance on these test trial types (1st Test, 2nd Test, Retest) was significantly above chance levels (25%) (single-sample t-test: 1st Test: $t(20) = 11.0$, $p < .001$; 2nd Test: $t(20) = 9.8$, $p < .001$; Retest: $t(20) = 9.5$, $p < .001$). There was no significant difference between performance on the 1st Test and Retest (paired samples t-test: $t(20) = -1.1$, $p = .29$); however, the difference in performance between the 1st Test and 2nd Test items was not significant (paired samples t-test: $t(20) = -9.8$, $p < .001$).

Participants' accuracy on Transferred Items 1 and Transferred Items 2 trials was 67% and 57% respectively. Performance on both of these sets of test trials was significantly above chance (single-sample t-test: Trans Item 1: $t(20) = 5.4$, $p < .001$; Trans Item 2: $t(20) = 5.3$, $p < .001$), indicating that the recency mappings for transferred words were successfully acquired. The 10% decrement in performance from Transferred Items 1 to Transferred Items 2 trials was no significant (paired samples t-test: $t(20) = 1.6$, $p = .12$).

We compared learning performance in Experiment 4c to performance in Experiment 4a. Every test trial type was compared: there were no significant differences in learning performance within any of the test trial types between experiments (First Test: $t(39) = -.19$, $p = .99$; Second Test: $t(39) = -.71$, $p = .43$; Retest: $t(39) = -.71$, $p = .3$; Transferred Items 1: $t(39) = .34$, $p = .21$; Transferred Items 2: $t(39) = .47$, $p = .19$).

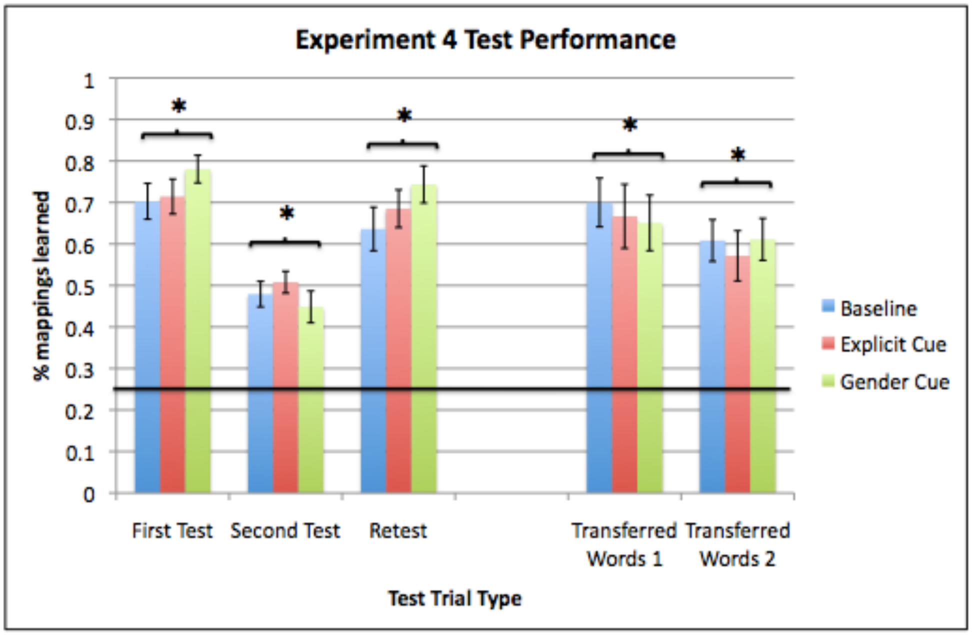

Figure 5.2 Performance on all test trial types in each condition of Experiment 4. Performance in each condition on each test trial type exceeded chance (25%). There were no significant differences between conditions within any test trial type.

5.13 Experiment 4c Confidence Ratings

Confidence ratings across the first familiarization of Experiment 4c rose from an average of 4.0 out of 9 at the first presentation to 6.8 at the sixth presentation. This difference was significant (paired samples t-test: $t(20) = 5.3$, $p < .001$). The average slope by participant across the first familiarization was calculated using the average confidence ratings from the sixth and first presentation of each object; this value was .56. Confidence ratings across the second familiarization were grouped by mapping type. Ratings for 1:1 mappings rose from an average of 2.9 to 6.6; this difference was significant ($t(20) = 6.9$, $p < .01$). Ratings for 2:1 mappings rose from an average of 2.7 to 6.9; this difference was significant ($t(20) = 6.2$, $p < .01$). The average slope of the confidence function for 1:1 mappings across the second familiarization was .73. The difference in the slope of the confidence function for 1:1 mappings between the first and second familiarization was marginally significant (paired samples t-test: $t(20) = 1.98$, $p = .061$). Examining confidence ratings for word-object pairs across the familiarizations by presentation number (i.e., 1-6), we found a significant difference only for the first presentation (paired-samples t-test: $t(20) = 2.2$, $p < .05$), such that ratings from the second familiarization were lower than those from the first.

