The Pennsylvania State University

The Graduate School

College of the Liberal Arts

THE INFLUENCE OF BILINGUALISM ON STATISTICAL LEARNING WITH MULTIPLE INPUTS

A Dissertation in

Psychology and Language Science

by

Timothy J. Poepsel

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

December 2016

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

Daniel J. Weiss
Associate Professor of Psychology and Linguistics
Dissertation Adviser
Chair of Committee

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

Kristin A. Buss
Professor of Psychology

Matthew T. Carlson
Assistant Professor of Spanish and Linguistics

Melvin M. Mark
Professor of Psychology
Head of the Department of Psychology

*Signatures are on file in the Graduate School.

ABSTRACT

Statistical learning is a fundamental component of language acquisition, yet to date, relatively few studies have examined whether these abilities differ in bilinguals, especially when the learning task presents multiple statistically independent distributions. Furthermore, extant work comparing monolinguals and bilinguals has been undertaken in a single statistical learning paradigm representing only one of many components of the process of language acquisition. In this dissertation, we addressed these gaps in the statistical learning literature by comparing the performance of monolinguals and late-learning bilinguals in two types of cross-situational statistical learning (CSSL) tasks that contained multiple statistical distributions.

In Experiment 1, learners (English monolinguals, English-Spanish bilinguals, and Chinese-English bilinguals) were asked to form both one-to-one and two-to-one word-object mappings, and were tested at three points during training. All groups performed identically on one-to-one mappings, but bilinguals outperformed monolinguals on two-to-one mappings, acquiring these mappings both more quickly and proficiently. In Experiment 2, learners (English monolinguals and English-Spanish bilinguals) were asked to form one-to-one mappings and two-to-one mappings that varied with respect to the amount of evidence supporting either side, and were tested at a single point after training. The average performance of monolinguals and bilinguals on both one-to-one and two-to-one mappings did not differ; however, bilinguals more frequently acquired both sides of two-to-one mappings when evidence for each side was equal, and also showed increased sensitivity to two-to-one mappings at lower thresholds of evidence. In Experiment 3, learners (English monolinguals and English-Spanish bilinguals) completed a speech segmentation task presenting multiple statistically distinct streams that differed in the degree to which their components (i.e., syllable inventories) overlapped. Here, we did not find a difference in the ability of monolinguals to acquire multiple inputs or manage the degree of statistical interference between two inputs, but we did note a significant effect of degree of overlap on learning, so that greater overlap correlated with decreased learning. In accord with previous research, the results of these studies suggest that the fundamental ability to track the statistics of language input may not be affected by bilingualism. However, we found distinct advantages for bilinguals in the acquisition of multiple distributions, and that this effect is modulated by the type of statistical learning task presented to learners. Overall, our results suggest that bilingual experience may impact a learner's response to variability in an input, and that the statistical learning mechanism may be comprised of multiple subcomponents that function and respond differently based on both a learner's experience and task-specific demands.

TABLE OF CONTENTS

List of Tables

List of Figures

Acknowledgements

Many people have contributed their efforts to the completion of this work, and for that l am very grateful. It is undeniable that my advisor, Dan Weiss, contributed most critically, and I attribute both the quality of this work and its very completion to his relentless guidance, patience, and insight. He has instilled in me an appreciation for thoroughness of thought and professionalism that will be invaluable to me throughout my career. Dan, for your incredibly hard work on my behalf, I thank you. The members of my committee – Judy Kroll, Giuli Dussias, Kristin Buss, and Matt Carlson each played an essential role, and their insistence on excellence is reflected in each chapter of this dissertation. I owe a debt to Chip Gerfen, who reinvigorated my passion for language science and allowed me to develop the independence and inquisitiveness at the core of my scientific urge. And were it not for my friends and peers, none of this would have been possible. I thank all of them who stood with me over the years: Cari Bogulski and her whimsical clubs; Jason Gullifer and his pizza parties; Amelia Dietrich and two-day drives to Colorado; my lab mate Federica Bulgarelli on whom I leaned for support and delicious baked goods; Bobby Marker, Nick Ray, and Peiyao Chen, who never failed to provide comic relief; and all of the RAs who have helped me collect and code mountains of data from seldom helpful participants.

My most stalwart friend and supporter over my entire time in State College isn't even human, and has absolutely no appreciation for statistical learning and language acquisition. Nonetheless, her contribution to this work has been enormous. No matter the time of day, no matter the height of my stress, she remained by my side (or rather, on it), reminding me that frequent breaks are an essential part of any healthy work day. Kitty, you will never know this, but I thank you.

My wife, Kaitlyn Litcofsky, is certainly the best and most undeserved thing that will ever happen to me, and her presence in my life has, without question, been the single most important factor in the successful completion of this degree. She handled with otherworldly grace the kaleidoscopic emotions that poured forth from me throughout my final years of graduate school. In the final months leading up to my defense, she continued to exemplify this loving dedication, cooking tasty meals while I typed, surprising me with cold beverages, serving as a one-person audience for my practice talks, implementing edits to my slides when I was too exhausted to do anything other than give dictation, and willingly walking up and down the quiet summer streets of State College whenever my walking muse struck. Her commitment to my work and sanity during this process made all the difference, and I can only hope that I will be able to provide such complete and comforting aid to her whenever necessary. She is both lovely and fantastic, and so I dedicate a healthy portion of this work to her.

The remaining portion I dedicate to my parents. There is no way to fully illustrate their influence, but it can be summed up as unwavering love. The seed of every good quality I have was planted and nurtured by them, and I have not yet found a limit on their commitment to my happiness and development as a complete human being. I am proud to call them my parents, and I hope that I can continue to make them proud in everything I do.

I have left out an incredible number of people from these acknowledgments. To all of those unnamed people who also lent their help, food, shoulders, and friendship during this time of my life I say, over and over again, thank you. This degree cannot be completed through the effort of single person, no matter how gifted or hard-working. I have learned and gained so much from this exceptionally talented community, and I hope to never forget the richness of this

experience. Although I’m excited to enter the next phase of my life, I am sad to be leaving behind such a unique and ever-welcoming place.

Chapter 1: Introduction

1.1 Statistical Learning

Language acquisition is arguably a process that relies on both experience independent and experience dependent mechanisms. That is, humans may be uniquely predisposed to acquire language by virtue of many cognitive and biological factors (Chomsky, 1965; Lenneberg, Chomsky, & Mark, 1967; Pinker, 1994; Wexler & Manzini, 1987). At the same time, humans undoubtedly require exposure to language for full acquisition, as it develops in limited form when exposure is severely limited or delayed, and learners acquire only the languages to which they are exposed (Lenneberg, 1967; Newport and Supalla, 1987; Mayberry & Lock, 2003). A prominent idea is that experience dependent acquisition may proceed via sensitivity to relationships between sounds or other elements of language input, as the reliability of such relationships (or lack thereof) may indirectly signal the presence of structure (Brent & Cartwright, 1996; Chomsky, 1955; Goodsitt, Morgan, & Kuhl, 1993; Harris, 1951; Hayes & Clark, 1970). Statistical learning (SL) has been proposed as an experience-dependent learning mechanism that facilitates the search for such structure by allowing a learner to implicitly track dependencies between events. Over time and with repeated exposure to stimuli, a statistical learner accrues knowledge about dependencies. Thus, within a brief stretch of experience, events that reliably co-occur can be acquired as units of a larger structure (e.g., sounds that form syllables, or syllables that are grouped into a word), while across many experiences, structures that frequently co-occur can be acquired as a coherent system (e.g., sounds and words that are grouped into a language). Evidence that learners can track dependencies between events has been found in both visual and auditory perception (Saffran, Aslin, & Newport, 1996; Fiser & Aslin, 2001), for linguistic and non-linguistic units (Saffran, Johnson, & Aslin 1999; McMurray, Dennhardt, & Struck-Marcell, 2008), in both infants and adults (Saffran et al., 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Saffran, Johnson, & Aslin, & Newport, 1999; Theissen & Saffran, 2003), and across species (e.g., Hauser, Newport, & Aslin, 2000; Newport, Hauser, Spaepen, & Aslin, 2006). Within the domain of language, statistical learning has been shown to facilitate the discovery of phoneme distributions (e.g., Maye, Weiss, & Aslin, 2008), speech segmentation (e.g., Saffran et al., 1996), word mapping (e.g., Yu & Smith, 2007) and syntactic patterns (e.g., Gomez & Gerken, 2000; Perfors, Tenenbaum, & Regier, 2006; Reeder, Newport, & Aslin; Meylan, Kurumada, Borschinger, Johnson & Frank, 2012).

Saffran, Newport, and Aslin (1996) carried out one of the first investigations of learners' ability to find words in an artificial speech stream by tracking the co-occurrence statistics of basic units (e.g., syllables). Infant learners were exposed to nonce words concatenated into a continuous stream, in which the only cues marking word boundaries were adjacent transitional probabilities (TPs) between syllables (i.e., the probability that one syllable would transition to another divided by all occurrences of the initial syllable). Within a statistical word, each syllable perfectly co-occurred with the following syllable (i.e., the transitional probability between syllables was 100%). Because words were randomly concatenated (with the restriction that a given word could not follow itself), the transitional probability from the offset syllable of one word to the onset syllable of another was 33%. If learners could track these statistics, they would be able to distinguish words as units of input over which transitional probabilities were high (i.e., where predictions about upcoming elements could be accurately formed), and word boundaries as points where transitional probabilities were low (i.e, where the accuracy of predictions about upcoming elements were far less predictable). In a test pitting words from the stream against non-words (groupings of three syllables present in the stream that had never co-occurred and whose internal TPs were thus 0) and part-words (groupings of three syllables that had co-occurred in the steam but spanned a word boundary and thus contained a dip in their internal TPs), learners reliably endorsed words over both non-words and part-words. Thus, the authors concluded that statistical learning was a plausible mechanism for solving the early challenges posed by speech segmentation. Subsequent studies have extended these findings to some non-adjacent dependencies (Creel, Newport, & Aslin, 2004; Newport & Aslin, 2004; Gebhart, Aslin, & Newport, 2009) and have attempted to address whether this mechanism can scale to real-world language learning environments. For example, statistical learning can operate on stimuli drawn from actual human languages (in which TPs are less exaggerated relative to typical artificial speech streams; Thiessen, Hill, & Saffran, 2005; Pelucchi, Hay, & Saffran, 2009), as well as in environments containing a greater number of words, variation in the number of syllables each word contains, and across Zipfian word frequency distributions (Kurumada, Meylan, & Frank, 2010; Frank, Tenenbaum & Gibson, 2013; but see Johnson & Tyler, 2010 for evidence to the contrary), confirming the plausibility of statistical learning as a vital learning mechanism for language acquisition..

1.2 Learning, Variability and Bilingualism

To date, a majority of SL studies have investigated the performance of monolingual learners with stimuli that contain a single distribution. Yet, more than half of the world is multilingual (Grosjean & Li, 2013) and their language input is thus derived from two or more underlying distributions. Consequently, there is a pressing need to understand how bilinguals may differ from monolinguals in statistical learning, and how these differences may interact with the number of statistical distributions, as well as the degree of statistical variability, present in the input.

The need to address bilingual performance in statistical learning can be viewed as part of a larger research agenda to understand the cognitive consequences of bilingualism (Bialystok & Martin, 2004; Prior & MacWhinney, 2010; Kroll & Bialystok, 2013). The results of several recent studies indicate that bilinguals may have advantages relative to monolinguals in areas such as executive function (e.g., inhibiting, updating, switching) and conflict monitoring (Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, 2009; Hernandez, 2013; Costa & Sebastian-Galles, 2014; Li, 2014). Additionally, a number of studies demonstrate that bilingual experience is correlated with differences in brain function and structure, and suggest that these changes can occur rapidly (i.e., on the order of days) and with little experience (Petitto, Berens, Kovelman, Dubins, Jasinka, & Shalinsky, 2012; Schlegel, Rudelson, & Tse, 2012; Martensson et al.,2012; for a review, see Li, Legault, & Litcofsky, 2014). The effect of bilingualism on cognition appears to benefit bilinguals of all ages and levels of proficiency, from early balanced bilinguals to late learning and L1 dominant bilinguals (Bialystok, 1986; Galambos & Goldin-Meadows, 1990; Hilchey & Klein, 2011; Tao et al., 2011; Schroeder & Marian, 2012).

Accordingly, a growing body of work has begun to compare the performance of monolinguals and bilinguals across a range of learning tasks to determine how these cognitive differences influence learning outcomes, especially in learning conditions that provide “bilingual-like” input (i.e., multiple inputs; Kovacs & Mehler, 2008; Byers-Heinlein & Werker, 2009; Bartolotti, Marian, Schroeder, & Shook, 2011; Yim & Rudoy, 2013; Bulgarelli & Weiss, 2016). For example, Kovacs and Mehler (2008) familiarized monolingual and bilingual infant learners to two rule-bound patterns (instantiated as auditory stimuli), each of which was associated with a different location for presentation of a visual reward. Monolingual infants were able to learn only one association between pattern and reward location, whereas bilingual infants learned both associations (however, monolinguals did acquire both patterns when provided with a contextual cue). In the domain of word learning, Byers-Heinlein and Werker (2009) conducted a fast-mapping experiment (in which learners were exposed to trials containing a single word-object pair) investigating whether mutual exclusivity (i.e., the expectation that words and objects form exclusive 1:1 mappings) differed as a function of language experience. They found that bilingual and trilingual infants were more likely to violate mutual exclusivity than their monolingual counterparts, an effect corroborated in Houston-Price, Calorghiris, and Raviglione (2010). Although mutual exclusivity has been proposed as an important word learning constraint (which effectively limits the number of possible mappings in a given learning environment), these results suggest that the development of mutual exclusivity may depend on the type of input a learner receives, in accordance with the modeling results of McMurray, Horst and Samuelson (2012; discussed in greater detail below). When the input largely contains mappings that are mutually exclusive, it follows that learners may assume that most or all mappings will follow suit. When the input contains many violations of mutual exclusivity, learners may discard mutual exclusivity or implement the constraint in more judicious form in new learning environments.

Although the empirical results mentioned above indicate that bilinguals may approach learning tasks with different assumptions about the types of input they will encounter, several studies directly comparing monolinguals and bilinguals have reported conflicting evidence of differences in statistical learning between monolinguals and bilinguals. Yim and Rudoy (2013) compared monolinguals and bilinguals (between 5 and 13 years of age) in both visual and auditory statistical learning tasks, and found no evidence that the groups differed in terms of success, nor that learning varied as a function of age. Bogulski (2013) tested English monolinguals and Spanish-English bilinguals in the paradigm of Gebhart et al. (2009), and found that bilinguals' performance was equivalent to monolinguals' in conditions containing a single language as well as multiple languages. Bartolotti, Marian, Schroeder, and Shook (2011) tested statistical learning in monolinguals and bilinguals using Morse code in conditions of low within language interference (in which certain cues were reliably aligned with word boundaries) and high within language interference (in which a set of cues conflicted with respect to signaling word boundaries). In the low-interference condition, bilinguals showed a segmentation advantage relative to monolinguals. This advantage disappeared, however, in the high interference condition, where the authors noted that inhibitory control was a better predictor of performance. This was an unexpected result in light of much research indicating that bilinguals may show advantages in certain areas of cognitive control that relate to inhibition and conflict resolution (e.g., Bialystok, 1999; Costa, Hernández, & Sebastián-Gallés, 2008). Wang and Saffran (2014) found that adult bilingual learners were advantaged relative to monolinguals when tracking an artificial speech stream that contained both predictive syllabic transitional probabilities and tonal cues to word boundaries. The authors attribute bilinguals' increased performance indirectly to bilingualism, suggesting that bilinguals' improved phonological working memory, application of different cue weighting strategies, and improved inhibitory control may have been driving the effect (see Service, Simola, Metsänheimo, & Maury, 2002; Majerus, Poncelet, van der Linden, & Weekes, 2008; Toro, Sebastian-Galles, & Mattys, 2009; Tyler & Cutler, 2009; Bialystok, 1999; Costa, Hernandez, & Sebastian-Galles, 2008). Finally, Nation and McLaughlin (1986) found a bilingual advantage for implicit learning using an artificial grammar-learning task, reporting that multilingual learners were better at acquiring the grammar when they did not explicitly attend to the rules (there was no advantage when they did). In sum, the results in the literature to date do not seem to converge on a firm conclusion regarding how bilingual experience influences statistical learning, nor is there a principled account regarding the types of inputs that elicit performance differences between monolinguals and bilinguals.

Given the variability in these results, along with the findings of Experiment 1 (see below), we have advanced a hypothesis that seems to synthesize these studies. We have proposed that bilingualism does not enhance the basic cognitive mechanisms involved in tracking statistical relationships from the input. Rather, the increased variability encountered by bilinguals, both from within individual languages, as well cross-linguistically, may fundamentally alter the assumptions regarding the number of causal models that generate newly encountered patterns. This hypothesis may accord with the recently proposed extraction and integration framework for statistical learning (Theissen, Kronstein, & Hufnagel, 2013). In this framework, statistical learning is composed of two different but related processes. Extraction is the binding together of smaller chunks of an input through attention to conditional probabilities, while integration is processing across these chunks to assess central tendency or distributional properties. One application of this framework to the question of how monolinguals and bilinguals may differ in statistical learning works as follows: while these groups may not differ in their ability to extract chunks from an input (i.e., perform the basic task of statistical learning) bilinguals may have an integration advantage, such that they are better able to determine the higher-level properties of individual inputs and similarities between multiple inputs (Erickson & Thiessen, 2015).

The learning to learn framework of behavioral learning asserts that exposure to frequent variability in rules or structure may change the way a learner approaches the task of learning. Several studies demonstrate that learners (typically rats) are able to adapt quickly to frequent reversals in reward rates associated with specific actions (e.g., pushing one of several levers)

(Krechevsky, 1932; Dufort, Gutman, & Kimble, 1954; Gallistel et al., 2001). Learners have been shown to achieve single error reversal proficiency (i.e., adaptation to a change in reward rate with only a single exposure to the changed rate) within ten such reversals (Williams, 1968). These results, found across a number of species, suggest that experience with frequent variability in the underlying reward structure of an environment influences the rate at which subsequent variability in structure is accommodated. In a direct test of this hypothesis, Gallistel et al. (2001) exposed rats to input containing frequent reversals of reward rates as well as long periods of rule stability. Their findings confirmed the above hypothesis: animals that had recently experienced frequent change more quickly accommodated new changes relative to animals that had recently experienced a period of rule stability. These findings have been thought to reflect basic tenets of learning and thus can be extended to human learners as well. Fine, Jaeger, Farmer, & Qian (2013), for instance, showed that learners' syntactic processing rapidly accommodates variability when they are frequently exposed to sentences containing syntactic ambiguities. The opposite was also true. When learners did not frequently encounter such ambiguities, they had difficulty accommodating unexpected syntactic structures. Frequent exposure to two languages, as well as the need to manage competition arising from nonselective activation (i.e., the idea that both of a bilingual's languages are always active and competing for selection; De Groot, Delmaar, & Lupker, 2000; Van Hell & Dijkstra, 2002; Dijkstra, 2003; Kroll & Dijkstra, 2000), may also work to bias a learner toward the expectation of non-uniform inputs. Accordingly, this may allow the learner to more readily view variability as a change in structure. In sum, the critical difference between monolinguals and bilinguals may be in the assumptions of how many causal models underlie the surface (i.e., observed) statistics. This hypothesis predicts monolinguals and bilinguals will perform identically in the basic task of tracking statistics, but the groups may differ with respect to how they treat variability in the input. Specifically, monolinguals may be more likely to assume the relationships in their input to arise from a single causal model, whereas bilinguals may have an increased expectation that encountering new structure and thus greater openness to finding and interpreting variability (i.e., resolving conflicts in the statistics) as arising from changes to the causal model itself.