5.14 Regression Models

Linear regression models were used to determine the influence of contextual cues on mapping and remapping in Experiment 4. As such, we compared learning in Experiment 4a, the baseline condition, to learning in the contextual cue conditions (Experiments 4b and 4c), which were combined into a single "contextual cue" condition for the purpose of this analysis. Three models compared learning in the baseline condition to the contextual cue condition; there was one model comparing confidence ratings for 1:1 mappings presented in the first familiarization of each condition; one model comparing confidence ratings for 1:1 mappings in the second familiarization of each condition; and one model comparing confidence ratings for 2:1 mappings in the second familiarization. Each model contained a factor for subject, presentation number of a word-object pair (1-6), object, response time, and condition (i.e., baseline or contextual cue). Confidence ratings were the dependent variables for each regression. The regressions were run over the full set of trial-by-trial confidence rating responses, from all of the participants, pertaining to each of the three comparisons outlined above.

The first model compared the learning of 1:1 mappings in the first familiarization between the baseline and contextual cue conditions. The full model was significant ($F(5, 6690) = 111.74$, $p < .001$), indicating that the set of factors used explained a significant amount of the variation in confidence ratings. The factor for presentation number was significant ($t = 22.9$, $p < .001$) indicating that there was a significant increase in confidence ratings across the first familiarization in both the baseline and contextual cue conditions. Critically, the condition factor did not reach significance ($t = 1.6$, $p = .102$), indicating that there was no difference in confidence ratings for 1:1 mappings between the baseline and contextual cue conditions. None of the other factors (subject, object, response time) reached significance in the model.

The second model compared the learning of 1:1 mappings in the second familiarization between the baseline and contextual cue conditions. The full model was significant ($F(5, 4334) = 197.34$, $p < .001$). The factors for object and response time did not reach significance. The factor for presentation number was again significant ($t = 30.9$, $p < .001$) indicating a rise in confidence ratings for 1:1 mappings in the second familiarization for both the baseline and contextual cue conditions. The factor for condition was also significant ($t =3.84$, $p < .001$), indicating a significant increase in confidence ratings from the baseline to the contextual cue condition. The subject factor was also significant ($t = -3.87$, $p < .001$) indicating a significant degree of variance in confidence ratings between participants in the second familiarization.

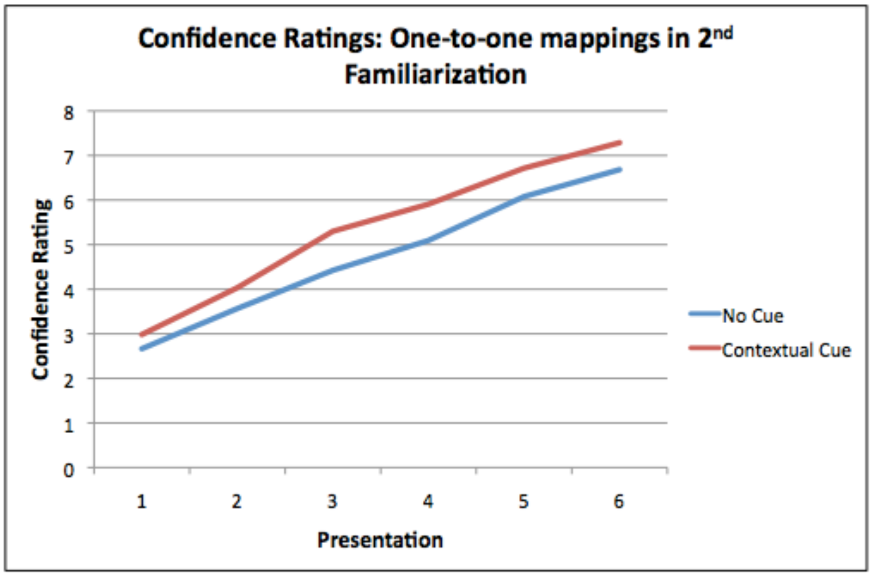

Figure 5.3 Trajectory of confidence ratings for words in 1:1 mappings in the second familiarization of Experiments 4a-c.

The third model compared the learning of 2:1 mappings in the second familiarization between the baseline and contextual cue conditions. The full model was significant ($F(5, 2164) = 100.9$, $p < .001$). The factor for response time did not reach significance ($t = -.49$, $p = .63$). The factor for presentation again reached significance ($t = 21.6$, $p < .001$), along with the factor for subject ($t = -3.5$, $p < .001$) and object ($t = 2.65$, $p < .01$). The factor for condition also reached significance ($t = 3.22$, $p < .01$) indicating a significant increase in confidence ratings for 2:1 mappings from the baseline to the contextual cue condition.