The desirable difficulty framework suggests that difficulties in learning and costs placed on processing may actually provide benefits to learners rather than hindering them. Findings from a number of studies (see below) provide evidence that responding to difficulty engages the processes that support learning (Bjork & Kroll, 2015). Examples of such difficulties include variability in training conditions, mixed training blocks, the distribution of training and testing across several periods rather than massed training and testing, and the use of tests and explicit feedback to assess learning rather than restudying (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1996; Glenberg, 1979; Hintzman, 1974; Melton, 1970; Soderstrom & Bjork 2015; Kerr & Booth, 1978). In relation to bilingualism, these findings are suggestive of the mechanisms by which bilinguals show advantages compared to monolinguals across a variety of cognitive tasks. Specifically, the literature on bilingualism demonstrates that each of a bilingual's languages is always active and competing in processes of comprehension and production, necessitating constant regulation of the languages and resolution of any conflicts that arise (e.g., Abutalebi et al., 2012; Kroll, Dussias, Bogulski, & Valdes-Kroff, 2012; Gold, Kim, Johnson, Kriscio, & Smith, 2013). Additionally, for the majority of bilinguals who are unbalanced, processing in the L2 requires effortful suppression of the dominant language. Bilingualism thus results in a state of constant difficulty for the learner, which may provide domain general benefits to cognition, especially in areas related to executive function, as noted above (Abutalebi et al, 2012; Kroll, Bobb, & Hoshino, 2014). Statistical learning has been shown to recruit resources related to executive function such as inhibitory control, ambiguity resolution, and working memory (Bartolotti & Marian, 2012; Umemoto, Scolari, Vogel, & Awh, 2010). As such, it follows that bilingualism itself may influence outcomes in statistical learning tasks.

1.3 Statistical Learning of Multiple Inputs

A growing body of SL research has begun to investigate how learners accommodate structural variability, given that real-world learning environments are typically variable (Qian, Jaeger, & Aslin, 2012). Outside the domain of language, for instance, patterns of traffic in a restaurant or cafeteria may change in relation to the time of day (i.e., breakfast, lunch, and dinner traffic) or time of year (i.e., seasonal demand for specific items), such that successfully learning the relationships depends on correctly determining and accommodating the points of transition between the underlying causal models (Qian et al., 2012). Because inputs unfold incrementally, quickly orienting to the emergence of a new input is an essential step in the process of learning. That is, failure to notice a transition may result in learners collapsing together patterns of inputs that are structurally incongruent, which may have deleterious effects on overall learning (French, 1999; Weiss et al., 2009, 2010; Qian et al., 2012). Within the domain of language, understanding how learners accommodate variability is particularly important in the case of bilinguals, who face a number of unique challenges such as acquiring the rules and perceptual patterns of two languages and forming distinct language representations despite receiving less input overall in each language relative to monolinguals (Weiss, Poepsel, & Gerfen, 2015; Gollan, Montoya, Cera, & Sandoval, 2008).

Several recent SL studies have begun to investigate how learning proceeds in tasks presenting multiple underlying inputs. These studies provide a foothold for understanding how learners accommodate multilingual input. Gebhart, Aslin, and Newport (2009) exposed adult learners sequentially to two artificial language streams with no pause between the streams, and with transitional probabilities as the only cue to underlying structure. When the two streams were presented in the same voice, learners showed a primacy effect, performing above chance at test on only the first of the two streams to which they had been familiarized. When the duration of the second stream was tripled relative to the first, and also when the two streams were differentiated with an explicit contextual cue (i.e., a pause as well as explicit mention of the presence of a second stream), learners were able to perform above chance at test on both streams. Weiss, Gerfen, and Mitchel (2009) presented learners with two artificial speech streams that were interleaved in twelve two-minute segments. When the statistics of each language were congruent, such that word boundaries were clearly marked by dips in TPs, both streams were learnable regardless of whether or not the streams were differentiated by a contextual cue (i.e., a change in speaker voice). However, when the statistics of each language were incongruent, such that combining their statistics increased statistical noise, the languages were only learnable when differentiated by a contextual cue. Franco, Cleeremans, and Destrebecqz (2011) also presented learners with two sequential artificial languages that fully shared a syllable inventory, finding that while both languages were learned above chance, learners still showed a significant primacy effect. The authors concluded that sequential learning of inputs was possible, but that shared syllabic structure may make learning of the second language more difficult after learning of the first. Mitchel and Weiss (2010), used the incongruent languages from Weiss et al. (2009) to demonstrate that visual contextual cues could also facilitate the acquisition of multiple structures. Learners were familiarized to each language while simultaneously watching a visual display of a talking face. When the same face was presented with each stream, the languages were unlearnable; however, when a different face was synchronized to each stream, both were learned above chance. In another statistical learning paradigm, Conway and Christiansen (2006) familiarized learners to either visual or auditory sequences generated from two separate grammars. When both grammars from a single domain (either auditory or visual) were instantiated in, for instance, shape sequences or tone sequences, learners only acquired sensitivity to the first of the two grammars. When the grammars within a domain were instantiated with distinct element sets (i.e., Visual Grammar 1 presented as a sequence of shapes, Visual Grammar 2 presented as a sequence of colors; Auditory Grammar 1 presented as a sequence of tones, Auditory Grammar 2 presented as a sequence of nonce words), learners were able to acquire both grammatical patterns.

The results of these studies suggest that success in the learning of multiple structures depends on a number of task-specific factors. As in Weiss et al., (2009), when structures are statistically congruent, learners are able to acquire both above the level of chance. When statistics are incongruent, however, learners may require additional information in order to determine that the underlying structure has changed (Weiss et al., 2009; Gebhart et al., 2009; Mitchel & Weiss, 2010), especially in the case that a learner has received extended exposure to the first structure and, consequently, reduced their attention to the input (e.g., Bulgarelli & Weiss, 2016). Notably, this additional information can take the form of more exposure to a second structure (Gebhart et al., 2009) or the addition of contextual cues that reliably mark and differentiate each structure (Gebhart et al., 2009; Weiss et al., 2009), such that learners redirect their attention to the point of transition and begin sampling the available structure again (Eramudugolla, Irvine, McAnally, Martin, & Mattingley, 2005; Qian et al., 2012). A number of other statistical learning studies demonstrate that in the absence of attention to an input, learning does not occur (Toro, Sinnett & Soto Faraco, 2005; Turk-Browne et al., 2005; but see Saffran, Aslin, Newport, Tunick, & Barreuco, 1997 for evidence to the contrary), providing additional support for the notion that once attention to an input has waned, a learner who is not expecting a structural change may interpret any variability as arising from within their initial representation of the input (Qian et al., 2012). This pattern of entrenchment in a first representation may continue until the learner notices the change in the underlying structure via the means discussed above. Alternatively, rapidly transitioning to a second structure following robust learning of the first can attenuate entrenchment effects (Bulgarelli & Weiss, 2016). This suggests that overlearning one structure may have detrimental effects on subsequent learning, perhaps by implicitly cueing the expectation of a statistically stable learning environment.

As noted above, many studies of both animals and humans suggest that the experience of frequent transitions between structures has different effects on a learner compared to periods of structural stability (Gallistel et al., 2001; Fine et al., 2010). The findings of Zinszer and Weiss (2014) further corroborate this phenomenon in the domain of SL. Using the stimuli of Gebhart et al. (2009), learners were exposed to multiple artificial languages in sequence. When there was only one transition between the languages, the authors replicated the primacy effect of Gebhart et al., (2009). In subsequent conditions that contained multiple transitions between languages, the authors found that learners were able to acquire both inputs. Further evidence for the assertion that the degree of variability in an input can serve to highlight stable structure comes from Gomez (2002) who exposed learners to three element sequences in which the dependency was between the first and third elements (i.e., a non-adjacent dependency). She found that when the number of intervening elements between first and last positions was large, learning of the non-adjacent dependencies was facilitated relative to fewer intervening elements. Thus, the uncorrelated nature of the intervening element served as a cue to learners regarding where structure was located. Taken together, the results across these SL paradigms suggest that variability in an input, instantiated in many different ways, may serve to highlight stable relationships, as well as facilitate change in the way that learners interpret the underlying structure of the input.

1.4 Cross-Situational Word Learning

The discussion of statistical learning to this point has focused largely on the tracking of statistics within speech segmentation tasks. Yet, during the course of language acquisition, many types of statistics are informative. One area of interest is word learning, as solving the word-world mapping problem is challenging since a given word can theoretically be mapped to any object, concept or place (Quine, 1960). Further, learners confront significant variability in mappings in both monolingual (e.g., homophones and homographs) and bilingual environments (e.g., translation equivalents and interlingual homographs). Among many other strategies, it has been proposed that learners may be aided in solving the mapping problem by a form of statistical learning in which they track the co-occurrence of words and objects across many different environments (Siskind, 1996; Yu & Smith, 2007; Smith & Yu, 2008; Vlach & Sandhofer, 2014).

In the cross-situational statistical word learning (CSSL) paradigm, learners are typically exposed to scenes in which arrays of objects are displayed and a corresponding number of labels are played (Yu & Smith, 2007). Words and objects are paired such that whenever an object is displayed in a scene, its label is also presented. Within any given scene, however, there is local mapping ambiguity. Words are presented in a random order with respect to the location of objects on the screen, and therefore learners cannot be sure which objects and words form discrete pairs. Across many scenes, though, the same words and objects co-occur reliably and frequently while spurious co-occurrences are unreliable and infrequent. A learner who tracks co-occurrence statistics across scenes is thus able to discover the stable word-object pairings. To date, many studies using the CSSL paradigm have found successful learning for both infants and adults (Yu & Smith, 2007; Smith & Yu, 2008; Fazly, Alishahi, & Stevensen, 2010; Fitneva & Christiansen, 2011; Kachergis, Yu & Shiffrin, 2009).

Although learners arguably use this form of statistical learning to map words to objects (for another proposal of word learning, see Medina, Snedeker, Trueswell, & Gleitman, 2011; also discussed in greater detail below), there are numerous other strategies that can be applied in tandem to reduce the complexity of the mapping problem. For example, learners may be influenced by a preference for labeling whole-objects (Markman, 1991), by social-pragmatic constraints (e.g., Baron-Cohen, 1995; Clark, 1987; Diesendruck & Markson, 2001; Tomasello & Barton, 1994), and by a preference for assigning novel labels to novel objects, termed “mutual exclusivity” (Markman & Wachtel, 1988). The mutual exclusivity (ME) constraint, at least conceptually, may bear some similarity to primacy effects found in the speech segmentation studies; specifically, once learners have formed a first association between a word and an object, evidence for secondary mappings between that word and other objects (and vice versa) may not lead to the formation of new, or multiple, mappings. This, in fact, is the result found in Ichinco, Frank, and Saxe (2008; and subsequently replicated in Poepsel, Gerfen, & Weiss, 2012; Poepsel & Weiss, 2014); learners were familiarized first to one set of word-object mappings, and in a second learning phase, a subset of these words and objects were given new mappings. Performance on these new mappings was not significantly above chance. Using a similar paradigm, Poepsel et al., (2012) probed the influence of contextual cues on statistical word learning. As in Ichinco et al., (2008), when the two learning phases were presented in the same voice (and otherwise undifferentiated), acquisition of the secondary mappings was blocked. However, when contextual cues were added, such that the voice in which each familiarization was presented varied, learners were significantly better at forming two-to-one mappings. This result accords with findings from the SL literature, in which changes in voice and other contextual cues allow learners to overcome the primacy bias and acquire a second input after exposure to a first (Gebhart et al., 2009; Weiss et al., 2009).

Interestingly, Yurovsky, Yu, and Smith (2013) demonstrated that when learners are exposed to both single and multiple mappings within a single CSSL learning phase, they are able to acquire both mapping types. Further, Yurovsky et al., (2013) showed that when initially presented with a mixture of mappings that either conform to or violate mutual exclusivity, learners may relax their reliance on ME in order to acquire all available mappings. However, when learners are familiarized only to mutually exclusive mappings, they may preferentially apply mutual exclusivity even in situations where multiple mappings are available. This suggests that the amount of mapping variability in the input may influence how learners approach the mapping problem. This logic may extend to bilingualism. That is, this finding is consistent with fast mapping and modeling studies showing that bilinguals, in comparison to monolinguals, rely less on the assumption of mutual exclusivity in the context of many-to-one word mappings (Kachergis et al., 2009; Byers-Heinlein & Werker, 2009; Houston-Price et al., 2010). As noted above, fast-mapping studies, have demonstrated that the application of mutual exclusivity (i.e., the expectation that words and objects form exclusive 1:1 mappings) differs as a function of language experience (Byers-Heinlein and Werker, 2009; Houston-Price et al.,2010). In a comparison of infant monolinguals, bilinguals and trilinguals, multilinguals were less bound by mutual exclusivity than monolinguals, and trilinguals were less reliant on ME than bilinguals

(Byers-Heinlein and Werker, 2009). This demonstrates that language experience influences learning in a rather straightforward way; greater experience with multiple languages results in less reliance on a constraint that arguably could hinder acquisition. Together, these findings support the hypothesis, in accordance with the learning principles from both the learning to learn and desirable difficulty literature, that individuals who have different experiences with variability (such as monolinguals and bilinguals) may show differences in performance when periods of structural stability in a learning environment are followed abruptly (and without cues) by periods of structural variability.

1.5 The Current Investigation

While studies of statistical learning have largely focused on monolinguals' sensitivity to stable relationships in the input, the research summarized above suggests that learners may vary in their performance as a function of language experience. The evidence seems to indicate that learners who frequently encounter variability, such as bilinguals, demonstrate processing differences across a range of cognitive tasks relative monolinguals. Thus, bilinguals may be better positioned to take advantage of variability that is correlated with structural changes in SL tasks. Nonetheless, the existing SL studies comparing monolinguals and bilinguals have a number of limitations that may contribute to the mixed outcomes from the previous studies reviewed above. This dissertation aims to address a number of these limitations and provide new evidence regarding how bilingual experience influences learning in SL tasks.

One limitation of the previous research is that the effect of bilingualism on SL has only been investigated within a single paradigm (segmentation of a streaming input; e.g., Yim and Rudoy, 2013; Bartolotti, et al., 2011; Wang & Saffran, 2014). Yet, as noted, SL plays a role in learning at many levels of acquisition, and the learning demands are far from uniform.

Consequently, there have been several recent attempts at providing models of SL that posit multiple, partially overlapping cognitive functions (Hsu & Bishop, 2011; Erickson & Thiessen, 2015, Frost et al., 2015). The extraction and integration model (described in detail above), for instance, proposes that SL is composed of two processes: the ability to track conditional statistics and form coherent chunks (extraction), and the ability to perform analyses across all of the extracted chunks and deduce distributional properties (integration; e.g., Thiessen and Erickson., 2013; Erickson & Thiessen, 2015). Speech segmentation tasks are characterized primarily as extraction tasks (Thiessen & Erickson, 2013), while cross-situational word learning tasks are characterized as integration tasks (in that learning requires the tracking and merging of information across multiple scenes – Erickson & Thiessen, 2015). The comparison of monolinguals and bilinguals in these tasks, which may differ in the cognitive processes they entail, provides an opportunity to test such models, and to determine if bilingualism differentially affects the processes underlying SL.

A second limitation of previous studies is that a majority present only a single input distribution to learners. Although the SL mechanism and its deployment may vary slightly between learners (e.g., Frost et al., 2015), there is little theoretical support for the claim that the learning of a single input, which is a task faced by all learners, should vary based on bilingual experience. Since even late bilinguals face are faced with the challenge of resolving additional ambiguity in their learning environments (such as multiple mappings corresponding to the labels generated by each language, or keeping track of how sounds pattern across languages, and so inferring the most salient cues to word boundaries in an L2), we hypothesize that this may influence their statistical learning abilities when the input affords the opportunity to form multiple statistical mappings. As a result of their experience encountering and accommodating variability, bilinguals may be more likely to assume that there are multiple causal models generating the surface statistics in a given situation, and so learning in such situations may be facilitated.

Consequently, the goal of the present dissertation is to compare monolingual and bilingual learners across multiple types of statistical learning tasks that present multiple inputs to learners. Specifically, in Experiments 1 and 2, we extend this work to cross-situational statistical word learning tasks, in which monolinguals and bilinguals have never been compared (even in the case of learning single mappings). Further, in Experiment 3, weadopt this approach with a speech segmentation task, augmenting and extending the few existing comparisons of these groups in the task of acquiring multiple inputs. Thus, within the context of both word mapping and speech segmentation tasks, this series of experiments (outlined below) can begin to test a specific prediction: that differences in the performance of monolinguals and bilinguals in statistical learning tasks may arise in more variable environments that contain multiple statistically distinct distributions. As noted above, this prediction of differences is also directional, such that we expect the cognitive consequences of bilingualism to result in improved performance for bilinguals when tasked with acquiring multiple inputs in variable learning environments. Further, recent work supports the hypothesis that statistical learning may have modality specific constraints (e.g., Frost et al., 2015), as well as recruiting different cognitive resources based on the specific task (e.g., Hsu & Bishop, 2011). Thus, by investigating two types of statistical learning tasks, an additional question that this dissertation can address is whether there is consistency across tasks in the manifestation of learning differences across monolinguals and bilinguals

An important consideration for testing the hypothesis outlined above is the type of bilinguals that will be compared to monolinguals. To date, a majority of studies comparing monolinguals to bilinguals have focused on early and high proficiency bilinguals (Kovacs & Mehler, 2008; Bartolotti et al. 2011; Yim & Rudoy, 2013; Nation & McLaughlin, 1986; Bogulski, 2013; Kalia, Wilbourn, & Ghio, 2014). As the results of several studies mentioned above suggest, the input that early bilinguals receive may alter the types of learning strategies they employ (e.g., early bilingualism may slow or inhibit the development of mutual exclusivity), such that comparing early bilinguals to monolinguals may introduce a number of confounds in relation to causation (e.g., Linck, Kroll, & Sunderman, 2009; Luk, De Sa, & Bialystok, 2011; Hernandez, Hoffman, & Kotz, 2007; Hull & Vaid, 2007; Mechelli et al., 2004). For instance, Hull & Vaid (2007) presents a meta-analysis of brain lateralization studies showing that early vs late bilinguals show different patterns of brain activation in response to a variety of behavioral tasks, suggesting that these two groups of learners may engage in fundamentally different processing to achieve similar ends. Mechelli et al. (2004) provides additional evidence that age of acquisition has differential effects on brain structure and function, showing that early vs late learners have a greater degree of structural reorganization (i.e., changes in grey matter density) in language-related brain regions (e.g., left inferior parietal cortex). Late learners, whose early developmental experiences mirror those of monolinguals, may provide a more rigorous test of whether bilingualism influences how learners approach multiple inputs in an SL task, specifically addressing the question of whether the process of becoming bilingual engenders differences in cognition at any age or level of experience (e.g., Pelham & Abrams, 2013). Consequently, this dissertation aims to examine the influence of bilingualism on statistical learning by comparing late learning and less proficient bilinguals to monolinguals across a range of statistical learning tasks.