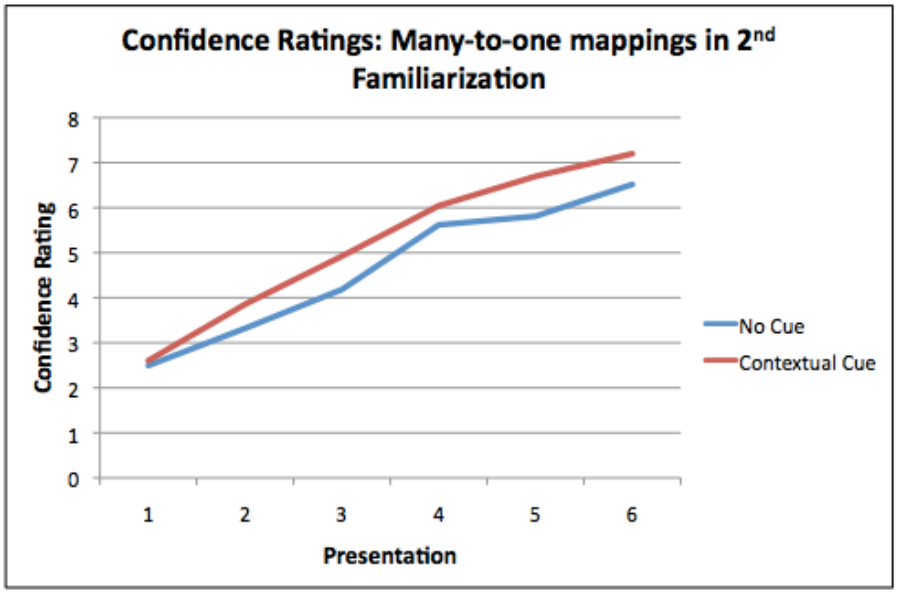

Figure 5.4 Trajectory of confidence ratings for words in 2:1 mappings in the second familiarization in Experiments 4a-c.

5.15 Experiment 4 Discussion

Experiments 4b and 4c, the contextual cue conditions of Experiment 4, explored the influence of contextual cues on the trajectory of learning in a cross-situational paradigm. Learning was significantly above chance for all test trial types across both familiarization of Experiments 4b and 4c. Furthermore, a comparison of learning in Experiment 4a, the baseline condition, to learning in Experiments 4b and 4c showed no significant differences for any test trial type. This suggests that contextual cues did not influence "end-state" knowledge of 2:1 mappings in Experiment 4. This result diverges from the findings of Experiment 3, but may be attributable to a change in the experimental paradigm that reduced mapping ambiguity in the second familiarization of Experiment 4.

However, the comparisons of confidence ratings in the baseline and contextual cue conditions suggest an effect of contextual cues on the trajectory of learning. In all conditions of

Experiment 4, confidence ratings rose across the first familiarization. A regression model compared the magnitude of the confidence ratings in the first familiarization across the baseline and contextual cue conditions of Experiment 4 and found no significant differences. Importantly, this establishes that any effects of contextual cues on learning in the second familiarization are not attributable to differences in learner confidence in the first familiarization. As noted, confidence ratings in the second familiarization of the baseline condition were lower compared to those of the first familiarization, a result that may have been caused by an initial period of confusion upon encountering 2:1 mappings. Two further regression models, one each for 1:1 and 2:1 mappings, compared confidence ratings in the second familiarization across the baseline and contextual cue conditions. Each model found that confidence ratings were significantly larger in contextual cue conditions compared to baseline. This finding suggests that the addition of contextual cues to the second familiarization reduced mapping ambiguity caused by the availability of 2:1 mappings. This reduction in ambiguity positively influenced learner confidence in both 2:1 and 1:1 mappings. In accordance with the results of Experiment 3, this finding suggests that contextual cues saliently signal a change in the statistics of the input, allowing for more rapid accommodation of variability in the structures to be learned.

6 General Discussion

In an initial series of three experiments, we investigated how contextual cues impact statistical word learning. In Experiment 1, we replicated the results reported by Yu and Smith (2007), demonstrating that learners can acquire word-object mappings by tracking co-occurrence statistics across scenes containing multiple objects and labels. In Experiment 2, we extended these results to demonstrate that learners exhibit a primacy effect in the CSSL paradigm, preferring to maintain 1:1 mappings even when the input supports the formation of 2:1 mappings. These methods represent a partial replication of Ichinco, Frank, and Saxe (2009) and a slight extension due to our sequencing constraints that precluded items from appearing in adjacent trials. In Experiment 3, we demonstrated that the primacy effect in statistical word learning can be overcome by adding contextual cues to the input. Specifically, we tested three contextual cues: a voice cue, a voice and accent cue, and an explicit cue. In each case, the addition of a contextual cue facilitated the acquisition of many-to-one mappings compared to a condition with no cue (Experiment 2). In Experiment 4, we found that contextual cues significantly decrease mapping ambiguity in conditions containing both 1:1 and 2:1 mappings. This attenuation of ambiguity appeared as a significant increase in the confidence with which participants rated their learning of word-object pairs, compared to a baseline condition containing no contextual cues.