This dissertation comprises three sets of experiments comparing monolinguals and bilinguals on different aspects of statistical learning in the context of multiple inputs. Experiment 1 explores the prediction that bilinguals should be faster to identify the presence of multiple mappings in a cross-situational statistical word learning task. Several recent studies of word learning examine learning in environments that contain multiple mappings (Ichinco et al., 2008; Poepsel et al., 2012; Yurovsky et al., 2013); however, the insights from these studies come only from samples of monolingual speakers, and probe learning at only a single point (after the familiarization period has been completed). Thus, in Experiment 1, I advance the word learning literature by comparing monolingual and bilingual learners in a cross-situational statistical word learning task that presents both stable (one-to-one) and variable (two-to-one) mappings to learners. An additional difference between this task and the extant literature on cross-situational learning is that performance is probed at three different points in time. This feature of the experiment allows for a more nuanced investigation of how the formation of multiple mappings unfolds over time, as well as a more precise account of how previous language experience influences the trajectory of learning for such mappings.

Experiments 2 asks whether bilinguals show greater sensitivity to weak or noisy evidence for the presence of multiple inputs relative to monolinguals. Real-world learning environments typically present learners with variability in the amount of evidence for a given structure or relationship. Yet, cross-situational statistical word learning studies (e.g.,Yu and Smith, 2007), typically do not represent such variability (all mappings, both 1:1 and 2:1, are presented an equal number of times across trials, and critically, each side of a 2:1 mapping has equal supporting evidence; e.g., Yurvosky et al., 2013; Ichinco et al., 2008; Poepsel, Gerfen, & Weiss, 2012). In Experiment 2, I investigate how variability in the amount of evidence that learners encounter for many-to-one word mappings (i.e., unequal evidence for each side of a 2:1 mapping, such that some mappings are high strength and some are low strength) influences learning. To date, this manipulation has only been implemented in a fast-mapping paradigm (e.g., Vouloumanos, 2008). Experiment 2 will thus contribute to an understanding of how statistical learners accommodate mapping variability that more reflective of the unbalanced input learners receive in real-world learning environments. I also investigate how the learning of mappings that vary in strength interacts with previous language experience. Specifically, I ask how bilingual experience modulates the threshold of evidence required for acquisition of two-to-one mappings, with the prediction (on the basis of the results of Experiment 1) that bilingual learners should outperform monolinguals on low strength many-to-one mappings as a result of their greater experience accommodating variability in language learning.

Experiment 3 aims to fill a number of existing gaps in the auditory statistical learning literature. First, in multi-stream speech segmentation experiments, the degree to which languages overlap in their sound inventories (and accordingly how congruent or incongruent the combined statistics of the language are) varies unsystematically across studies. Given that many studies on multi-stream learning to date use the stimuli of Gebhart et al. (2009) that overlap by 50% in their syllable inventory, and that other studies containing multiple inputs do not systematically vary the degree of overlap, a question about how the degree of language overlap influences learning in such paradigms logically arises. In Experiment 3, therefore, I vary the degree of overlap between languages from 0% of their syllabic inventory to 100% in steps of 25% in order to investigate how variability in the degree of similarity of two languages in a speech segmentation task influences learning outcomes. At low levels of overlap (0%, in which the two languages do not share any syllables), detecting a transition may pose a lesser challenge to learners (since the change in phonetic inventories itself becomes a cue to the change in structure), while at high levels of overlap transitions should be more difficult to detect, as transitional probabilities collapsed across languages may become considerably noisy (i.e., if statistics are combined across languages, within and between word transitional probabilities will converge and become less informative with respect to the location of potential word boundaries). This manipulation will critically contribute to an understanding of how similarity between languages, as well as statistical congruence and incongruence, influences the degree to which learners are able to acquire sequential inputs. Experiment 3 will also investigate whether bilingualism modulates a learner's response to language overlap and statistical incongruence at higher levels of overlap, with the prediction that bilingual leaners, as a result of increased expectation of changes between inputs and improved conflict resolution, should find the noisier statistics less difficult to accommodate. Although SL studies comparing monolinguals' and bilinguals' learning of multiple inputs have reported mixed effects to date (see below), Experiment 3 represents the first auditory SL study that carefully controls how sequentially presented languages overlap, and so may be better suited to capturing variability in learning between monolinguals and bilinguals.

Chapter 2: The Influence of Bilingualism on Cross-Situational Learning of Multiple Inputs

Statistical learning can be described as the process of detecting structure by monitoring distributional information available in the sensory input. For the past two decades, research on statistical learning has had a dramatic impact on our understanding of language acquisition. Yet despite many advances in this line of inquiry, very few investigations have approached this problem from the perspective of bilingualism. In order to acquire two languages, bilinguals must be able to establish and maintain multiple statistical representations. This experience could influence how bilinguals approach new statistical information (see Weiss, Gerfen, & Poepsel, 2015). Consequently, in Experiment 1 we endeavor to explore whether there are consequences of bilingualism for statistical word learning.

As noted in the Introduction, only a handful of studies have compared statistical learning in bilinguals relative to monolingual abilities, and so far the results have been mixed. Yim and Rudoy (2013) tested monolingual and sequential bilingual children between 5 and 13 years of age on a nonlinguistic auditory tones task, as well as a visual statistical learning task. There was no advantage for bilinguals on either task as learning was equivalent across both groups. This finding suggests that the most fundamental sequential statistical learning abilities may not be influenced by multi-language exposure. By contrast, Wang and Saffran (2014) found that adult bilingual learners were advantaged relative to monolinguals when tracking an artificial speech stream that contained compatible syllabic transitional probabilities and tonal cues to word boundaries. The authors note that the tones appear to have increased the difficulty of the segmentation task rather than simplified it, and therefore may have required suppression in order to successfully segment the stream. This conjecture accords with the observation that bilinguals who are not proficient in a tonal language outperformed Chinese monolinguals on this task. Bartolotti, Marian, Schroeder and Shook (2011) presented participants with a statistical learning task using International Morse Code. Participants listened to two Morse Code languages in the context of either a high or low interference condition (a competing pause cue conflicted with the statistics in one condition and reinforced it in the other; see also Weiss, Gerfen & Mitchel, 2010). Bilingual experience improved performance in the low interference condition, and inhibitory control correlated with improved learning when interference was high. The authors suggest that the improvement shown by bilingual learners may stem from a bilingual advantage in phonological working memory (e.g., Majerus, et. al, 2008; see also Misyak & Christensen, 2007). Similarly, Nation and McLaughlin (1986) found a bilingual advantage for implicit learning in an artificial grammar-learning task. They found that bilingual learners were better able to acquire the grammar when they did not explicitly attend to the rules, and no advantage when they did. To summarize, the differences between monolinguals and bilinguals reported so far for statistical learning have been mixed, and often quite nuanced. The goal in the present experiment was to extend this literature in two ways: first, by comparing functionally monolingual and bilingual performance in a new domain of inquiry, namely statistical word learning; and second, by providing learners with the opportunity to acquire multiple sets of statistics, a situation that may mirror the real-world challenges confronting bilinguals.

2.1.2 Statistical Word Learning

A primary challenge for learning words is mapping them to their correct referents. This task is complex because words can potentially refer to any object, feature, or event in an environment (e.g., Quine, 1960). Accordingly, a prominent suggestion in the literature has been that learners may be constrained in the types of word-object mappings that they will consider.

For example, it has been proposed that language learners may have a preference for assigning novel labels to novel objects (Markman & Wachtel, 1988), a preference for labeling whole-objects (Markman, 1991) and may also be limited by social-pragmatic constraints (e.g., Baron-Cohen, 1995; Clark, 1987; Diesendruck & Markson, 2001; Tomasello & Barton, 1994). However, constraining the problem space is not the only tool for word learners to alleviate the word-world mapping problem. Statistical learning has recently been proposed as another mechanism that helps learners overcome the challenge of indeterminacy (e.g., Yu & Smith, 2007). Word meanings may seem ambiguous in the context of one learning environment, yet if learners can aggregate information across multiple environments then statistical information (such as co-occurrence probabilities) may help them disambiguate which words belong with which objects.

This idea was modeled using a cross-situational statistical learning (CSSL) paradigm introduced by Yu and Smith (2007). In their initial study, participants were shown multiple scenes in which two to four objects were displayed on a computer screen while their corresponding labels were played in random order (note that the location of an object on the screen was not related to the position of its label in the auditory stream). Due to this randomization, learners could only assign words to their objects by aggregating information across multiple scenes. That is, since words and objects appeared multiple times in different visual and auditory contexts throughout familiarization (i.e., with different non-target objects and thus with different sets of labels), learners could infer that the most frequently and reliably co-occurring words and objects cohered as pairings. This task has yielded successful learning by both adult and child learners (Fazly, Alishahi & Stevensen, 2010; Fitneva & Christiansen, 2011;

Kachergis, Yu & Shiffrin, 2009; Smith & Yu, 2008; Vlach & Sandhofer, 2014; Yu & Smith, 2007).

It should be noted that there has been considerable debate as to whether learning in this task is best described by statistical accumulation of multiple label-object pairings across trials (e.g., Vouloumanos, 2008; McMurray, Horst, & Samuelson, 2012; Yurovsky, Fricker, Yu, & Smith, 2014) or by forming hypotheses related to individual referents (e.g., Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell et al., 2013). One possibility is that task difficulty might determine which strategies learners adopt, as many of the aforementioned studies use different experimental paradigms (see Yurovsky & Frank, in review). While this debate is outside the scope of the present study, it seems that the modified procedures employed here are most consistent with studies that are thought to rely on statistical accumulation rather than hypothesis-testing (e.g., see Yu & Smith, 2012).

2.1.3 Bilingual Word Learning

For bilinguals, the challenges of word learning are compounded by multiple mappings. These can take the form of translation equivalents (e.g., learners must realize that 'dog' and 'chien' both describe a four-legged pet canine) as well as interlingual homographs (i.e., "false friends", such as the word 'tuna' which refers to a fish in English and a pear in Spanish). While monolingual learners are also confronted with similar challenges in the form of synonymy and polysemy, for bilinguals such multiple mappings are compounded as they are encountered both within each language as well as across languages. Since at least half of the world's population is bilingual, an important question for word-learning research is how learners accommodate bilingual input which routinely violates assumptions of mutual exclusivity (Byers-Heinlein & Werker, 2009; Grosjean, 2008; Grosjean, 2010; Marian & Shook, 2012). One possibility is that bilingual learners are not constrained in the same manner as monolinguals when approaching the word-learning situation. In that vein, a number of recent word-learning studies suggest that the extent to which mutual exclusivity develops may depend on the input that a learner receives. For example, in a study with monolingual, bilingual and trilingual infants, Byers-Heinlein and Werker (2009) demonstrated that 17-18 month-old infants with exposure to multiple languages showed less disambiguation in the context of many-to-one word mappings. Furthermore, this effect was greater for trilinguals than bilinguals, suggesting that increased exposure to language variation predicts less reliance on an assumption of mutual exclusivity in mapping. Houston-Price, Caloghiris and Raviglione (2010) noted a similar finding in a study with monolingual and bilingual infants using a broader age range (17-22 months). These results are consistent with the computational modeling efforts of McMurray, Horst, and Samuelson (2012). In their model, the development of a mutual exclusivity preference crucially depends on how many translation equivalents are encountered. An important issue for the present experiment is that, to the best of our knowledge, the studies suggesting bilinguals may relax the mutual exclusivity constraint have focused on early or simultaneous bilinguals, and thus it is unknown whether later exposure to a second language might similarly impact learning style.

In the broadest sense, the relaxation of the mutual exclusivity constraint by early bilinguals can be understood within the framework of "learning to learn", a concept that dates back to the early behavioral learning literature. Several discrimination learning studies have demonstrated that when learners (in these studies, rats) receive repeated reversal training, they are more likely to reverse their choice when they encounter a new reversal (Krechevsky, 1932; Dufort, Gutman & Kimble, 1954; Williams, 1988; summarized in Gallistel, Mark, King & Latham, 2001). More recently, Gallistel and colleagues (2001) extended these findings by testing how learners adapt to variability in reward rates and found that the frequency of change in the environment was strongly predictive of the adaptation rate. That is, the learners that experienced more frequent change were able to accommodate change faster than those who experienced less frequent change. Thus, at a very fundamental level, it can be argued that developing a prior expectation for change in a learning environment may enhance the ability to detect changes in new environments (see Qian, Jaeger, & Aslin, 2012 for further discussion of this topic).

In Experiment 1, we investigated whether the statistical learning mechanisms that facilitate word learning might similarly be impacted by the nature of the input to learners. Specifically, we sought to determine whether late bilingual learners perform differently than monolinguals in the cross-situational statistical learning paradigm. Since even late bilinguals contend with an added layer of variability in their mappings (corresponding to the labels generated by each language), we hypothesized that this may impact their statistical learning abilities. In particular, we were interested in exploring this phenomenon when the input affords the learner an opportunity to form multiple statistical mappings, such as when multiple objects could be mapped to a single word. Bilinguals may be more likely to assume that there are multiple causal models generating the surface statistics. To the best of our knowledge, this notion has yet to be formally tested in the context of statistical learning. Consequently, we approached this problem by comparing a group of functional monolinguals with two groups of sequential bilinguals (Chinese-English and English-Spanish) who acquired their L2 subsequent to mastering their L1. This provided a rigorous test of whether proficiency with a second language could impact statistical learning even in the absence of early learning experiences with two languages that have been shown to result in a relaxation of the mutual exclusivity constraint (e.g., Byers-

Heinlein & Werker, 2009; Houston-Price et al., 2010). We first explored whether there are differences between functional monolinguals and bilinguals on cross-situational learning in the context of one-to-one mappings (Experiment 1a). Next, we tested whether these groups differed when the input afforded two-to-one mappings for a subset of objects and labels (Experiment 1b).

2.2.1 Experiment 1a

In Experiment 1a we explore whether functionally monolingual learners, Chinese-English bilinguals, and English-Spanish bilinguals might differ in their abilities to track statistical information across scenes in a CSSL task. There were three conditions that varied in the number of items presented simultaneously (ranging between two and four).

2.2.2 Participants

Seventeen students (11 female, 6 male; mean age 19.7 years, SD=1.4) at Penn State University were given course credit for their participation in this experiment. Based on language history questionnaire (LHQ) data, these participants were native speakers of English who self-rated their English proficiency at an average of 9.6 (SD = .79) on a 10-point scale, on which 10 indicated maximum proficiency. Due to a foreign language requirement at Penn State University for undergraduates, all participants indicated exposure to a second language in the course of their education. Participants self-rated their second-language proficiency at an average of 1.2 (SD = 1.4) on the ten-point scale, and all participants rated below a 4. As such, we considered these participants to be functionally monolingual.

Sixteen Chinese-English bilinguals (14 female, 2 male; mean age 22.2 years, SD = 1.4) from Beijing Normal University in Beijing, China also participated for payment. Participants self-rated their Mandarin proficiency at an average of 9.3 (SD = 1.2) on the same 10-point scale used above. Bilinguals began learning English at the age of 11.1 years (SD = 2.1) and self-rated their English proficiency at an average of 6 (SD = 1.3).

Sixteen Penn State students (15 female, 1 male; mean age 21 years, SD = .78) who were English-Spanish bilinguals also participated for payment. These participants self-rated their English proficiency at an average of 9.9 (out of 10; SD = .27), their Spanish proficiency at an average of 6.9 (SD = .89) out of 10, and began learning their L2 at an average age of 10.9 years (SD = 4.4).

2.2.3 Stimuli

The stimuli for Experiment 1a 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 and 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 consisted of an equal distribution of monosyllabic, disyllabic, and trisyllabic items chosen from the English Lexicon Project (ELP) non-word database (http://elexicon.wustl.edu) (see Table 1 for a full listing of the nonce words used in Experiment 1a and 1b). All nonce words had American English phonological patterns, were between 4 and 10 characters in length, and based on data from the ELP had an average of 2.2 orthographic neighbors and a bigram mean of 2022. The words were created in a female American English voice (*Crystal*) via the AT&T Natural Voices text-to-speech synthesizer (http://www.naturalvoices.att.com), and converted into WAV files sampled at 22050 Hz. The fifty-four word-object pairs were separated into three non-overlapping sets of eighteen pairs, in which word length was equally distributed.

Monosyllabic	**Bisyllabic**	**Trisyllabic**

barsh	briskle	baturate
blep	crinklow	calorix
chost	dounger	caprion
crid	durrow	clamoreck
daint	grinter	coronick
drock	haser	haterfront
dulch	lattle	interlade
feech	masset	jatterside
frane	mubble	latercress
glack	murler	naureate
glink	pangle	overlood
gotch	patchet	perminal
plock	peadle	rentacle
plunt	pedline	tanderer
scown	pritter	thermistar
slute	tallot	todular
sunch	tarren	tonogram
veam	thecker	ventuker

Table 1. Nonce words used in Experiments 1 & 2, organized by syllable count.

2.2.4 Procedure

During familiarization, participants watched objects appear on a computer screen while listening to words presented over speakers. Each participant completed three familiarization phases, each containing 18 unique word-object pairs, distinguished by the number of words and objects presented in a trial; there was a 2x2 condition (in which participants saw two objects and heard two words), a 3x3 condition, and a 4x4 condition. The order in which participants encountered these familiarizations was randomized, and the set of words and objects presented in each familiarization was non-overlapping. Preceding each trial, a fixation cross appeared for 750ms. During the trial, two to four 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 in that the end of one trial cued the presentation of the next. The order of the trials was pseudo-randomized such that no word-object pair appeared in two consecutive trials. Within each condition, every word-object pair was presented 6 times, and across all trials participants saw a total of 108 objects and heard 108 words. Accordingly, the total number of trials varied by condition: there were fifty-four trials in the 2x2 condition, 36 in the 3x3, and 27 in the 4x4. Total time of familiarization was constant at 320 seconds across all conditions. Prior to beginning the experiment, participants were told that they would be learning novel names for novel objects.

Following each familiarization phase, participants completed a 4 alternative forced-choice (4AFC) test consisting of 18 test trials (i.e., one trial for each word-object pair presented during familiarization). 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 and labeled with a number (1-4). Participants were asked to press the number key corresponding to the correct object without any time limit.

2.2.5 Results

Functionally monolingual English participants in Experiment 1a learned 87% (SD = 17%) of word-object pairs in the 2x2 condition; 70% (SD = 27%) of pairs in the 3x3 condition, and 53% (SD = 19%) of pairs in the 4x4 condition. Chance performance was 25% (since there were four alternatives at test). Single-sample t-tests confirmed that performance in each of the three conditions was significantly above chance (2x2: $t(16) = 15.3$, $p < .001$; 3x3: $t(16) = 6.8$, p

<. 001; 4x4: $t(16) = 5.9$, $p < .001$). Chinese-English bilinguals learned 78% (SD = 19%) of word-object pairs in the 2x2 condition; 66% (SD = 24%) of pairs in the 3x3 condition, and 38% (SD = 13%) of pairs in the 4x4 condition. Single-sample t-tests confirmed that performance in each of the three conditions was significantly above chance (2x2: $t(15) = 11.2$, $p < .001$; 3x3: $t(15) = 7.1$, $p < .001$; 4x4: $t(15) = 4.3$, $p < .01$). English-Spanish bilinguals learned 86.5% (SD = 11%) of word object pairs in the 2x2 condition, 73% (SD = 16%) of pairs in the 3x3 condition and 56% (SD = 20%) of pairs in the 4x4 condition. Single-sample t-tests confirmed that performance in all three conditions for English-Spanish bilinguals was significantly above the level of chance (2x2: $t(15) = 23.2$, $p < .001$; 3x3: $t(15) = 11.4$, $p < .001$; 4x4: $t(15) = 6.2$, $p < .001$) (See Figure 1).

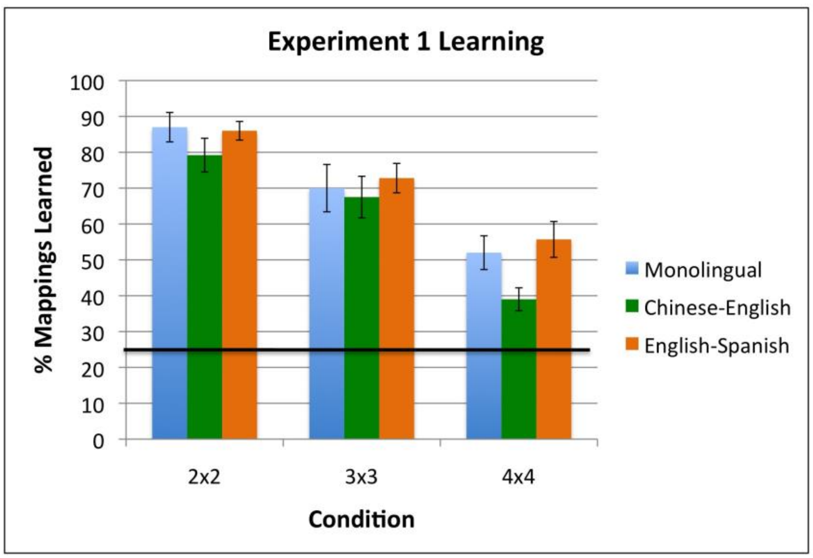

Figure 1 Results from all conditions in Experiment 1a for English monolinguals, Chinese-English bilinguals, and English-Spanish bilinguals, where the black line represents chance performance (25%). Learning is equivalent among the groups in the 2x2 and 3x3 conditions, although in the 4x4 condition the Chinese-English bilinguals acquired significantly fewer mappings (see text). Error bars represent 1 SE of the mean.

A Shapiro-Wilk test confirmed that participants' performance across Experiment 1a was normally distributed ($W(49) = .97$, $p = .22$). We used a 3 (Group) x 3 (Condition) repeated measures ANOVA to investigate factors that influence learning in the CSSL task. Group

(English monolingual, Chinese-English bilingual, English-Spanish bilingual) was a between-subjects factor, while Condition (2x2, 3x3, or 4x4) was a within-subjects factor. There was a main effect of test ($F(2,92) = 71.3$ $p < .001$, $\eta^2 = .6$), such that learning was greatest in the 2x2 condition (M = 84%, SE = 2.3%), followed by the 3x3 condition (M = 70%, SE = 3.3%) and the 4x4 condition (M = 50%, SE = 2.5%). The interaction between test and group was not significant ($F(4,92) = .94$, $p = .44$, $\eta^2 = .04$), nor was the between-subjects factor of Group ($F(2,46) = 1.9$, $p = .19$, $\eta^2 = .07$), indicating that English monolinguals, Chinese-English bilinguals and English-Spanish bilinguals were not significantly different in their learning performance across the three conditions of Experiment 1 (L1 English: M = 70.3%, SE = 3.7%; Chinese-English bilinguals: M = 61.2%, SE = 3.7%; English-Spanish bilinguals: M = 72.4%, SE = 3.8%). A series of one-way ANOVAs compared performance within each learning condition (i.e., 2x2, 3x3, 4x4) between the samples of English monolinguals, Chinese-English bilinguals, and English-Spanish bilinguals. The difference was not significant in the 2x2 condition ($F(2,46) = 1.7$, $p = .2$) or 3x3 condition ($F(2,46) = .32$, $p = .73$), but was significant in the 4x4 condition ($F(2,46) = 4.3$ $p < .05$). Bonferroni corrected post-hoc tests confirmed that there were no differences between the groups in either the 2x2 or 3x3 conditions (all *p*s > .23), no differences between monolinguals and English-Spanish bilinguals in the 4x4 condition ($p > .99$), and significant differences between Chinese-English bilinguals and both monolinguals ($p = .04$) and English-Spanish bilinguals ($p = .012$) in the 4x4 condition, as the Chinese-English bilinguals acquired fewer mappings than the other two groups.

A final analysis examined whether the number of syllables in a word as well as the first language of the participant (i.e., English or Chinese) influenced learnability. We used a 3 (Number of Syllables) x 2 (L1 English or L1 Chinese) ANOVA to investigate this issue. Number of Syllables was a within-subjects factor, while L1 was a between-subjects factor. There was a main effect of number of syllables ($F(2,92) = 5.6$, $p < .01$, $\eta^2 = .11$) indicating that learning of monosyllabic words was significantly higher than bisyllabic and trisyllabic words (mono: M = 71.5%, SE = 2.4%; bi: M = 67%, SE = 2.6%; tri: M = 67.1%, SE = 2.6%). The between-subjects factor of L1 was also significant ($F(1,46) = 5.5$, $p < .05$, $\eta^2 = .11$), such that learning for L1 English participants (M = 72.4%, SE = 3%) was more robust than that of L1 Chinese participants (M = 61.2%, SE = 4%) across all conditions. The interaction between Number of Syllables and L1 was also significant ($F(4,92) = 4.9$, $p < .01$, $\eta^2 = .1$). Planned follow-up tests investigating this interaction showed that L1 English participants did not perform differently based on the number of syllables ($F(2,62) = .97$, $p = .39$, $\eta^2 = .03$; monosyllabic: M = 73.3%, SE = 3.1%; bisyllabic: M = 73%, SE = 3.1%; trisyllabic: M = 70.3%, SE = 3.3%), while L1 Chinese participants performed significantly higher on monosyllabic words relative to bisyllabic or trisyllabic words ($F(2,32) = 6.9$, $p < .01$, $\eta^2 = .3$; monosyllabic: M = 68%, SE = 3.4%; bisyllabic: M = 55%, SE = 4.3%; trisyllabic: M = 61%, SE = 4.1%).

2.2.6 Discussion

The overall findings from Experiment 1a suggest that monolinguals and bilinguals do not significantly differ with respect to their ability to engage in cross-situational statistical learning. The only difference in performance between the three groups emerged in the 4x4 condition, which was the most difficult condition for all participants. Given that performance in the 2x2 and 3x3 conditions was equivalent across groups, and that the English-Spanish bilinguals performed equivalently to the English monolinguals on all conditions, we doubt that the difference in the 4x4 condition reflects a true difference in statistical learning abilities. Rather, the stimuli conformed to English phonology and this may have advantaged native English-speaking participants, as evidenced by the decrement in performance on multisyllabic words that occurred for Chinese-English bilinguals but not for the native English speakers. In the 4x4 condition, the probability of encountering a multisyllabic word in every scene was higher than in the other conditions. This could explain why this condition was particularly hard for the Chinese-English bilinguals relative to the English monolinguals and English-Spanish bilinguals. Future experiments using words that conform to the native phonology of the Chinese-English bilinguals could further elucidate the source of this performance difference.

2.3.1 Experiment 1b

Having discovered relatively similar levels of performance by functionally monolingual and bilingual learners on the standard version of CSSL, in Experiment 1b, we extended the paradigm to provide learners with the opportunity to form multiple mappings. Several previous studies have investigated whether learners obey mutual exclusivity in CSSL when they have the opportunity to map one item with multiple objects or labels. The results of these studies suggest that there is a bias toward mutual exclusivity.

Yurovsky, Smith, and Yu (2013), for example, exposed learners to an initial set of 1:1 mappings and in a subsequent training session remapped a subset of words to new objects. Learners were able to acquire both the first (primacy) and second (recency) referent of remapped words, but in direct preference tests demonstrated a bias towards the primacy referent. A follow-up condition found that learners could acquire two mappings for a word within the same training session. Similarly, Ichinco, Frank and Saxe (2009) presented learners with a set of 18 one-to-one mappings in a 3x3 design, followed by a second training phase that included a new set of one-to-one mappings as well as a subset of items transferred from the first familiarization phase. These items appeared with their original mapping but could also be mapped to one of the new items. In this way, learners were afforded the opportunity to form a second mapping for the transferred item, but still encountered information consistent with the primacy mapping. Learners in this paradigm preferred the primacy mapping and did not acquire the new (recency) mapping. A more recent study from our lab replicated this result and also found that the addition of contextual cues to the second familiarization (e.g., a change in speaker voice) facilitated acquisition of the recency mapping and significantly reduced participants' preference for primacy mappings (Poepsel, Gerfen & Weiss, 2012). Finally, Kachergis, Yu and Shiffrin (2012) modeled CSSL in environments presenting learners with two-to-one mappings, and concluded that learners could modulate their reliance on mutual exclusivity in response to the proportion of mappings in the input that either followed or violated ME. Increased exposure to mappings that violated mutual exclusivity predicted better acquisition of new mappings.

In Experiment 1b we compared the learning performance of a group of functionally monolingual English speakers to late-learning Chinese-English and English-Spanish bilinguals using a CSSL task that presented learners with a mix of one-to-one and two-to-one mappings. As noted previously, there is empirical support in the developmental literature for the idea that bilinguals may be more likely to assume multiple underlying structures (here mappings) relative to monolinguals as a consequence of frequent exposure to multiple languages (e.g., Byers-Heinlein & Werker, 2009; Houston-Price et al., 2010; Kovacs & Mehler, 2008), but this was in the context of tasks that did not require statistical learning and the participants all had early exposure to multiple languages. We chose to study late learning bilinguals to investigate whether experience with multiple languages could influence statistical learning, even for learners who shared similar early L1 experiences with monolinguals (i.e., sequential bilinguals presumably did not relax the mutual exclusivity constraint during development).

We chose to implement a modified version of the task used by Yurovsky, Yu, and Smith (2013), as this paradigm facilitated learning of multiple mappings (with a preference for the mutual exclusivity mapping) and could therefore permit measures of learning related to the amount of exposure. In order to assess whether there might be differences in how quickly the mappings were acquired, we provided participants with a series of three distinct familiarization phases, each followed by a test. Following Yurovsky, Yu and Smith (2013), we provided learners with multiple words mapped to a single object. This conferred a practical advantage relative to mapping multiple objects to a single word in that the new objects appeared on the screen throughout the trial, whereas novel words would have been highly transient (and in our previous work, remapping words was more effective; Poepsel, Gerfen, & Weiss, 2012). With respect to bilingual acquisition, interlingual homographs are known to be challenging, particularly when presented in conjunction with cognates (see Brenders, van Hell, & Dijkstra, 2011). Since interlingual homographs occur less frequently than translation equivalents, this manipulation arguably provided a subtler test of differences in mapping abilities between monolinguals and late-learning bilinguals.

2.3.2 Participants

Sixteen English monolinguals (11 female, 5 male) from Penn State University who did not participate in Experiment 1 participated for course credit. Based on language history questionnaire data, these participants had a mean age of 18.6 years (SD = 0.65) and self-rated their proficiency in English at 10 (SD = 0) on the ten-point scale used above. Six of these participants had been classroom learners of Spanish, who began receiving instruction at a mean age of 13.2 years (SD = 1.5) and self-rated their L2 proficiency at a mean of 2.5 (SD = 1.2) on the same ten-point scale.

Sixteen Chinese-English bilinguals (13 female, 3 male) from Beijing Normal University in Beijing, China who did not participate in Experiment 1 participated for payment. Based on language history questionnaire data, these participants had a mean age of 22.9 years (SD = 2.4), began learning English at an average age of 10 (SD = 4), and self-rated their proficiency in English at 6.6 (SD = 1.4) on a ten-point scale.

Sixteen English-Spanish bilinguals (14 female, 2 male) from Penn State University who did not participate in Experiment 1 participated for payment. These participants had a mean age of 20.2 years (SD = 1.1) and self-rated their overall proficiency in English at 10 out of 10 (SD = 0). These participants indicated that they began learning their L2 at a mean age of 12.6 years (SD = 1.8) and rated their L2 proficiency at an average of 6.4 out of 10 (SD = 1.8).

2.3.3 Stimuli

Stimuli consisted of 18 nonce words and 24 novel objects, grouped into twelve 1:1 (one object to one word) mappings and six 2:1 (two objects to one word) mappings. All stimuli were chosen from the 54 word-object pairs used in Experiment 1a. Due to the results of Experiment 1a, words chosen for this experiment were all monosyllabic, between four and five characters in length, and contained between three and five phonemes. They followed one of four syllable patterns (CVC, CCVC, CVCC, CCVCC). Based on data from the ELP, nonce words had an average of 4.3 orthographic neighbors and a bigram mean of 4983.

2.3.4 Procedure

Participants completed three familiarization phases, each of which was followed by a test phase (see Figure 2). Within each familiarization phase, participants were exposed to the same set of twelve 1:1 and six 2:1 object-word mappings across twenty-four training trials. Before each trial, a fixation cross appeared for 500 ms. Trials presented three objects and three words as in the 3x3 condition of Experiment 1a. Participants were exposed to four instances of each 1:1 mapping within a familiarization. Each pairing for the 2:1 mappings (e.g., object A - word 1; object B – word 1) was presented twice. The ordering of the 24 trials within each familiarization was pseudo-randomized such that no word-object pair appeared in consecutive trials.

After each familiarization phase, participants completed a 2AFC test in which the learning of both 1:1 and 2:1 mappings was assessed. Test trials in Experiment 2 were largely similar in their presentation to those of Experiment 1a. On each test trial, participants heard a single word and saw two objects, one a distractor and one the correct referent of the presented word. The order of trials was randomized in all three tests.

For each test following the first and second familiarizations, participants completed eighteen test trials. All twelve 1:1 mappings were tested once. Participants also received one test trial for each of the six 2:1 mappings; each possible referent (i.e., primacy and recency) was tested once across the first two tests, and the order in which both referents were tested was pseudorandomized and counterbalanced across participants. That is, if the primacy mapping was tested after the first familiarization (i.e., the first label for an object encountered by the learner), the recency mapping (i.e., the second label for an object) was tested after the second familiarization, and vice versa. This procedure was instantiated in order to ensure that participants were not explicitly cued to the presence of multiple mappings for some objects and not reinforced for one mapping over another. Moreover, in each of the first two tests, half of the 2:1 mapping trials probed primacy mappings, while the other half probed recency mappings. In the test following the third familiarization, participants completed 36 test trials; each 1:1 mapping was tested twice, and both the primacy and recency referents of words with 2:1 mappings were tested once, as the aforementioned concern was no longer relevant. Participants were given as much time as needed to make each response.

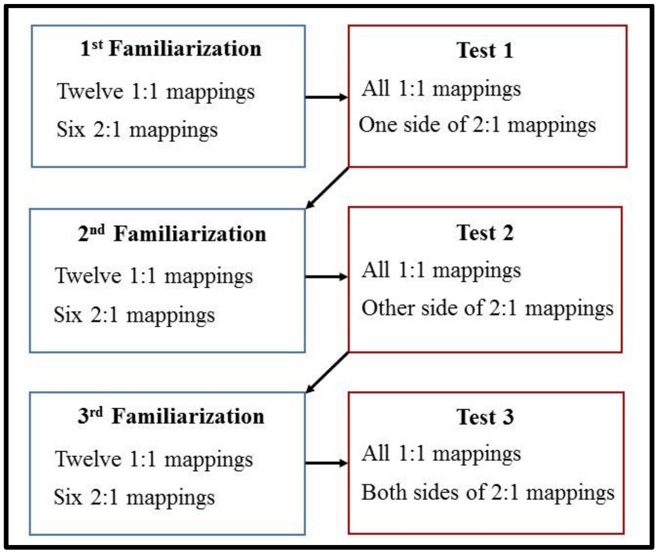

Figure 2 Experiment 1b was composed of three familiarization phases, each of which was followed by a test. In tests 1 and 2, participants completed a trial for either the primacy or recency mapping of each 2:1 mapping. In test 3, both the primacy and recency mappings were tested.

2.3.5 Experiment 1b Results

In the test following the 1st familiarization, English monolinguals learned 70% (SD = 13%) of 1:1 object-word mappings and were accurate on 56% (SD = 21%) of trials probing 2:1 mappings, Chinese-English bilinguals learned 71% (SD = 13%) 1:1 mappings and were accurate on 61% (SD = 20%) of trials probing either side of the 2:1 mappings, and English-Spanish bilinguals learned 72% (SD = 15%) of 1:1 mappings and were accurate on 66% (SD = 15%) of trials probing either side of the 2:1 mappings. In the test following the 2nd familiarization, English monolinguals learned 81% (SD = 14%) of 1:1 mappings and were accurate on 70% (SD = 20%) of trials probing 2:1 mappings, Chinese-English bilinguals learned 84% (SD = 14%) of 1:1 mappings and were accurate on 82% (SD = 15%) of trials probing either side of the 2:1 mappings, and English-Spanish bilinguals learned 92% (SD = 12%) of 1:1 mappings and were accurate on 79% (SD = 20%) of trials probing 2:1 mappings. In the test after the third familiarization, English monolinguals learned 92% (SD = 16%) of 1:1 mappings and were accurate on 77% (SD = 22%) of trials probing 2:1 mappings, Chinese-English bilinguals learned 93% (SD = 10%) of 1:1 mappings and were accurate on 88% (SD = 12%) of trials probing 2:1 mappings, and English-Spanish bilinguals learned 97% (SD = 7%) of 1:1 mappings and were accurate on 94% (SD = 7%) of trials probing 2:1 mappings.

The learning averages reported above were compared against chance (which was set at 50% as a result of the 2AFC design) in a series of single sample t-tests. For monolinguals, learning was significantly above chance for 1:1 mappings in all three tests and for 2:1 mappings in both the second and third tests (all *p*s < .01). Monolinguals did not exceed chance performance on 2:1 mappings following in the first test ($t(16) = 1.1$, $p = 0.27$). For both Chinese-English (C-E) and English-Spanish (E-S) bilinguals, learning was significantly above chance for 1:1 and 2:1 mappings in all tests (C-E bilinguals all *p*s < .05; E-S bilinguals all *p*s < .01).