Previous work in statistical learning has found that learners have difficulty acquiring a second distribution after exposure to a first. Specifically, when transitions between distributions are not cued, learners exhibit either primacy or interference effects. These results implicate a learning mechanism that either accumulates statistics without regard to internal consistency (i.e., a mechanism that combines all statistics into a single representation, regardless of the number of distributions from which they arise), or a mechanism that requires additional sources of information in order to disambiguate inputs and accommodate each with a separate representation. One source of this information may be more implicit in nature, involving changes in the duration or sequencing of multiple inputs. For example, Gebhart, Aslin and Newport (2009) showed that increasing the duration of a second distribution relative to a first facilitates acquisition of both distributions, even if both are presented in the same voice. Likewise, Zinszer and Weiss (in review) showed that increasing the number of switches between distributions guides learners to the conclusion of multiple inputs (again, even without a change in voice), thereby facilitating acquisition. In a word learning study, Kachergis, Yu and Shiffrin (2010) showed that the formation of many-to-one mappings was positively correlated with the number of exposures to a second mapping after a first had been learned. Thus, as learners receive more input supporting a second mapping, or experience regular but uncued changes (i.e., switches) from one distribution to another, the strength of primacy and interference effects is reduced. These findings regarding implicit cues provide a number of topics for future research on the learning of multiple inputs. Specifically, we might begin to explore how differences in other properties of distributions, such as phonemic or syllabic inventory, phonotactics, phonology, or prosody, influence the ability of learners to detect the presence of multiple inputs. Importantly, these properties map onto actual differences between languages, and may inform an understanding of how the present research scales up to natural language.

The results of the present work in word learning, as well as the studies from speech segmentation reviewed above, suggest that contextual cues as well may signal a change in language context to learners, thereby facilitating accommodation of the distinct inputs present in their learning environment. Specifically, we propose that unique cues (or sets of cues) which co-occur with distributions serve to identify those distributions, as well as highlight points of transitions between them. While both implicit and explicit cues serve to facilitate the acquisition of multiple inputs, here we propose that explicit contextual cues serve as shortcuts for learners, either removing or significantly attenuating the primacy and interference effects noted in previous research on the acquisition of multiple inputs, and significantly reducing the overall amount of exposure necessary for the acquisition of multiple inputs. Interestingly, although we tested three distinct cues in the present study, we found no significant differences in the degree to which these cues supported the formation of multiple mappings. This finding seemingly supports the conclusion that changes in context serve primarily as shortcuts to learning, and further, that learners are able to take advantage of a range of systematic contextual variation as long as that variation is correlated with differences in statistical structure. Despite this, previous studies have noted selectivity with regard to the contextual cues that learners are able to use to segregate inputs (Gebhart, Aslin and Newport, 2009; Mitchel and Weiss, 2010). However, the present study largely employed cues that have been found effective and highly salient in paradigms containing multiple inputs. Future research may explore a broader range of contextual cues, especially with respect to their salience in natural language as markers of change, to determine how the properties of specific cues influence their effectiveness in signaling transitions between inputs.

Additionally, it is worth noting that previous studies have found that changes in speaker voice, found here to facilitate segregation of inputs, are also useful for generalizing across speakers and forming well-defined categories of linguistic structure for individual phonemes or words (Lively et al., 1993; Bradlow and Bent, 2003). This finding suggests a flexibility on the part of learners in interpreting the local function of a contextual cue. Specifically, while the individual cues studied so far may be capable of supporting conclusions of both segregation and accommodation, experienced learners may be capable of inferring through sets of other cues (such as phonemic inventory, phonology, or lexicon) the consistency of the underlying statistics of language input. If learners determine through these cues that language input is not changing, a change in speaker voice may be more viably employed to generalize across speakers. However, sets of cues converge on the conclusion of a change in language input, surface cues such as a change in speaker voice may become stronger invitations for learners to segregate inputs. In contrast to experienced learners, infant learners may be incapable of adequately using other cues to determine the consistency of language input, and as a result may be more inclined to use salient contextual cues such as changes in speaker voice to segregate inputs, even when segregation in inappropriate (Houston et al., 1998).