As our primary goal was to compare the statistical learning abilities of our monolingual participants to those of our bilinguals, we carried out several analyses to determine whether the Chinese-English and English-Spanish bilinguals were statistically equivalent in their performance in Experiment 1b. Thus we compared their performance using a 2 (Group) x 3 (Test) x 2 (Mapping Type) repeated-measures ANOVA. Test and Mapping Type were within-subjects factors, while Group (C-E and E-S bilinguals) was a between-subjects factor. There was a main effect of Test ($F(2, 60) = 59.3$, $p < .001$, $\eta^2 = .66$) indicating that learning performance increased from Test 1 to Test 3. There was also a main effect of Mapping Type ($F(1, 30) = 21.2$, $p < .001$, $\eta^2 = .41$), demonstrating that performance on 1:1 mappings was higher than that for trials probing either side of the 2:1 mappings. Critically, the between-subjects factor of Group was not significant ($F(1, 30) = 1.38$, $p = .25$, $\eta^2 = .04$), suggesting that both groups of bilinguals did not perform differently across both mapping types in Experiment 1b. Furthermore, no interaction terms reached significance (all Fs < 1.31, all *p*s > .28, all η^2 s < .04). To further verify that the bilingual groups were equivalent in performance, we ran a follow-up 2 (Group) x 3 (Test) ANOVA for each mapping type. As above, Test was a within-subjects factor and Group was a between-subjects factor. For both 1:1 mappings and 2:1 mappings, there was a significant main effect of Test (1:1 mappings: $F(2, 60) = 43.4$, $p < .001$, $\eta^2 = .59$; 2:1 mappings: $F(2, 60) = 25.4$, $p < .001$, $\eta^2 = .46$) such that performance increased from Test 1 to Test 3. Again, the between-subjects factor of Group did not reach significance for either mapping type (1:1 mappings: $F(1, 30) = 1.96$, $p = .17$, $\eta^2 = .06$; 2:1 mappings: $F(1, 30) = .54$, $p = .47$, $\eta^2 = .01$), nor did the interaction of Test and Group for either mapping type (1:1 mappings: $F(2, 60) = .72$, $p = .49$, $\eta^2 = .02$; 2:1 mappings: $F(2, 60) = .79$, $p = .46$, $\eta^2 = .03$) suggesting that English-Spanish and Chinese-English bilinguals performed equivalently on all mappings in Experiment 2.

Having established this equivalence, we combined these groups into a single bilingual group for comparison against the sample of English monolinguals. Subsequently, we used a 2 (Group) x 3(Test) x 2 (Mapping Type) repeated-measures ANOVA to determine the influence of bilingual experience on learning in Experiment 1b. A Shapiro-Wilk test confirmed that participants' performance across Experiment 1b was normally distributed ($W(48) = .97$, $p = .23$). Group (monolingual, bilingual) was a between-subjects factor, while Test (performance on post-tests 1, 2 and 3) and Mapping Type (1:1 or 2:1) were within-subjects factors. There was a main effect of Test ($F(2,92) = 62.9$, $p < .001$, $\eta^2 = .58$), indicating that learning increased over the course of the experiment (Test 1: M = 65.2%, SE = 2%; Test 2: M = 79.9%, SE = 1.9%; Test 3: M = 88.6%, SE = 1.7%). We also found a main effect of Mapping Type ($F(1,46) = 52.9$, $p < .001$, $\eta^2 = .54$), indicating that participants' performance was significantly higher on trials probing 1:1 mappings than 2:1 mappings (1:1 mappings: M= 82.8%, SE = 1.4%; 2:1 mappings: M = 73%, SE = 1.7%). The interaction of Mapping Type and Group was also significant ($F(1,46) = 6.2$, $p < .05$, $\eta^2 = .12$) as bilinguals exhibited a smaller difference in performance between 1:1 and 2:1 mappings relative to monolinguals. The between-subjects factor of Group was significant ($F(1,46) = 7.0$, $p < .05$, $\eta^2 = .13$) providing evidence that bilinguals performed better overall in Experiment 1b relative to monolinguals (bilinguals: M = 81.7%, SE = 1.7%; monolinguals: 74.1%, SE = 2.3%).

Planned follow-up tests further explored the significant Group factor, directly comparing monolinguals' and bilinguals' performance on trials probing 1:1 and 2:1 mappings in separate ANOVAs. Here we found no significant difference in performance on 1:1 mappings between the groups ($F(1,46) = 2.2$, $p = .14$, $\eta^2 = .04$), but a significant difference in performance on trials probing 2:1 mappings ($F(1,46) = 9.8$, $p < .01$, $\eta^2 = .18$), as bilinguals outperformed monolinguals on this test trial type (monolinguals: M = 67.5%, SE = 2.9%; bilinguals: M = 78.5%, SE = 2%). The factors for Test x Group, Test x Mapping Type, and Test x Mapping Type x Group did not reach significance (all Fs < .85, all *p*s > .43).

The analyses carried out above for 2:1 mappings were based on participants' accuracy on trials probing one of the two sides of these mappings; this does not necessarily provide information on how frequently participants were accurate on both sides of a 2:1 mapping. Across the first two tests, participants completed trials probing both the primacy and recency mappings for each 2:1 mapping. In the third test, both the primacy and recency mappings were probed (see Methods for further detail). In assessing performance across both sides of the 2:1 mapping, we again compared English monolinguals to a combined group of Chinese-English and English-Spanish bilinguals. Across the first two tests, monolinguals were accurate on both sides of a 2:1

mapping in 36% (SD = 23%) of cases, while bilinguals were accurate on both sides in 50% (SD = 21%) of cases. This difference was significant ($t(46) = 2.03$, $p < .05$), suggesting that bilinguals learned more 2:1 mappings than monolinguals across the first two familiarizations (see Figure 3). In the third test, monolinguals were accurate on both sides of a 2:1 mapping 59% (SD = 33%) of the time , while bilinguals achieved an accuracy of 83% (SD = 19%). This difference was also significant ($t(46) = 3.18$, $p < .01$), again suggesting that bilinguals acquired more 2:1 mappings than monolinguals by the end of training in Experiment 1b.

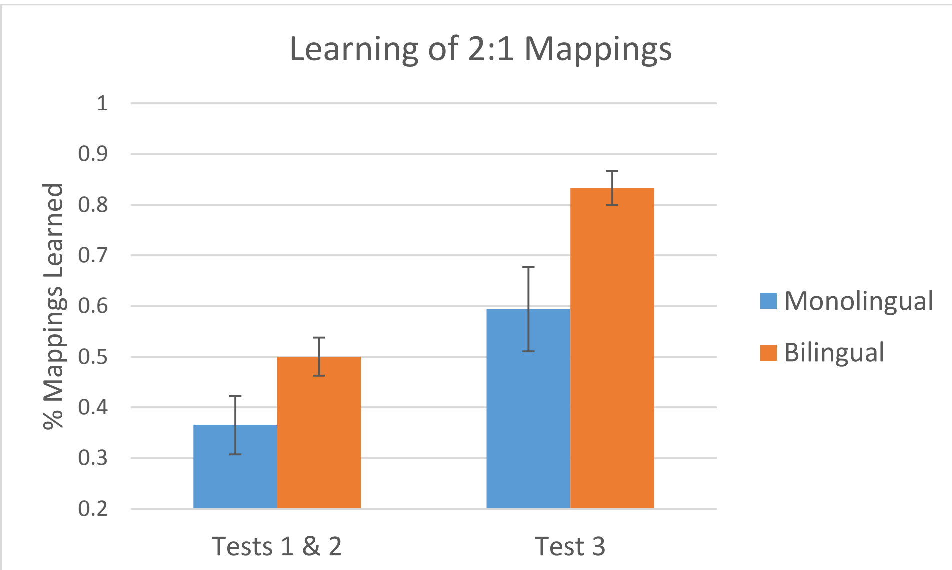

Figure 3 Bilinguals acquired significantly more 2:1 mappings than monolinguals across the first two tests and in the third test. Error bars represent 1 SE of the mean.

For words with 2:1 mappings, we were also interested in whether participants exhibited a bias toward either primacy or recency mappings and whether bilingual experience impacted mapping preferences. We used a 3 (Test) x 2 (Mapping Type) x 2 (Group) ANOVA to investigate the factors that influenced acquisition of primacy and recency mappings in Experiment 1b. Test (first, second, third) and Mapping Type (primacy, recency) were within-subjects factors, while Group (monolingual, bilingual) was a between-subjects factor. There was a main effect of Test ($F(2,92) = 22.4$, $p < .001$, $\eta^2 = .33$) demonstrating that overall performance increased across the three tests (Test 1, M = 60.7%, SE = 2.8%; Test 2: M = 75.5%, SE = 2.8%; Test 3: M = 83.9%, SE = 2.3%). There was also a main effect of Mapping Type ($F(1,46) = 12.6$, $p < .001$, $\eta^2 = .22$), as performance on primacy mappings was higher relative to recency mappings (primacy: M = 77.7%, SE =2.1%; recency: M = 68.9%, SE = 2.1%). The between-subjects factor also reached significance ($F(1,48) = 10.8$, $p < .01$, $\eta^2 = .19$) suggesting that bilinguals acquired more mappings (both 1:1 and 2:1) than monolinguals across Experiment 1b (bilingual: M = 78.5%, SE = 2.9%; monolingual: M = 67.5%, SE = 2%). The factors of Test x Mapping Type, Test x Group, Mapping Type x Group, and Test x Mapping Type x Group did not reach significance (all Fs < 1.7, all *p*s > .18).

Planned follow-up tests further investigated the interaction of Test x Mapping Type within the monolingual and bilingual groups. For monolinguals, the interaction of Test x Mapping Type was not significant ($F(2,30) = 1.9$, $p = .17$, $\eta^2 = .1$), while for bilinguals, this interaction was significant ($F(2,62) = 6.6$, $p < .01$, $\eta^2 = .18$), indicating that over the course of learning, bilinguals' performance on recency and primacy showed a greater trend of convergence than monolinguals (see Figure 4). Another set of planned follow-up tests compared the groups' performance on each mapping type (primacy and recency). For both primacy and recency mappings, the between-subjects factor of Group was significant (primacy: $F(1,46) = 6.5$, $p < .05$, $\eta^2 = .12$; recency: $F(1,46) = 7.6$, $p < .01$, $\eta^2 = .14$), such that bilinguals performed better on each mapping type than monolinguals across Experiment 1b.

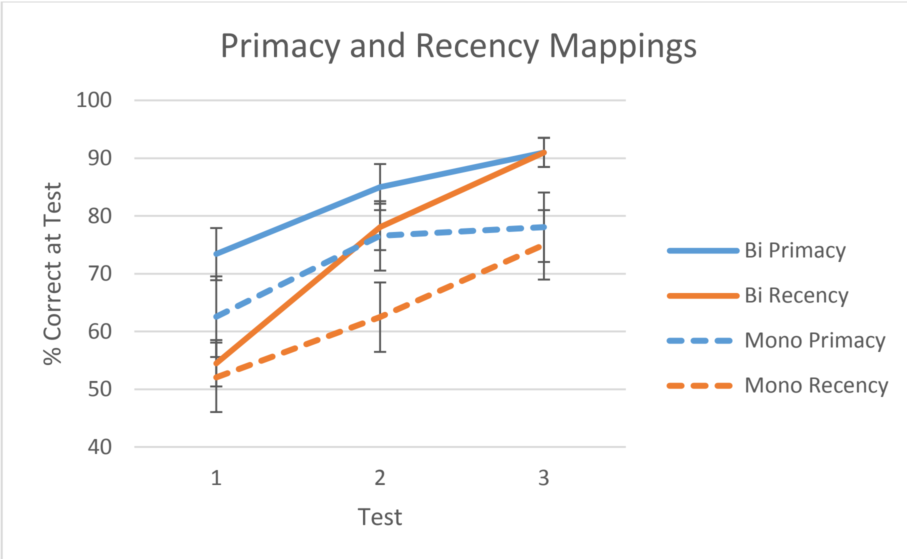

Figure 4 Performance on primacy and recency mappings across tests for monolinguals and bilinguals. Error bars represent 1 SE of the mean.

A final analysis compared the performance of the two groups of English L1 participants (English monolinguals and English-Spanish bilinguals) in Experiment 1b. Here we used a 3 (Test) x 2 (Mapping Type) x 2 (Group) ANOVA, set up similarly to those presented above. There was a main effect of Test ($F(2,60) = 48.6$, $p < .001$, $\eta^2 = .62$) indicating an increase in performance across tests, and a significant interaction between Mapping Type and Group ($F(1,30) = 6.0$, $p < .05$, $\eta^2 = .17$). The between-subjects factor of Group was also significant ($F(1,30) = 5.2$ $p < .05$, $\eta^2 = .15$). Two follow-up tests investigated this significant between-subjects factor, comparing both groups in their performance on each mapping type. We found no difference in performance between these groups on 1:1 mappings ($F(1,30) = 1.3$, $p = .26$, $\eta^2 = .04$), but a significant difference in performance on 2:1 mappings ($F(1,30) = 7.6$, $p < .05$, $\eta^2 = .2$), demonstrating that English-Spanish bilinguals outperformed their monolingual counterparts on this mapping type.

2.3.6 Discussion

Across Experiment 1a and 1b, we compared a group of functional monolinguals to two groups of late-learning bilinguals (Chinese-English and English-Spanish) to determine whether there were differences in the acquisition of 1:1 and 2:1 mappings in cross-situational statistical word learning tasks. In Experiment 1, we replicated the CSSL study of Yu and Smith (2007) with our own set of stimuli. Participants learned unique sets of 1:1 mappings in three conditions that varied with respect to how many word-object pairings appeared at once. Overall, the results of Experiment 1a suggest that the three groups are very similar in their ability to acquire 1:1 mappings in a standard CSSL task. Across all conditions, there was only one significant difference in performance across the three groups. Chinese-English bilinguals had lower performance relative to the other two groups in the most difficult 4x4 condition. As noted above, this might be attributable to the fact that the stimuli were presented in the phonology of their L2 (in particular, multi-syllabic stimuli posed difficulties for these bilinguals).

In Experiment 1b, three new samples of participants drawn from the same populations were familiarized with a mixed set of 1:1 and 2:1 mappings in a 3x3 cross-situational word learning design with three consecutive familiarization phases, each of which was followed by a test. As in the 3x3 condition of Experiment 1a, all groups achieved equivalent levels of performance in learning the 1:1 mappings, providing further evidence that these groups were matched in their core cross-situational statistical learning abilities. However, Chinese-English and English-Spanish bilinguals acquired significantly more 2:1 mappings than the monolinguals across all three tests. Further, bilinguals' performance on both primacy (the first label paired with an object and recency mappings (the second label paired with a given object) was significantly higher than that of the monolinguals across all familiarizations, and bilinguals showed a stronger trend of convergence in their performance on primacy and recency mappings compared to monolinguals. Taken together, these results suggest that bilinguals may have acquired true 2:1

mappings earlier than monolinguals, who showed a more consistent bias toward primacy mappings. In sum, despite broad similarities in performance on 1:1 mappings, we found that Chinese-English and English-Spanish bilinguals seemed to acquire 2:1 mappings with less exposure and greater overall proficiency than monolinguals. Thus, our results suggest that when tracking novel statistical inputs over time, late bilingual learners appear to be more open to the possibility of multiple mappings in the input.

These findings extend the current knowledge regarding bilingualism and statistical learning in several respects. As noted in the Introduction, to date there has been mixed evidence regarding whether monolinguals and bilinguals differ with respect to their statistical learning abilities (Yim & Rudoy, 2013; Bartolotti et al., 2011; Wang & Saffran, 2014). Whereas the previous studies have focused on the task of speech segmentation, here we extend the investigation to statistical word learning, which may involve different cognitive processes (see below). One interesting parallel that emerged between our studies and previous work is that monolinguals and bilinguals seem to perform similarly on the most straightforward versions of these tasks. For example, when studies required learners to track only the transitional probabilities between adjacent elements without any sources of interference, the results of bilinguals mirror monolinguals (Yim & Rudoy, 2013 with children; see also Bogulski, 2013 with adults). In Experiment 1a, we found that on the 1:1 version of the CSSL task, performance was largely similar across two populations of late-learning bilinguals and our functional monolinguals (also for 1:1 mappings in Experiment 1b), including Chinese-English bilinguals for whom English phonology may have posed an additional challenge (as evidenced by our analysis of performance based on the number of syllables in a word). While more data is required to draw a firm conclusion, evidence thus far supports the idea that the most basic forms of statistical learning, involving either tracking sequential probabilities or accruing associative information over time, may be relatively unaffected by experience with second language learning.

By contrast, the previous statistical learning studies that do report differences between monolingual and bilingual learners have all contained multiple cues to segmentation. Arguably, in these studies learners must suppress one set of cues in order to correctly segment the stream. In the case of Bartolotti et al. (2011) transitional probability statistics were tracked along with a competing pause cue, whereas in the case of Wang and Saffran (2014) a congruent suprasegmental tone cue appeared to hinder performance, particularly for monolinguals (relative to prior studies in which learners only tracked similar transitional probabilities between adjacent syllables without additional cues) and the authors suggest that learners may have had to inhibit one cue to follow the other. We note that it is unknown whether the bilingual groups tested in the latter study would have also maintained an advantage had the suprasegmental cues been removed. Our study did not involve a second cue type, but did offer learners the opportunity to form a second mapping for one of the elements. In this situation, bilinguals also appear advantaged relative to functional monolinguals, being better able to overcome a mutual exclusivity bias that has been evidenced with adults in previous cross-situational learning studies using a similar experimental paradigm (e.g., see Yurovsky, Smith, & Yu, 2013). Future work in this area will need to determine whether all of these advantages arise as a function of bilinguals possessing a general advantage in implicit learning abilities (e.g., Klein, 1995; Nation & McLaughlin, 1986), or whether they emerge due to other cognitive advantages associated with bilingualism such as improved phonological working memory (e.g., Service et al., 2002; Majerus et al., 2008; Adesope et al., 2010; Bartolotti et al., 2011) or an advantage in inhibitory control (e.g., Bialystok, 1999; Bialystok et al., 2004; Costa, Hernández, & Sebastián-Gallés, 2008; Tao et al., 2011; Wang & Saffran, 2014), or some combination thereof. It is possible that in Experiment 1b, learners may have had to inhibit the first mapping (i.e., inhibitory control) in order to acquire the second. Future work will include measures of inhibitory control and working memory that may begin to address this issue. Irrespective, another contribution of the present work is that it is the first demonstration, to the best of our knowledge, that late-learning sequential bilinguals are more open to remapping during accumulative statistical learning relative to functional monolinguals.

As noted in the Introduction, the term statistical learning encompasses many forms of learning, and thus it is quite possible that the different types of statistical learning involve a diverse set of cognitive processes. For example, Hsu and Bishop (2010) suggest that the statistical learning involved in word learning may differ from that of grammar learning and also from nonverbal sequence learning (see Figure 1 of Hsu & Bishop, 2010). Correspondingly, deficits such as Specific Language Impairment are thought to impact the different types of statistical learning to varying degrees (see Hsu & Bishop, 2010; Ullman & Pierpont, 2005). One attempt to formalize this kind of distinction has been proposed by Thiessen, Kronstein, and Hufnagel (2013) who differentiate between statistical learning tasks involving extraction and those involving integration. Extraction involves holding two elements in working memory and binding them into a chunk (Perruchet & Vintner, 1998) whereas integration involves combining information across chunks to deduce a central tendency (see Erickson & Thiessen, 2015). These distinctions may provide a useful framework for interpreting the results of statistical learning studies comparing monolinguals and bilinguals. Previous studies exploring statistical learning and bilinguals have all involved sequence-learning, which is best characterized as extraction (Erickson & Thiessen, 2015) whereas the present study involve processes related to integration (since information is stored across trials to deduce the correct associations). Thus, it is possible that the bilingual advantage in forming 2:1 mappings may be evidenced in processes involving integration, but not those involving extraction. For example, data collected by Bogulski (2013) suggest that bilinguals may be prone to the same primacy effect as monolinguals when asked to track sequential regularities across two artificial speech streams presented consecutively without any cues to the change in structure (see Gebhart, Aslin, & Newport, 2009; see also Weiss, Gerfen, & Mitchel, 2009). Arguably, this dual-stream task also involves remapping, since there is a fifty percent overlap in elements used in both languages (and thus learners would have to recognize that the initial set of transitional probabilities across elements no longer apply to the overlapping elements found in the second stream). Thus, it is possible that statistical learning abilities involving extraction are equivalent across populations (unless a second cue is added to the stream, see above) while tasks that require accruing statistics over time may be approached differently by bilinguals, who are more open to the possibility of multiple distributions. In our view, the extraction-integration framework could provide some traction in understanding when differences are likely to emerge between monolinguals and bilinguals in statistical learning tasks and our future work will explore this possibility more directly.

With respect to mutual exclusivity, the present findings contribute to an existing literature that has largely studied simultaneous or early sequential bilinguals. For example, several studies have reported that bilingual infants are less likely to adhere to the mutual exclusivity constraint relative to monolinguals (e.g., Byers-Heinlein & Werker, 2009; Davidson & Tell, 2005; Houston-Price et al., 2010). By contrast, the bilingual participants in this study were late L2 learners, and consequently their experience acquiring their L1 was likely equivalent to their monolingual peers with respect to adhering to word-learning constraints. Thus, these findings provide suggestive evidence that this constraint may become relaxed (at least in the context of a statistical learning task) even for sequential bilinguals whose acquisition of the L2 comes later in life. This accords with the notion put forth by Markman and Wachtel (1988) that mutual exclusivity likely persists into adulthood (e.g., Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Halberda, 2006) but weakens with age as learners come to experience more overlap in terms of mappings. For bilinguals, who experience translation equivalents and interlingual homographs as well as synonymy and polysemy within each language, the mutual exclusivity constraint may consequently become significantly weaker relative to monolinguals. This also is consistent with a recent model of word learning that demonstrated decrements in mutual exclusivity when multiple labels were present for a single object (though interestingly not when multiple meanings were tested, see McMurray, Horst, & Samuelson, 2012). Thus, our findings add to a growing literature demonstrating that the cognitive impacts of late sequential bilingualism mirror some of the changes associated with simultaneous or early bilingualism (e.g., Vega-Mendoza, West, Sorace, & Bak, 2015).

We also note the importance of testing multiple populations of bilingual learners to compare with functional monolinguals. In order to be confident that the observed differences arose as a consequence of proficiency with a second language, it was necessary to decrease the likelihood that our findings could arise as a function of idiosyncrasies associated with a particular language. For example, speakers of Chinese must resolve homophonous relationships frequently as a result of great overlap in Chinese characters and syllables (e.g., Chang, 1993; Perfetti & Tan, 1998; Kuo et al.,2004), and thus it is conceivable that any Chinese speaker (including monolinguals) could expect greater success in forming 2:1 mappings relative to English monolinguals who do not encounter homophony on such a scale. Therefore, the finding that English-Spanish bilinguals perform equivalently to the Chinese-English bilinguals (and also outperform functional monolinguals) in forming 2:1 mappings in Experiment 2 offers very suggestive evidence that the effect is language-independent and arises as a consequence of proficiency with two languages.

Returning to the "learning to learn" framework mentioned in the Introduction, this study provides evidence that experience with multiple languages may fundamentally influence the assumptions made in new learning environments with respect to how many causal models underlie the observed statistics. This notion finds support in developmental studies of early bilingualism (e.g., Kovacs & Mehler, 2008), and, as we have shown here, even when proficiency with a second language occurs after the first language has already been mastered. Based on our findings along with previous studies (Yim & Rudoy, 2013; Bogulski 2013), we argue that the core distributional learning abilities evidenced in transitional probability tasks or cross-situational statistical learning tasks may be unaffected by proficiency with more than one language; just as being exposed to multiple rules is unlikely to impact the most basic principles of learning (e.g., Gallistel, et al., 2001). Rather, the differences in performance are more likely to become evident in novel non-stationary environments when learners must determine the number of mappings or structures that underlie the surface input.

Chapter 3: The Influence of Bilingualism on Ambiguity Resolution in Cross-Situational Word Learning with Multiple Inputs

Mapping words to objects is a difficult problem faced by language learners. There are many constraints that have been posited to facilitate this task, such as the whole-object bias (Markman, 1991), the Principle of Contrast (Clark, 1983), and pragmatic constraints (Diesendruck & Markson, 2001). As noted above, it has recently been proposed that learners may employ a form of statistical learning to help overcome indeterminacy (i.e., the near infinite number of possible objects that can be considered for a given word). The logic of this assertion is that word meanings may be ambiguous within the context of a single scene, but aggregating co-occurrence information about words and objects across multiple scenes can help disambiguate which words belong with which objects. Many experiments in the cross-situational statistical learning (CSSL) paradigm have demonstrated that learners are sensitive to such co-occurrence statistics, and can use them to successfully map novel words to novel objects (e.g., Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Vlach & Johnson, 2013; Yu & Smith, 2007).

Although a majority of CSSL experiments to date have examined the acquisition of 1:1 mappings (i.e., mappings in which a single word is paired with a single object), recent work has begun to investigate how learners accommodate 2:1 mappings (i.e., mappings in which two objects are paired with a single word or vice versa; e.g., Yurovsky, Fricker, Yu, & Smith, 2015; Yurovsky & Yu, 2013; Ichinco, Frank, & Saxe, 2009; Poepsel, Gerfen, & Weiss, 2012; Poepsel & Weiss, 2016). The findings from these studies suggest that learning is influenced by the chosen paradigm. In Ichinco et al. (2009), learners were exposed to an initial learning block that contained only 1:1 mappings. Following this, learners were exposed to a second block containing novel words and objects that formed 1:1 mappings, as well as a set of items from the first block (either words or objects) that could also be mapped to the novel words and objects. In this way, multiple mappings were available to learners in the second block, but not forced, as learners could choose to ignore the previously learned items and map only novel words to novel objects. Learners in this paradigm exhibited a mutual exclusivity bias (Markman, 1994), choosing to form only 1:1 mappings and ignoring the evidence for 2:1 mappings. Put another way, learners chose not to remap previously learned items, a finding that has been replicated by several other studies using the same paradigm (Poepsel, Gerfen, & Weiss, 2012; Poepsel & Weiss, 2014). Based on these results, one possible conclusion is that initial exposure to a learning environment in which mutual exclusivity is not violated may bias a learner toward applying mutual exclusivity in subsequent environments, even if this strategy does not perfectly accommodate all of the available evidence. This accords with the modeling work of McMurray et al. (2012), as well as two fast-mapping studies showing that infant monolingual learners (whose learning environments may exhibit a high degree of mutual exclusivity) are more likely to follow mutual exclusivity compared to bilingual and trilingual infants (whose learning environments = contain relatively more violations of mutual exclusivity; Byers-Heinlein & Werker, 2009; Houston-Price et al., 2010).

Context has also been shown to be a powerful disambiguating cue in cases where multiple relationships may exist. For example, when two speakers produce different descriptions for a novel object (such that voice or gender serves as a contextual cue), there may be no cost to learners in terms of associating multiple descriptions with a single object. However, when both descriptions are produced by a single speaker, there is a cost associated with violating the initial description (e.g., Metzing and Brennan, 2003; Trude and Brown-Schmidt, 2012). Such a finding is broadly consistent with work in speech segmentation with multiple inputs, where multiple findings suggest that learning is facilitated when the inputs are differentiated by a contextual cue (e.g., Weiss et al., 2009; Gebhart et al., 2009; Mitchel & Weiss, 2010). In a CSSL paradigm, contextual cues that differentiate two possible mappings have also been shown to decrease reliance on mutual exclusivity and support the acquisition of many-to-one mappings; thus, the primacy effect reported in Ichinco et al. (2008) can be mitigated by changing the voice in which new mappings to previously learned words or objects are presented (Poepsel et al., 2012).

One feature of these studies is that the learning of multiple mappings occurs across two blocks, so that first one mapping is introduced and tested, followed by another. Learners may not always encounter evidence for multiple mappings so discretely, however. This has also been explored in a CSSL paradigm. In mixed block presentation in which both 1:1 and 2:1 mappings were interleaved within a single learning phase, and in which learners receive explicit evidence for alternative mappings (in contrast to the paradigm by in Ichinco et al., 2009), both Yurvosky, Yu, & Smith (2013) and Poepsel and Weiss (2016) found that learners were able to overcome this mutual exclusivity bias and form 2:1 mappings without the addition of contextual cues. Several factors may explain the improved learning of many-to-one mappings in mixed-block presentation. First, learners may be more apt to violate mutual exclusivity when provided with explicit evidence that it no longer holds. Accordingly, mutual exclusivity was evident in Ichinco et al. (2009) as learners could ignore evidence for multiple mappings and adhere to a single mapping that was consistent throughout familiarization. However, in Yurovsky and Yu (2013) and Poepsel and Weiss (2016) forming multiple mappings was necessary to accommodate all of evidence in a given learning trial. More specifically, learners may not become entrenched in certain mappings when learning of a first mapping is quickly followed by exposure to a second mapping. This idea finds some support from a statistical learning by Bulgarelli and Weiss

(2016), who presented multiple streams to segment. When learners were overtrained on a first stream and then exposed to a second stream, the authors found a primacy effect (i.e., only the first stream was learned); however, when learners were exposed to the second stream immediately after having acquired the statistics of the first, they were able to learn both. Further support is found in the "learning to learn" framework, in which learners who encounter frequent variability in rule structure during learning (here, reward rates in studies of animal learning) are better able to accommodate subsequent rule change relative to learners who initially experienced long periods of rule stability (Gallistel et al, 2001).

This logic regarding variable inputs may extend to life experiences outside of the laboratory. As noted above, bilinguals, who arguably encounter greater variability in their language input relative to monolinguals, also demonstrate processing differences across a range of cognitive tasks relative to monolinguals, such as selective attention, task switching, inhibitory control, phonological working memory, and word learning (Bialystok, 1999; Colzato et al., 2008; Bruck & Genesee, 1995; Kaushanskaya & Marian, 2009; Bialystok et al., 2003; Papagno & Vallar, 1995; Van Hell & Mahn, 1997). Correspondingly, bilingual experience has also been demonstrated to influence learning in a CSSL task containing 2:1 mappings (Poepsel & Weiss, 2016; based on Experiment 1). As evidenced in Chapter 2, bilingual experience may facilitate the acquisition of 2:1 mappings, such that bilinguals acquire more 2:1 mappings, and show earlier evidence of learning 2:1 mappings relative to monolinguals. Our proposed explanation of this facilitation is greater flexibility in interpreting variability and positing the emergence of new inputs (explained below), which comes as a result of bilinguals' experience with the natural variability in their input. Specifically, bilinguals may more frequently encounter switches between statistically distinct distributions relative to monolinguals, and even in the absence of such switching, continuously contend with competition (e.g., in structure, phonology, word selection) that arises from having competition arising from the nonselective activation of multiple languages (Gollan & Kroll, 2001; Jared & Kroll, 2001; Costa & Santesteban, 2004; Bialystok, 2005; Kroll, Bobb, & Wodnieka, 2006; Schwartz, Kroll, & Diaz, 2007; Blumenfeld & Marian, 2007). This nonselective access and the need to manage two language continuously is hypothesized to tune the brain networks for cognitive control and executive function (e.g., (Abutalebi et al., 2012; Gold, Kim, Johnson, Kriscio, & Smith, 2013), and in this way may contributes to the greater flexibility that bilinguals seems to exhibit when learning in environments that contain multiple mappings. As noted, a sizeable body of research indicates that frequent exposure to variability both increases a learner's prior expectation of encountering subsequent variability (e.g., Gallistel et al., 2001; Qian et al., 2012; Qian, Jaeger, & Aslin, 2015) and can reduce or even eliminate processing costs attributed to encountering unexpected structures or events, such as those that occur at a point of transition between two inputs (e.g., Fine, Jaeger, Farmer, & Qian, 2013). Taken together, it follows that bilingual experience may serve to shift the amount of variability that learners must encounter before positing non-stationarity (i.e., the presence of multiple underlying causal models) in order to avoid conflating multiple distinct statistical patterns. Specifically, bilinguals may react more quickly than monolinguals to increased variability and prediction-error by positing a context-shift and forming a new representation.

Experiment 1 above presented 2:1 mappings in which evidence for each mapping was equal (i.e., in each familiarization, participants saw two scenes that supported mapping A and two that supported mapping B). To the best of our knowledge, all studies of CSSL with multiple mappings following the Yu and Smith (2007) CSSL paradigm present equivalent evidence (i.e., the same number of trials in which a word-object mapping appears) for each mapping. However, such balanced evidence for many-to-one mappings may not frequently be encountered in natural learning environments (e.g., Quine, 1960; Barrett, 1978; Merriman & Bowman, 1989; Medina et al. 2011). As such, one remaining area of inquiry in CSSL is determining how the acquisition of multiple mappings is influenced by variability in the amount of evidence supporting each side. Further, given the finding in Experiment 1 that bilinguals outperform monolinguals on 2:1 mappings when evidence for each side is equal, a remaining question is whether this advantage persists when the evidence for both sides of a 2:1 mapping is unequal. Experiment 1 found evidence that bilinguals may begin to show sensitivity to 2:1 mappings with sparser evidence. That is, while we did not directly manipulate the amount of evidence presented for each mapping, training was spaced over three learning phases, and we found that monolinguals required more training in order to achieve the same level of performance as bilinguals. Experiment 2 seeks to provide a more direct test of the hypothesis that bilinguals learn with sparser evidence (i.e., form multiple mappings based on weaker evidence relative to monolinguals), by directly manipulating the amount of evidence supporting each side of a 2:1 mapping.

A related issue pertains to the type of learning strategy employed by learners. Most CSSL experiments assume that learners closely track and aggregate co-occurrence statistics of words and objects across scenes, so that many possible relationships between words and objects are represented at a given moment (e.g,, Siskind, 1996; McMurray et al., 2012; Yurovsky et al., 2013; Yu & Smith, 2007; Xu & Tenenbaum, 2007). However, tracking all of the relationships encountered across time may not be cognitively plausible as a means of overcoming indeterminacy (Medina, Snedecker, Trueswell, & Gleitman, 2011). An alternative account, evidenced in studies of both children and adults, suggests that learners may be able to form only a limited number of mapping hypotheses, typically for pairings that are the strongest or of highest frequency (Medina et al., 2011; Trueswell, Medina, Hafri, & Gleitman, 2013; Woodard, Gleitman, & Trueswell, 2016). In the strongest version of this account, learners propose only one mapping conjecture for each word or object at a time. If subsequent evidence confirms this conjecture, it is retained; if evidence disconfirms it, a new mapping conjecture is made. These two accounts of learning make very different predictions for learning in a CSSL paradigm, especially in the context of increased mapping ambiguity. Namely, closely tracking the statistics predicts that learners will show sensitivity to a range of possible mappings between words and objects (which has been shown in the context of a fast-mapping study explained in greater detail below; e.g., Vouloumanos, 2008), while single-conjecture learning predicts sensitivity only to the strongest mappings. Because the majority of CSSL studies present deterministic word-object pairings (i.e., 1:1) to learners, their results leave to question which account more accurately describes learning in a specific task.

Although no CSSL studies have directly compared these accounts of learning in the context of high mapping ambiguity, this has been undertaken in a fast mapping paradigm. In Vouloumanos (2008), learners viewed a series of pairings between words and objects, presented individually (i.e., one word and one object per trial). Over the course of the experiment, learners were exposed to deterministic pairings of a single novel word and a single novel object, as well as probabilistic pairings in which two words were paired with a single object. These 2:1 pairs occurred probabilistically with one mapping occurring either 80%, 60%, 20% or 10% of the time. Thus, the resulting input contained both high and low strength 2:1 mappings. At test, learners completed a series of 2AFC trials, in which they heard a single word and saw two objects, and were asked to decide which object went “best” with the word. If learners could propose and track only a limited number of possible word mappings for each word or object (consistent with the idea of single-conjecture learning), the author hypothesized that learners would acquire only the highest-strength mappings and show little sensitivity to alternatives. On the other hand, if learners were able to closely track co-occurrence and so encode multiple sets of statistics for each word or object, they might show gradient sensitivity to many possible mappings (i.e., sensitivity in accordance with the amount of evidence that had been encountered for each mapping). She found that learners were able to encode both high and low frequency mappings, and that there was a positive correlation between the strength of the mapping and learners’ performance on that mapping, suggesting that learners were closely tracking the statistics of the input and representing mappings at all levels of strength (from 80% to 10%). With these results, the author concluded that word learning is not an all-or-nothing process (i.e., that learners acquire only the highest strength mappings and ignore evidence of gradience in mapping probability), and further suggested that constraints such as mutual-exclusivity might be applied in a more graded fashion, so that learners don’t fully ignore novel mappings to previously learned words or objects, but assign greater and lesser probabilities to mappings based on how frequently they appear. The suggestion here is that the local context with respect to mapping variability plays a role in how learners approach the task of word learning. This idea bears some similarity to the conclusion reached in Byers-Heinlein and Werker (2009) and Houston-Price et al. (2010), that mutual exclusivity is applied less rigorously by bilingual learners (who more frequently encounter violations of mutual exclusivity) compared to monolingual learners (who more frequently encounter evidence that accords with the predictions of ME).

Vouloumanos (2008) also suggests that real-world situations, as well as constraints on working memory and attention (e.g., Kareev, 1995; Turk-Browne et al., 2005) may render the complete tracking of statistics occasionally unviable. In a follow-up condition, Vouloumanos (2008) found that the degree of mapping variability in the learning environment seemed to change the amount of information about possible mappings that learners retained; less variability (i.e., an environment rich in high strength and deterministic word-object pairs) resulted in fine-grained sensitivity for both low and high-strength mappings, while high statistical variability (i.e., an environment containing predominantly low-strength mappings) resulted in a learning bias for high-strength mappings. Yurovsky and Frank (2015) also suggest that the complexity of the learning environment may factor in the amount of information about possible mappings that learners retain. Across two cross-situational experiments, the authors manipulated the complexity of the learning environment by changing the number of possible referents for each word. Specifically, on each trial, learners heard a single word and then, depending on the condition, saw either 2, 4, 6, or 8 objects on the screen (one of which was the correct mapping). In conditions that presented many mapping competitors on each trial, learners tended only to represent information about their most preferred mapping. When there were fewer competitors present, learners retained more information about the frequency with which words and objects co-occurred across trials. Together, these results indicate that learners' preferred strategy for disambiguating mappings may be directly tied to the degree of ambiguity that they encounter.