Regardless, the results from word learning and speech segmentation studies suggest that the statistical learning mechanism is sensitive to context, and that learners can exploit this sensitivity to optimize processing in environments which contain an unknown number of inputs. This finding may have implications for bilingual learners: specifically, despite receiving less input for each of their languages compared to monolingual learners, bilinguals are not delayed relative to monolinguals in their achievement of linguistic milestones. The proposed role of contextual cues as shortcuts to the acquisition of multiple inputs may partially explain this lack of a developmental delay for bilinguals, suggesting that bilinguals close the gap in language exposure by the use of contextual cues to quickly disambiguate and acquire distinct inputs

The evidence presented so far from the statistical learning literature suggests that contextual cues play an important role in acquisition in environments that contain multiple inputs. Learners are sensitive to changes in both linguistic and extralinguistic cues, and to date, the influence of a subset of these cues (including speaker gender, pitch, as well as visual context) has been tested in statistical learning paradigms (Weiss, Gerfen and Mitchel, 2009; Gebhart, Aslin and Newport, 2010; Mitchel and Weiss, 2010). However, contextual cues may not be necessary for the formation of multiple inputs. As noted, several previous studies demonstrate that additional exposure to a second distribution or mapping facilitates the formation of multiple representations. Additionally, learners may be guided by their prior expectation for a change in language context. That is, previous exposure to environments that contain multiple inputs may condition learners to more readily anticipate the presence of multiple inputs. This increased prior expectation, in turn, may facilitate the discovery of transitions between inputs, even when no contextual cues saliently mark those transitions. For instance, a number of word-learning studies (Byers-Heinlein and Werker, 2010; Houston-Price 2012) have found that bilinguals more easily form many-to-one linguistic associations compared to monolinguals, and furthermore have demonstrated that reliance on mutual exclusivity decreases as the number of languages a learner has been exposed to increases (i.e., trilinguals are less bound by ME than bilinguals). These studies have been limited to lexical learning so far; however, it may be reasonable to assume that this flexibility in forming complex associations might also facilitate phonological and syntactic learning as well.

6.1 Statistical learning or Hypothesis Testing?

The findings from this study also have implications for a recent discussion within the word learning literature. The initial studies employing CSSL have assumed that learning is achieved by tracking the co-occurrence of words and objects across remembered scenes (e.g., Yu & Smith, 2007). The core assumption of this reductionist approach to word learning is that the variability in the environment of word learners (e.g., how often a word is heard in the presence of its referent, the length of the interval between occurrences of a word, etc.) may be overcome through the accumulation of evidence over a series of encounters. However, this assumption has recently been questioned on several grounds pertaining to the extent to which this paradigm scales up to the problems that must be overcome in the real world (see Medina, Snedeker, Trueswell, & Gleitman, 2011). In a series of experiments that adopted a somewhat more naturalistic approach to word learning, Medina et al. (2011) concluded that in situations with multiple objects and labels available, learners do not compute statistics over multiple scenes. Rather, the authors propose that learners entertain a single meaning conjectured for each word, which carries forward as the learner seeks confirming or disconfirming evidence. As such, no alternative meanings are considered nor are there statistics retained across epochs. This view differs sharply from the proposed mechanisms that form the basis for the CSSL paradigm of word learning.

Of particular interest here, the paradigm adopted in Experiment 3 may shed light on this debate. Learners in Experiment 3 encountered statistically supported word-object pairings during the first familiarization. Six of those objects were then transferred to the second familiarization and paired with both the old object as well as a new object. Unlike Experiment 2, in which participants displayed consistent mutual exclusivity preferences, participants in Experiment 3 were willing to consider the new mapping for the transferred word at above chance levels, presumably as a consequence of a contextual cue that encouraged the formation of a separate statistical representation. At first glance, these results are at odds with the premise of hypothesis testing endorsed by Medina and colleagues. Specifically, learners never encountered disconfirmatory evidence for an initial mapping hypothesis for transferred words, as multiple mappings were available but not forced in the second familiarization. However, it is important to note that while learners in all conditions of Experiment 3 were more likely to entertain a possible new mapping, they still frequently demonstrated awareness of, and preference for, the original mapping (as shown by performance on test trials pitting the old object mappings against the new.

Consequently, these differing accounts of learning may be more similar than they first appear, given that under either view they rest on the premise that learners form contextually sensitive hypotheses about mappings that take into account prior encounters with word-object pairings.

As a final point of note here, it may be the case that the experimental paradigm employed by Medina et al. significantly influenced their finding regarding learner strategy. Specifically, Medina et al. required learners to entertain a hypothesis about word-object mappings after every trial. Requiring learners to make hypotheses may, in turn, have contributed to the finding of processing that suggests a hypothesis-testing strategy. Future research on this topic may remove this requirement and assess learner's responses to mapping competitors, instead (e.g., words that co-occur with a certain object frequently but which are not, in fact, mapped to that object). The prediction here is that competitors would not exert a noticeable influence on processing under a hypothesis-testing account of learning, as only the current hypothesis regarding a mapping is retained in memory. However, a statistical learning account might predict that learners track co-occurrence statistics for all words and objects. Thus, competitors might be expected to exert a noticeable influence on processing.