One possibility is that this pattern (i.e., of processing differently based on the variability in the learning environment) emerges independently of differences in cognition across learners, and so that all populations will show this effect. Alternatively, and as suggested by Vouloumanos (2008), differences in aspects of executive function, such as working memory and attention, may influence the emergence of this effect. An abundant literature on the cognitive consequences of bilingualism consistently finds that advantages for bilinguals in aspects of executive function, leading to the prediction that bilinguals may outperform monolinguals in contexts that present a high degree of mapping ambiguity. To date, however, these results come from studies that have left the language learning experience of their participants as an uncontrolled variable. An interesting question, then, is whether bilinguals react to mapping variability and environmental complexity in the same way as monolinguals. Because bilinguals often receive unbalanced exposure to each of their languages (e.g., Myers-Scotton, 2008; Bialystok et al., 2010), and thus unbalanced evidence for each side of a translation equivalent, they may approach to task of learning in environments with a high degree of mapping ambiguity differently than monolinguals. In addition, many studies have reported that bilinguals exhibit cognitive control advantages relative to monolinguals (e.g., Bialystok, 2005; Kroll, 2008; Costa et al., 2008). These advantages may impact statistical learning abilities to some extent, as certain aspects of cognitive control, such as working memory, attention, and inhibitory control, have been shown to play a critical role in learners' performance on statistical and sequence learning tasks (e.g., Kareev, 1995; Turk-Browne et al., 2005; Toro, Sinnett, & Soto-Faraco, 2005; Bartolotti et al., 2011; Awh, Vogel, & Oh, 2006; Engle, 2002). Although there is no reason to hypothesize that differences in cognitive control (arising from bilingual experience or natural variation between individuals) will influence performance on 1:1 mappings with respect to the two approaches to learning outlined above (as 1:1 mappings are always high strength and deterministic), these observations do suggest that bilingual experience may factor significantly in the learning of many-to-one relationships in complex environments. Specifically, the advantages in cognitive control associated with bilingualism, in addition to a bilingual's increased expectation of encountering non-mutually exclusive relationships, may facilitate the tracking and retention of lower-strength mappings. This accords with the findings of Experiment 1, in which bilinguals acquired many-to-one mappings faster and more proficiently than monolinguals, leading to the conclusion that bilinguals may be more flexible with respect to the ability to posit and track 2:1 statistical relationships.

In Experiment 2, learners were presented with one-to-one mappings and two-to-one mappings in a single familiarization phase. Two-to-one mappings varied from strong support for only one side of the mapping (i.e., encountering one mapping 7 out of 8 times) and conversely weak support for the alternative mapping (i.e., the second mapping occurred only once out of 8 times viewing the object) to equal support for two both sides of the mapping (four and four). Following familiarization, leaners were tested on their knowledge of both 1:1 and 2:1 mappings. Learning of 2:1 mappings is assessed in two ways: by performance on trials probing learning of either side of a 2:1 mapping; and by comparing the number of cases in which learners formed true 2:1 mappings (i.e., learned both sides of the mapping).

Experiment 2 thus extends the incremental association paradigm adopted by Vouloumanos (2008) to the CSSL paradigm. One benefit of this extension is that it provides a convenient way of examining how learners represent mapping variability in a more naturalistic learning environment (as learners rarely encounter balanced evidence for mappings). Much recent work has focused on the type of processing that best characterizes learning in the CSSL paradigm (Medina et al., 2011; Yurovsky & Frank, 2015). As summarized above, this work has investigated whether learners entertain the possibility of many alternative mappings at a given time, or prefer a more limited approach in which very few mapping conjectures (or possibly only one) are entertained at a given time. If learners in Experiment 2 are able to track statistics for many possible mappings, we predict a significant linear relationship between mapping strength and performance on trials probing 2:1 mappings, such that learning is positively correlated with strength (i.e., the frequency with which evidence for a mapping is encountered). This trend should be apparent not only in the aggregated data, but also for individual learners, as aggregation of data can produce a gradual learning curve even if individuals show trends indicative of single-conjecture learning (Yurovsky & Frank, 2015). If the complexity of the learning environment presents too much information for participants to closely track, we predict, as in Vouloumanos (2008), that learners may only encode the highest-strength mappings (such as the deterministic 1:1 mappings and 2:1 mappings in which one mapping is heavily favored over the other), showing no sensitivity to infrequent mapping alternatives.

Additionally, this paradigm provides the opportunity to compare the performance of monolinguals and bilinguals in learning environments with systematic variability in mapping ambiguity, addressing the question of how bilingual experience influences the way learners contend with high degrees of mapping ambiguity. Experiment 2 thus extends the Vouloumanos (2008) study to a CSSL task, comparing monolinguals and bilinguals on a task that manipulates the strength of many-to-one mappings. Given that bilinguals learned more true 2:1 mappings in Experiment 1, and acquired sensitivity to trials probing either side of 2:1 mappings more quickly than monolinguals, we predict that bilinguals may also be more capable learners when confronted with highly unbalanced 2:1 mappings, outpacing their monolingual peers who may only acquire mappings at a higher threshold of evidence. Alternatively, the bilingual learning advantage evidenced in Experiment 1 may not equate to accepting unequal evidence for a 2:1 mapping in Experiment 2. That is, bilinguals may be better able to take advantage of unambiguous (i.e., balanced) evidence for the presence of many-to-one mappings, while uneven and highly ambiguous evidence for two-to-one mappings (such as when a majority of evidence supports one mapping) may require additional exposure or explicit cueing in order to be accommodated.

3.1 Participants

24 functionally monolingual English speakers (3 male) participated in Experiment 2 for course credit. Based on language history questionnaire data, these participants had a mean age of 19.2 years (SD = 2.2) and self-rated their proficiency in English at 10 (SD = 0) on the ten-point scale used above. Ten of these participants had been classroom learners of a second language, who began receiving instruction at a mean age of 11.8 years (SD = 3.2) and self-rated their L2 proficiency at a mean of 3.4 (SD = 1.8) on the same ten-point scale.

25 English-Spanish bilinguals (10 male) also participated in Experiment 2 for course credit. These participants had a mean age of 19.1 years (SD = 1.0) and self-rated their proficiency in English at 9.5 (SD = .8) on the ten-point scale used above. They began learning their L2 at a mean age of 10.7 (SD =1.7) years, and self-rated their L2 proficiency at a mean of 7.5 (SD = 2.1) on the same ten-point scale.

3.2 Stimuli

Stimuli for Experiment 2 consisted of a subset of the words and objects presented in Experiment 1.

3.3 Design and Procedure

Experiment 2 consisted of a single familiarization phase that was followed by a single test phase. During familiarization, participants learned ten 1:1 mappings and eight 2:1 mappings (i.e., a set of eight objects with two word mappings each) across forty-eight trials that presented three words and three objects each. 1:1 mappings were presented eight times. 2:1 mappings were also presented a total of eight times, but the frequency of each of side of the 2:1 mapping varied. Out of eight 2:1 word-object mappings, two were structured such that one word mapping was encountered seven times and the other word mapping only once (i.e., 7 to 1), two were 6 to 2, two were 5 to 3, and two were 4 and 4. The order of presentation was pseudo-randomized such that no word or object appeared multiple times within a single trial or in consecutive trials. All details of the visual and auditory array, as well as the timing of word and object presentation within each trial, were identical to those of Experiment 1.

Following familiarization, learners completed a 4AFC test consisting of thirty-six trials. In each test trial, participants saw four objects, one in each corner of the screen, and heard a single word. Three of these objects were distractors, while the remaining object was matched to the word. Each object was labeled with a number, and participants selected their answer by pressing the corresponding number on a keyboard. As in the familiarization phase, no object appeared multiple times within a single trial or in consecutive trials. 1:1 mappings were tested twice (for a total of twenty 1:1 trials), and each side of 2:1 mappings was tested once (for a total of sixteen 2:1 trials). Participants were given as much time as needed to make a response for each trial.

3.4 Results

In Experiment 2, monolinguals' average performance on 1:1 mappings was 43.8% (SD=20%), and average performance on trials probing either side of 2:1 mappings (across all levels of strength) was 36.1% (SD = 17%), both of which exceeded the level of chance (1:1 Mappings: $t(23) = 4.6$, $p < .001$; 2:1 Mappings: $t(23) = 3.2$, $p < .01$). The difference in performance between 1:1 and 2:1 mappings was also significant ($t(23) = 2.1$, $p < .05$). Bilinguals' average performance on 1:1 mappings was 49.8% (SD = 25%), and average performance on trials probing either side of 2:1 mappings was 36.8% (SD = 21%). Both of these also exceeded chance (1:1 Mappings: $t(24) = 5.0$, $p < .001$; 2:1 Mappings: $t(24) = 2.9$, $p < .01$). Bilinguals performed significantly higher on 1:1 mappings than 2:1 mappings ($t(24) = 4.3$, $p < .001$). A comparison of monolingual and bilingual performance on both 1:1 mappings and 2:1 mappings collapsed across all strengths showed no differences (1:1 mappings: $t(47) = .93$, $p = .36$; 2:1 mappings: $t(47) = .11$, $p = .91$).

We used an 8 (Strength of Mapping) by 2 (Group) ANOVA to investigate the factors that influenced mapping performance in Experiment 2. Strength of Mapping was a within-subjects factor, while Group (i.e., monolingual or bilingual) was a between subjects factor. There was a main effect of Strength of Mapping ($F(7,329) = 6.86$, $p < .001$, $\eta^2 = .13$), such that performance was positively correlated with strength (i.e., participants performed better on mappings that appeared more frequently in training). The Strength of Mapping x Group interaction did not reach significance ($F(7,329) = .96$, $p = .45$, $\eta^2 = .02$), nor did the between-subjects factor of Group ($F(1,47) = .03$, $p = .86$, $\eta^2 = 0$) suggesting that monolinguals' and bilinguals' overall learning performance did not differ in Experiment 2.

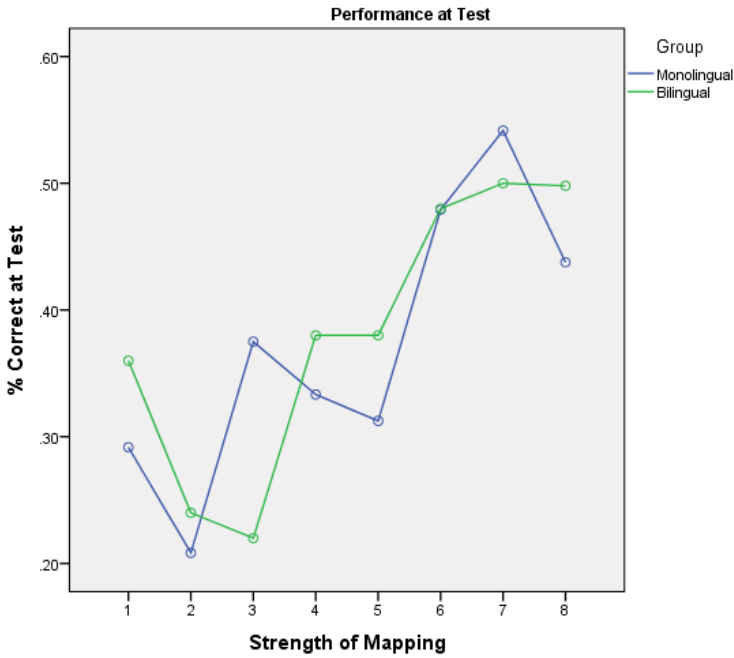

Figure 5 Performance for each group at each level of mapping strength.

The main effect of Strength of Mapping in the above ANOVA suggests a positive relationship between performance at test and the strength of a mapping. This finding was confirmed by a contrast analysis showing a significant linear term for the relationship between performance and strength ($F(1,384) = 26.38$, $p < .001$). To further determine whether this trend was reflective of the performance of individual learners or an effect of aggregating data, we correlated each participant's performance with mapping strength at each level of strength. The average correlation was .28 (SD = .36). Across all 49 participants in Experiment 2, 33 showed a positive correlation (i.e., Pearson's r greater than zero) between performance and mapping strength (M = .49, SD = .21). A binomial test on this proportion showed a significant result ($p < .05$), suggesting that significantly more participants than expected by chance followed the trend identified in the ANOVA above. Of the 33 participants who showed a positive correlation, 29 had a correlation above the group mean of .28, which was again significantly more than expected by chance (binomial test: $p < .01$). Taken together, these analyses suggest that a majority of participants followed the trend identified in the ANOVA and, further, that this relationship was robust.

An additional analysis focused on the point at which performance at a given threshold of evidence for trials probing 2:1 mappings significantly exceeded chance (see Table 2). Across all learners, this point was at four presentations of a mapping, suggesting that learners required at least equal evidence for both possible mappings in order to show significant evidence of learning. We performed this same analysis for each group, as well, in order to determine whether bilingual experience influenced the threshold of evidence required for developing sensitivity to 2:1 mappings. Monolinguals' performance became significant at six presentations for a mapping, while bilinguals' performance became significant at four. This result provides evidence that bilinguals may have been able to develop sensitivity to 2:1 mappings at a lower threshold of evidence compared to monolinguals.

Group	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
All	$t(48) = 1.5$ $p = .15$	$t(48) = .61$ $p = .54$	$t(48) = 1$ $p = .32$	$t(48) = 2.9$ $p < .01$	$t(48) = 2.2$ $p < .05$	$t(48) = 4.6$ $p < .001$	$t(48) = 5.4$ $p < .001$	$t(48) = 6.7$ $p < .001$
Mono	$t(23) = .62$ $p = .54$	$t(23) = .7$ $p = .49$	$t(23) = 1.7$ $p = .11$	$t(23) = 1.8$ $p = .09$	$t(23) = .94$ $p = .35$	$t(23) = 3.0$ $p < .01$	$t(23) = 4.4$ $p < .001$	$t(23) = 4.6$ $p < .001$
Bi	$t(24) = 1.4$ $p = .18$	$t(24) = .17$ $p = .87$	$t(24) = .59$ $p = .56$	$t(24) = 2.2$ $p < .05$	$t(24) = 2.1$ $p < .05$	$t(24) = 3.4$ $p < .01$	$t(24) = 3.3$ $p < .01$	$t(24) = 5.0$ $p < .001$

Table 2 Comparison of performance against chance (25%) for each mapping strength.

Reaction times to test trials were also collected in Experiment 2, and were examined using an additional 8 (Strength of Mapping) by 2 (Group) ANOVA. Strength of Mapping was a within-subjects factor, while Group was a between-subjects factor. Prior to analysis, we removed absolute outliers (RTs below 300 ms and above 10000 ms – i.e., the mean + 2.5 SD) from the data of all participants, which accounted for 1.9% of responses. There was a main effect of Strength of Mapping ($F(7,329) = 8.3$, $p < .001$, $\eta^2 = .15$) such that RTs (for both correct and incorrect responses) were negatively correlated with strength. There was no significant interaction between Strength of Mapping and Group ($F(7,329) = 1.2$, $p = .3$, $\eta^2 = .02$), but the between-subjects factor of Group did reach significance ($F(1,47) = 5.7$, $p < .05$, $\eta^2 = .11$), such that bilinguals' RTs were greater than monolinguals' across all mapping strengths (mono: M = 3668.3, SD = 192; bi: M = 4316.4, SD = 189.1). Planned follow up tests examined whether this difference was driven by RTs from correct or incorrect responses. For RTs from correct responses, the between-subjects factor of Group did not reach significance ($F(1,47) = 1.8$, $p = .18$, $\eta^2 = .04$; mono: M = 3738.3, SD = 198.9; bi: M = 4057.5, SD = 187.4), while it did for RTs from incorrect responses ($F(1,47) = 7.4$, $p < .01$, $\eta^2 = .14$; mono: M = 3737.4, SD = 223.9; bi: M = 4639.4, SD = 219.5), indicating that the significant between-groups factor in the main ANOVA was driven primarily by RTs from incorrect responses.

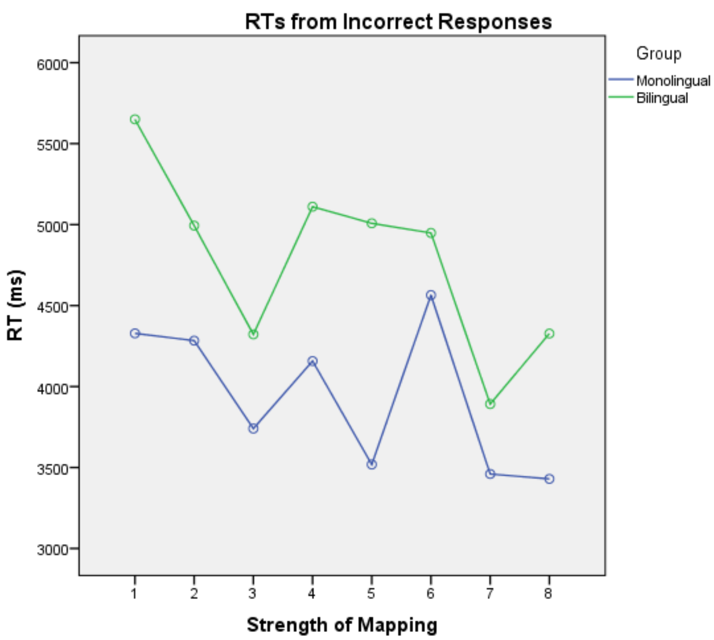

Figure 6 RTs from incorrect responses by group (monolingual vs bilingual).

We also examined the number of cases in which participants learned both sides of a 2:1 mapping. On average, monolinguals learned both sides of a 2:1 mapping in 13% (SD = 17.5%) of cases, and bilinguals learned both sides in 17% (SD = 20%) of cases. Across all mapping strengths, there was no significant difference in the learning of both sides of 2:1 mappings between monolinguals and bilinguals ($t(47) = .55$, $p = .58$). As noted above, 2:1 mappings in Experiment 2 were grouped into four bins that differed in terms of variability in the amount of evidence for each side of a mapping (i.e., 7:1, 6:2, 5:3, and 4:4). A further analysis compared monolingual and bilingual learning of both sides of 2:1 mappings within each of these bins (see Figure 7). There were no significant differences in performance between the groups when evidence for 2:1 mappings was unequal (7 to 1: $t(47) = .37$, $p = .71$; 6 to 2: $t(47) = .88$, $p = .39$; 5 to 3: $t(47) = .07$, $p = .94$). When evidence for a 2:1 mapping was equal (i.e., in the 4 to 4 bin), bilinguals learned both sides of the 2:1 mapping significantly more often than monolinguals (Mono: M = 8.3%, SD = 19%; Bi: M = 22%, SD = 25%; $t(47) = 2.13$, $p < .05$).