6.2 Mutual Exclusivity

While this work demonstrates that primacy effects in statistical learning can be overcome by contextual cues, it does not rigorously address the origin of primacy effects nor the precise mechanism by which contextual cues mediate primacy. Previous attempts to describe mutual exclusivity have invoked an inferential learning mechanism, under which account ME is a hard-wired constraint that awakens at a certain point in development. The application of ME is considered to be critical for resolving referential ambiguity and facilitating "in the moment" mapping of words to objects. Thus, through the function of ME and a number of other proposed heuristics (e.g., the whole-object preference), the indeterminacy of word-mapping is resolved.

Word learning, however, may not require a specialized set of language-specific learning mechanisms. A number of recent studies have shown that a realistic model of word recognition and novel word learning can result from a simple associative learning or Bayesian inference mechanism (Regier, 2005; Mayor and Plunkett, 2010; Xu and Tenenbaum, 2007; Frank et al., 2009; McMurray, 2007). An account of mutual exclusivity consistent with these data-driven learning mechanisms might describe the principle as an emergent property of learning within a statistically structured environment. For instance, in a model that did not explicitly incorporate a rule for mutual exclusivity, training input that contained predominantly 1:1 mappings resulted in the emergence of a general learning principle that favored the acquisition of such 1:1 mappings. When the model was trained on input containing both 1:1 and 2:1 mappings, the strength of the bias towards 1:1 mapping diminished as a function of the proportion of 2:1 mappings in the input. That is, as more 2:1 mappings were encountered, learning depended less on an assumption of mutual exclusivity.

Thus, the nature of the input itself determined the learning biases that emerged across training. Interestingly, this finding parallels experimental work on the development of mutual exclusivity in infant populations of monolinguals and bilinguals. Byers-Heinlein and Werker (2009) and Houston-Price et al. (2010) showed that use of mutual exclusivity in a word-mapping task was correlated with prior language experience. That is, bilinguals, as a result of frequent exposure to multiple languages and experience acquiring translation equivalents, were significantly less likely to show processing consistent with mutual exclusivity, compared to monolinguals. The findings of both modeling and experiment studies, then, suggest that in addition to contextual cues, prior experience with language input significantly modulates processing in environments containing multiple inputs. Thus, bilingual or bidialectal learners may approach the challenge of language learning with an increased prior expectation for a change in the underlying statistics of their input. This in turn may allow such learners to more quickly detect and accommodate changes in the input, even in the absence of explicit contextual cues.

6.3 Conclusions

As noted above, studies of statistical learning have made a significant impact on the field of language development. Yet, with few exceptions (see Weiss et al., 2009; Gebhart et al., 2010; Mitchel and Weiss, 2010; Franco, Cleeremans, and Destrebecqz, 2011), almost all of the studies of statistical learning have assumed that the underlying structure is uniform. In reality, however, language acquisition proceeds despite substantial variability in the input. Some variability maybe characterized as statistical noise, but other variability could suggest that the input is either more complex or may be restricted to a particular context (indicating that there is more than one underlying statistical structure). The findings reported here, along with the previous work in the domain of speech segmentation, inform our understanding of this central challenge that confronts the naïve language learner. Namely, the learner must be capable of identifying when it is suitable and/or optimal to aggregate information across a corpus of input and when to create a new representational structure. This challenge may be as pertinent for monolinguals as it is for bilinguals, and for visual statistical information as well as auditory input (see Weiss et al., 2009;

Gebhart et al., 2010). When considered together, this body of research has begun to suggest that learners are in fact capable of tracking statistical information in a contextually sensitive manner. Given that we can demonstrate that learners have this capability, a related challenge is determining which cues in the signal denote a change in the underlying context. Our previous work suggests that individual identity cues may facilitate this process irrespective of whether they are auditory (Weiss et al., 2009) or visual (e.g., Mitchel and Weiss, 2010). Future research will focus on determining what types of cues (including linguistic cues such as stress patterns, phonotactic patterns, etc.) may be used by learners to successfully form multiple representations. Likewise, the goals of the larger research initiative are to determine the developmental and phylogenetic roots of these sensitivities.

References

Bartolotti, James, and Viorica Marian. "Language learning and control in monolinguals and bilinguals." *Cognitive science* 36.6 (2012): 1129-1147.

Bradlow, A. R., & Bent, T. (2003). Listener adaptation to foreign accented English. In M. J. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 2881-2884). Barcelona: Futurgraphic.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some eﬀects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101(4), 2299–2310.

Bortfeld, H., Morgan, J., Golinkoff, R., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Manuscript in preparation*.

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, 12, 815–823.

Carey, Susan, and Elsa Bartlett. "Acquiring a single new word." (1978).

Chen, Chi-hsin, et al. "Statistical Word Learning and Object Categorization: A Cross-Linguistic Study in English and Mandarin." *Proceeding of the 31st Annual Conference of the Cognitive Science Society*. 2009.

Chomsky, Noam. Aspects of the Theory of Syntax. Vol. 119. The MIT press, 1965.

Chomsky, N. (1980). The linguistic approach. In M. Piattelli-Palmarini (Ed.), *Language and Learning*. Cambridge, MA: Harvard University Press, 109–116.

Chomsky, N. (1981) Lectures on Government and Binding, Foris, Dordrecht.

Christoffels, Ingrid K., Annette de Groot, and Judith F. Kroll. "Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency." *Journal of Memory and Language* 54.3 (2006): 324-345.

Cole, Ronald A., Jola Jakimik, and William E. Cooper. "Segmenting speech into words." *The Journal of the Acoustical Society of America 67* (1980): 1323.

Elman, Jeffrey L. "Learning and development in neural networks: The importance of starting small." *Cognition* 48.1 (1993): 71-99.

Fiser, József, and Richard N. Aslin. "Unsupervised statistical learning of higher-order spatial structures from visual scenes." *Psychological science* 12.6 (2001): 499-504.

Franco, Ana, Axel Cleeremans, and Arnaud Destrebecqz. "Statistical learning of two artificial languages presented successively: how conscious?." *Frontiers in Psychology* 2 (2011).

Gebhart, Andrea L., Richard N. Aslin, and Elissa L. Newport. "Changing structures in midstream: Learning along the statistical garden path." *Cognitive science* 33.6 (2009): 1087-1116.

Goldwater, Sharon, and Mark Johnson. "Learning OT constraint rankings using a maximum entropy model." *Proceedings of the Stockholm workshop on variation within Optimality Theory*. 2003.

Golinkoff, Roberta Michnick, Carolyn B. Mervis, and Kathryn Hirsh-Pasek. "Early object labels: The case for a developmental lexical principles framework." *Journal of child language* 21 (1994): 125-125.

Gomez, Rebecca L., and LouAnn Gerken. "Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge." *Cognition* 70.2 (1999): 109-135.

Houston, D., Jusczyk, P.W., & Tager, J. (1998). Talker-specificity and the persistence of infants' word representations. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), Proceedings of the 22nd Annual Boston University Conference on Language Development, (vol. 1, pp. 385–396). Somerville, MA: Cascadilla Press.

Houston‐Price, Carmel, Zoe Caloghiris, and Eleonora Raviglione. "Language experience shapes the development of the mutual exclusivity bias." *Infancy* 15.2 (2010): 125-150.

Ichinco, Denise, Michael C. Frank, and Rebecca Saxe. "Cross-situational word learning respects mutual exclusivity." *Proceedings of the 31st annual meeting of the cognitive science society*. Vol. 31. 2009.

Joshi, Aravind K. "Processing of sentences with intra-sentential code-switching." *Proceedings of the 9th conference on Computational linguistics-Volume 1*. Academia Praha, 1982.

Jusczyk, Peter W., Elizabeth A. Hohne, and Angela Bauman. "Infants' sensitivity to allophonic cues for word segmentation." *Perception & psychophysics* 61.8 (1999): 1465-1476.

Kovács, Ágnes Melinda, and Jacques Mehler. "Cognitive gains in 7-month-old bilingual infants." *Proceedings of the National Academy of Sciences* 106.16 (2009): 6556-6560.

Kovács, Ágnes Melinda, and Jacques Mehler. "Flexible learning of multiple speech structures in bilingual infants." *Science* 325.5940 (2009): 611-612.

Legate, Julie Anne, and Charles D. Yang. "Empirical re-assessment of stimulus poverty arguments." *Linguistic Review* 19.1/2 (2002): 151-162.

Lidz, Jeffrey, Sandra Waxman, and Jennifer Freedman. "What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months." *Cognition* 89.3 (2003): 295-303.

Lively, S. E., Logan, J. D., and 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,’’ *J. Acoust. Soc. Am*. 94, 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. 1994. ‘‘Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories,’’ *J. Acoust. Soc. Am*. 96, 2076 – 2087.

Mattys, Sven L., and Peter W. Jusczyk. "Phonotactic cues for segmentation of fluent speech by infants." *Cognition* 78.2 (2000): 91-121.

Maye, Jessica, Janet F. Werker, and LouAnn Gerken. "Infant sensitivity to distributional information can affect phonetic discrimination." *Cognition* 82.3 (2002): B101-B111.

Maye, Jessica, Daniel J. Weiss, and Richard N. Aslin. "Statistical phonetic learning in infants: Facilitation and feature generalization." *Developmental Science* 11.1 (2008): 122-134.

Mayor, Julien, and Kim Plunkett. "A neurocomputational account of taxonomic responding and fast mapping in early word learning." *Psychological Review*117.1 (2010): 1.

McMurray, Bob. "Defusing the childhood vocabulary explosion." *Science*317.5838 (2007): 631-631.

Medina, Tamara Nicol, et al. "How words can and cannot be learned by observation." *Proceedings of the National Academy of Sciences* 108.22 (2011): 9014-9019.

Mitchel, Aaron D., and Daniel J. Weiss. "What's in a face? Visual contributions to speech segmentation." *Language and Cognitive Processes* 25.4 (2010): 456-482.

Monaghan, Padraic, and Karen Mattock. "Integrating constraints for learning word–referent mappings." *Cognition-International Journal of CognitiveScience*123.1 (2012): 133.

Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran. "Learning in reverse: Eight-month-old infants track backward transitional probabilities." *Cognition*113.2 (2009): 244-247.

Perfors, Amy Francesca, Joshua B. Tenenbaum, and Terry Regier. "Poverty of the stimulus? A rational approach." *The Annual Conference of the Cognitive Science Society (28th: 2006: Vancouver, Canada)*. 2006.

Perruchet, Pierre, and Sebastien Pacton. "Implicit learning and statistical learning: One phenomenon, two approaches." *Trends in cognitive sciences* 10.5 (2006): 233-238.

Pinker, Steven. "Learnability and cognition: The acquisition of argument structure." (1989).

Polka, L., Sundara, M., & Blue, S. (2002). The role of language experience in word segmentation: A comparison of English, French, and bilingual infants. *The 143rd Meeting of the Acoustical Society of America: Special Session in Memory of Peter Jusczyk*, Pittsburgh, PA.

Qian, Ting, T. Florian Jaeger, and Richard N. Aslin. "Learning to represent a multi-context environment: more than detecting changes." *Frontiers in Psychology* 3 (2012).

Quine. W.V.O. (1960). Word and object. Cambridge. MA: MIT Press.

Rebuschat, Patrick, and John N. Williams. "Implicit Learning in Second Language Acquisition." *The encyclopedia of applied linguistics* (2012).

Regier, T. (2005). The emergence of words: Attentional learning inform and meaning. Cognitive Science, 29, 819–865.

Regier, Terry, and Susanne Gahl. "Learning the unlearnable: The role of missing evidence." *Cognition* 93.2 (2004): 147-155.

Rost, Gwyneth C., and Bob McMurray. "Speaker variability augments phonological processing in early word learning." *Developmental Science* 12.2 (2009): 339-349.

Rost, Gwyneth C., and Bob McMurray. "Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning." *Infancy* 15.6 (2010): 608-635.

Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport. "Statistical learning by 8-month-old infants." *Science* 274.5294 (1996): 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.

Shriberg, E. (1999). Phonetic consequences of speech disfluency. In Proceedings of the XIVth International Congress on Phonetic Sciences (pp. 619–622). San Francisco.

Tang, Chaoju, and Vincent J. van Heuven. "Mutual intelligibility of Chinese dialects experimentally tested." *Lingua* 119.5 (2009): 709-732.

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science*, *34*, 1092-1106.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209–253.

Tree, Jean E. Fox. "The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech." *Journal of memory and language* 34.6 (1995): 709-738.

Trude, Alison M., and Sarah Brown-Schmidt. "Talker-specific perceptual adaptation during online speech perception." *Language and Cognitive Processes* 27.7-8 (2012): 979-1001.

Vouloumanos, Athena. "Fine-grained sensitivity to statistical information in adult word learning." *Cognition* 107.2 (2008): 729-742.

Weiss, Daniel J., Chip Gerfen, and Aaron D. Mitchel. "Speech segmentation in a simulated bilingual environment: A challenge for statistical learning?" *Language Learning and Development* 5.1 (2009): 30-49.

Xu, Fei, and Joshua B. Tenenbaum. "Word learning as Bayesian inference." *Psychological review* 114.2 (2007): 245.

Yim, Dongsun, and John Rudoy. "Implicit statistical learning and language skills in bilingual children." *Journal of Speech, Language and Hearing Research* 56.1 (2013): 310.

Yu, Chen, Smith, Linda B. Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*. 2007;18(5):414–420

Yu, Chen, and Linda B. Smith. "What you learn is what you see: using eye movements to study infant cross-situational word learning." *Developmental Science* 14.2 (2011): 165-180.

Yurovsky, Daniel, and Chen Yu. "Mutual exclusivity in crosssituational statistical learning." *Proceedings of the 30th annual conference of the Cognitive Science Society*. 2008.