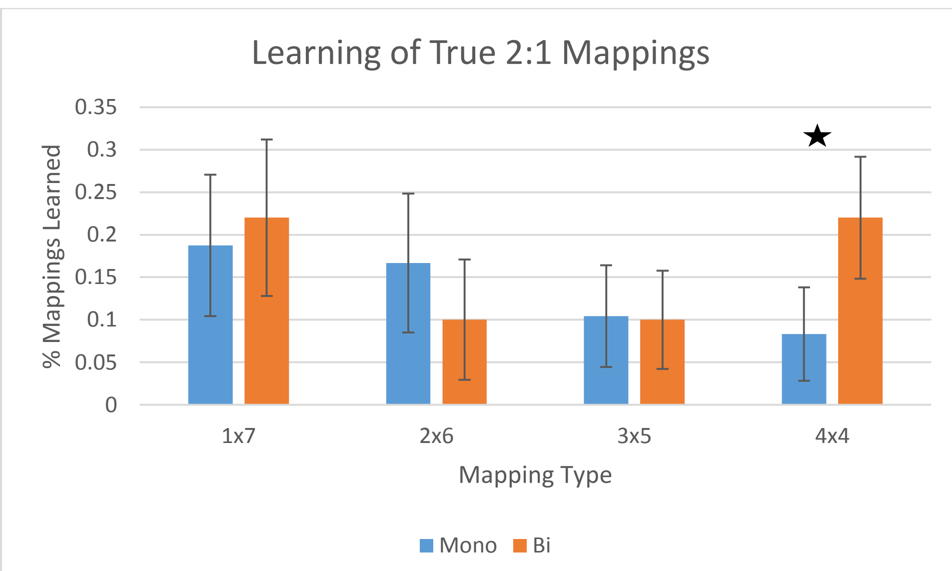

Figure 7 Frequency with which monolinguals and bilinguals learned both sides of a 2:1 mapping for each 2:1 mapping type.

3.5 Experiment 2 Discussion

In Experiment 2, functional monolingual and late-learning bilingual were familiarized to a mixed set of 1:1 (object:word) and 2:1 mappings in a single familiarization phase of a cross-situational statistical learning task. The 2:1 mappings varied along a continuum in the frequency with which either side of the mapping was supported during training, from high frequency for one side and low for the other (e.g., 7:1), to equal frequency for both (4:4). Participants then completed a 4AFC test, in which each 1:1 mapping was tested twice and each side of a 2:1 mapping was tested once. Overall learning for both 1:1 and 2:1 mapping was significantly above chance, indicating that learners were able to succeed in the CSSL task, even in the face of high mapping ambiguity. Notably, monolinguals and bilinguals were largely equivalent in their performance on both 1:1 and 2:1 mappings. One exception was that bilinguals significantly outperformed monolinguals when the evidence for both sides of a 2:1 mapping was equal (i.e., the 4:4 condition), a finding that paralleled the results reported in Experiment 1. Although direct comparisons of each groups' performance at each level of strength showed no significant differences, further analysis showed that bilinguals' performance on trials probing 2:1 mappings exceeded chance at a lower threshold of evidence compared to monolinguals (i.e., at 4 presentations of a mapping for bilinguals vs 6 for monolinguals).

One goal of Experiment 2 was to investigate the effect of variation in mapping strength for 2:1 mappings (i.e., the amount of evidence for either side of a 2:1 mapping) in a statistical word learning task. Here we compared competing theories of statistical learning using CSSL. One possibility is that learners are constrained in the types of statistics they can track, with retention only for mappings that reach a threshold of sufficiently strong evidence (see Vouloumanos, 2008). Alternatively, as has been reported for fast-mapping (Vouloumanos, 2008), statistical learning in the cross-situational context may contain a high degree of granularity, with retention of both high and low strength mappings. Across all learners, performance on trials probing 2:1 mappings was positively correlated with mapping strength, indicating that learners may be capable of tracking fine-grained statistics of word-object co-occurrences and retaining multiple mapping conjectures for each word or object at a given time. A significant linear contrast term for performance by mapping strength provided further evidence that learners closely tracked the co-occurrence statistics, bolstered also by the finding that a majority of individual learners in Experiment 2 evidenced a moderate positive correlation between performance and mapping strength. These findings address the concern of Yurovsky & Frank (2015) that aggregated data may indicate gradience in an effect (i.e., that averaging across learners may spuriously produce a smooth learning curve) while individual learners exhibit no such gradient behavior. Thus, these analyses strongly suggest that the results of Experiment 2 reflect learners' exquisite sensitivity to variation in mapping strength during cross-situational statistical learning and pose somewhat of a challenge for the idea of single-conjecture learning (outlined above) in CSSL paradigms (e.g., Medina et al., 2011).

A second goal of Experiment 2 was to extend the logic of Experiment 1 to the current paradigm by comparing the performance of functional monolingual adult participants to late learning bilinguals, both with respect to gradient effects (i.e., the extent to which they might learn mappings that occurred fewer times) as well as in the overall acquisition of both sides of 2:1 mappings (extending the paradigm of Experiment 1 to unbalanced mapping pairs). Given the results of Poepsel & Weiss (2016; Experiment 1) indicating that late learning bilinguals are more open to the possibility of 2:1 mappings in a CSSL task, we sought to explore whether a similar population of bilinguals might also more readily form 2:1 mappings from sparser evidence (such as when one mapping appeared less than half of the time) relative to monolinguals. On trials probing either side of 2:1 mappings, we found that bilinguals first achieved above chance performance at a lower threshold of evidence than monolinguals (i.e., at four presentations of a mapping for bilinguals compared to six for monolinguals), suggesting that bilingualism may facilitate the beginning stages of multiple mapping acquisition under conditions of high ambiguity. As in Experiment 1, we also looked at cases in which learners acquired both sides of a 2:1 mapping. Here, we found no differences between the groups when the evidence presented for 2:1 mappings was unequal (i.e., the 1:7, 2:6, and 3:5 mappings), but still a significant difference when evidence was equal (i.e., the 4:4 mappings), corroborating the findings of Experiment 1, and also demonstrating that bilingualism may not provide immediate learning advantages for 2:1 mappings in all contexts (i.e., when evidence is unequal) within this paradigm.

Reaction time measures provided another method of assessing performance in Experiment 2. Reaction times for responses at test were negatively correlated with mapping strength; as the strength of a mapping increased the RT for producing a response decreased. This correlation provides additional supporting evidence that the strength of a mapping has a gradient effect on performance, and further suggests that learners accrued statistical information for many possible mappings rather than tracking only high strength mappings. This correlation was found despite the fact that learners could take as much time as needed to respond, which further argues for the validity of the relationship between RTs and mapping strength (see similar unspeeded task effects in Pashler, 1989; Eskes, Klein, Dove, Coolican, & Shore, 2007). With respect to the effect of language experience, bilingual participants' RTs were significantly slower than monolinguals' across all mapping strengths. Further analysis showed that while there was a clear difference between monolingual and bilingual RTs for incorrect responses, there was no difference for correct responses. The lack of a difference in RTs for correct responses suggests that bilinguals weren't simply slower than monolinguals in rendering responses (i.e., if this were the case, the effect should exist regardless of whether a response was correct or incorrect). One possible, and speculative, explanation of this effect is that bilingual participants' slower RTs for incorrect responses may be indicative of greater sensitivity regarding the variability in the statistics of co-occurrence for 2:1 mappings in Experiment 2. This possibility will be discussed further below.

To summarize, the results of Experiment 2 provide evidence that adult learners are highly sensitive to co-occurrence statistics, and can make and retain multiple mapping hypotheses (as opposed to tracking only high-strength mappings), thereby extending the findings of Vouloumanos (2008) to a CSSL paradigm. Consistent with Vouloumanos (2008), our results suggest that statistical word learning of multiple mappings may not be an all-or-nothing process subject to the mutual exclusivity constraint (e.g., Markman, 1994). Rather, learners are able to form gradient associations between words and objects in accordance with how frequently they are evidenced in the input. This conclusion is broadly consistent with the results of several studies that have demonstrated that the application of mutual exclusivity is highly dependent on a learner's experience (Regier, 1996; Byers-Heinlein & Werker, 2009; McMurray et al., 2012). For instance, Byers-Heinlein and Werker (2009) tested monolingual, bilingual, and trilingual infants application of a disambiguation heuristic, which describes the tendency to associate a novel noun with a novel object instead of a familiar one. Monolinguals showed strong evidence of disambiguation, bilinguals showed intermediate evidence, and trilinguals showed no evidence. The authors interpreted these results as evidence suggesting that language experience, and in this case, the experience of mapping multiple words to one object, influences the development of disambiguation. The results further suggest an intimate relationship between language experience and development, as trilinguals showed less evidence of disambiguation than monolinguals. Thus, the more frequently a learner encounters evidence that multiple mappings exist, the less likely they are to learn in a mutually exclusive way or otherwise apply heuristics that function best with the stable input of monolinguals.

Experiment 2 both corroborates and extends the findings of Experiment 1 with regard to the effect of bilingual experience. Specifically, we have found evidence that bilingual experience facilitates the acquisition of true many-to-one mappings when evidence for both sides is equal (as in Experiment 1). Moreover, our findings with respect to performance on trials probing either side of a 2:1 mapping show that bilinguals may require a lower threshold of evidence in order to posit and begin acquiring sensitivity to multiple mappings. Experiment 1 provided indirect evidence for such a claim by showing that bilinguals outperformed monolinguals on trials probing 2:1 mappings at an earlier point in time (i.e., with less exposure); however, all 2:1 mappings in Experiment 1 were equally evidenced (i.e., analogous to the 4:4 mappings of Experiment 2), such that the locus of the bilingual advantage was unclear. Specifically, it could have been the case that bilinguals require less exposure than monolinguals to acquire multiple mappings (i.e., they could detect these multiple mappings based on overall sparse evidence such as only 2 instances of a mapping out of 8 total trials with a particular object). The findings of Experiment 2 are quite nuanced in that sense. On the one hand, the results demonstrate that neither group is able to acquire multiple mappings when the evidence is sparse within the confines of a single familiarization (such as a mapping that occurs twice out of a total of eight exposures). This hypothesis could only be tested by systematically varying the strength of the mappings. Nonetheless, if participants were provided with subsequent experience (i.e. another familiarization period), it is possible that differences between groups could arise, as the 4-4 mappings in Experiment 1 did increase over subsequent familiarizations (discussed further below). Consistent with the results of Experiment 1, however, the results of the 4-4 mappings do suggest that bilinguals are sensitive to sparser evidence for multiple mappings than monolinguals in the sense that they can acquire them significantly better than monolinguals in the variable learning environment of this paradigm (since bilinguals acquired them within the single session even given the additional variability relative to Experiment 1). The results of Experiment 1 suggest that, over additional exposure, monolinguals would likely perform at above chance, though likely still outpaced by their bilingual peers. Taken together, Experiment 2 lends more evidence to the claim that bilingualism facilitates learning in environments that contain multiple inputs, and also suggest that late L2 learners, despite being of lower proficiency and, in general, having less experience using and managing two languages, may still benefit from language learning experience as is often evidenced in highly proficient and early bilinguals (e.g., Bialystok and Martin, 2004; Carlson and Meltzoff, 2008; Kovacs & Mehler, 2009; Hernández et al., 2010; Soveri et al., 2010; Salvatierra and Rosselli, 2011).

The recently proposed extraction and integration framework (Thiessen, Kronstein & Hufnagel, 2013) may provide some explanatory power for the differences observed here, and in other experiments, between monolinguals and bilinguals. This framework allows for a more refined understanding of performance in statistical learning tasks by offering a two-process model of learning. Extraction depends on conditional statistics, and describes the process of binding together smaller units into chunks that function as exemplars for further processing. Integration depends on distributional statistics, and describes the process of inferring a central tendency, or similarity, across exemplars. In this way, extraction is a process that attends to local, smaller structures and outputs discrete representations (such as syllables and words), while integration operates across these discrete representations and provides sensitivity to cues and distributional information (such as the degree of similarity of smaller units and how they function together). Using the insights of this account of statistical learning, one hypothesis is that while extraction processes may not differ as a function of bilingual experience (that is, the fundamental task of SL may be unaffected by bilingual experience: e.g. Yim & Rudoy, 2013), integration processes might (see Poepsel & Weiss, 2016). That is, bilinguals may be better able to assess the statistical coherence of their input in order to determine the number of distributions from which it arises. In word-learning experiments that present multiple inputs, this might translate into bilinguals more rapidly concluding that mutual exclusivity no longer applies, such that the formation of multiple mappings can take place more quickly (as evidenced in Experiment 1) and with less evidence for a given mapping (as evidenced in Experiment 2). In speech segmentation tasks to date, this claim is not evidenced (for example, Bogulski, 2013 found no evidence that adult Spanish-English bilinguals outperformed English monolinguals in a sequential learning task). We will revisit this idea in Chapter 4 with a novel manipulation.

One challenge to the insights drawn from the extraction and integration model of statistical learning is that bilinguals in Experiment 2 did not outperform monolinguals in overall performance on trials probing either side of a 2:1 mapping, or in the learning of both sides of 2:1 mappings supported by unequal evidence. A possible explanation, alluded to above, is that the task itself was more difficult than Experiment 1, not only in terms of increased variability but also in overall less exposure. That is, learning in Experiment 1 was spread across three familiarization phases, while learning in Experiment 2 took place in a single familiarization. The findings of many studies suggest that distributed training and testing on new material (i.e., spread over several familiarizations), as opposed to massed training and testing, facilitates recall of the material and improves performance on measures of long-term retention (Izawa, 1966; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Karpicke & Roediger, 2008; Soderstrom & Bjork, 2014). This is known as the spacing effect. Spaced training has also been noted to improve the learning of more complex material (Carlson & Yaure, 1990; Rawson & Kintisch, 2005). Although the expectation is that both monolinguals and bilinguals would benefit from spaced as opposed to massed training, it may be the case that bilinguals' increased flexibility in positing the presence of multiple mappings renders the effect of spaced training more likely or faster to emerge. That is, with more opportunities to integrate the extracted units and thereby determine distributional similarity or dissimilarity (i.e., how many underlying representations are present in the input), bilinguals may be more likely to assume that mutual exclusivity does not apply and thereby outperform monolinguals. Additionally, experience learning and using multiple languages has been linked to improved ability to learn words' form-meaning links (e.g., Cenoz & Valencia, 1994; Keshavarz & Astaneh, 2004), as well as improved phonological working memory (e.g., Service et al., 2008; Majerus et al., 2008), which has also been shown to improve statistical learning of word forms (Misyak & Christiansen, 2007). Coupled with the finding that novel word learning is facilitated by frequent encounters with those words (e.g., Osterhout, McLaughlin, Pitkanen, Frenck-Mestre, & Molinaro, 2006), the hypothesis that bilinguals may benefit more from distributed training than monolinguals seems plausible. Taking all of this evidence together, a tentative conclusion, corroborated by the results of Experiment 1 as well as the finding here that bilinguals may show sensitivity to 2:1 mappings at a lower threshold of evidence, is that given further training on low and high strength mappings, bilinguals might begin to outperform monolinguals in acquiring true 2:1 mappings.

Overall, the increased variability in Experiment 2 clearly negatively impacted performance relative to Experiment 1, as well as previous studies. Yurovsky and Yu (2013), for instance, reported higher learning for both 1:1 and 2:1 mappings in a similar paradigm with less ambiguity (i.e., 2:1 mappings always had equal evidence for both sides). Similarly, Poepsel and Weiss (2016; Experiment 1) found learning of 1:1 mappings to be 20% above chance levels after a single familiarization in which learners were exposed to 4 instances of each mapping; this same level of performance on 1:1 mappings was found after exposure to 8 instances of each mapping in Experiment 2 (i.e., learners in Experiment 2 required twice the exposure to achieve a similar threshold of learning compared to learners in Experiment 1, although differences in the number of competitors at test in Experiments 1 and 2 should be noted). Yurovsky, Fricker, Yu, & Smith (2013) suggests that partial knowledge of word mappings facilitates the acquisition of other mappings by reducing the amount of mapping ambiguity present in the learning environment. That is, when learners achieve a certain threshold of knowledge for a subset of the mappings available in learning environment, the number of possible mappings for unknown words and objects is reduced, and the mapping task is made accordingly easier. As average performance on deterministic 1:1 mappings in Experiment 2 was low (below 50% for both groups), a further suggestion is that learners were unable to leverage their knowledge of these higher strength mappings in order to reduce the complexity of the task (and accordingly improve performance on lower strength mappings). Together, these findings suggest that learning in Experiment 2 was highly demanding, and that the difficulty of mapping under conditions of high ambiguity may have prevented us from finding an effect of previous language experience on the acquisition of multiple mappings.

We make a final speculative note regarding the RT differences between monolinguals and bilinguals for incorrect responses. It is possible that these findings are indicative of bilinguals possessing greater sensitivity to the presence of statistical ambiguity in an input relative to monolinguals. Our previous work has uncovered rather subtle effects confidence ratings reported by learners that can be altered even when overall performance on the retrospective task is unaffected (Poepsel & Weiss, 2014). So too, it is possible that these RT differences are highlighting a sensitivity that ultimately precedes learning, in much the same way as there are transitional knowledge states in other domains of language learning. For example, Goldin-Meadow and Wagner (2005) describe how gesture-speech mismatch (i.e., when what is expressed verbally conflicts with what is expressed gesturally) indexes a transitional knowledge state, in which a learner may be more ready to learn and so may exhibit larger benefits from additional, disambiguating input. Given the results of Experiment 1 that clearly show learning effects across additional exposure, these RT differences may forebode differential learning effects if additional input were provided. Future studies should address this issue as it may produce a more nuanced understanding of the differences between monolinguals and bilinguals, as well as an insight into how RT can be used to index learning in CSSL studies.

Chapter 4: The Influence of Structural Overlap and Bilingualism on Learning Multiple Inputs in Auditory Statistical Learning

For proficient language users speech is perceived as a series of discrete words, though in reality, speech streams comprise a continuous acoustic waveform. One task facing naïve learners of a language, then, is to discover where words begin and end, so that the next steps of language learning (such as mapping words to their referents) can occur (e.g., Saffran et al., 1996; Graf Estes, Evans, Alibali, & Saffran, 2007). There are many cues within a language that help learners detect word boundaries, such as prosody, phonetics, and suprasegmental factors (Jusczyk, 1997). However, these cues vary inconsistently across languages, and thus may not factor prominently in the initial stages of language acquisition (Cole & Jakimik, 1980). An invariant cue to structure, however, is the frequency with which different sounds co-occur (Harris, 1951). In this way, words can be thought of as sounds that frequently co-occur (i.e., sequences of sounds that are highly predictable), while the boundaries between words can be thought of as points where predictions about upcoming sounds are difficult or impossible to make (see Christiansen, Allen, & Seidenberg, 1998; Lu, 2010). Statistical learning describes the ability of learners to track these predictive relationships between smaller units of an input, and, as noted in Chapter 1, has been shown to be an important part of learning across a variety of domains, such as language (e.g., Saffran et al., 1996; Pacton, Fayol, & Perruchet, 2005; Maye, Weiss, & Aslin, 2008), vision (e.g., Fiser & Aslin, 2002; Hunt & Aslin, 2001), and non-linguistic auditory processing (e.g., Saffran, Johnson, Aslin, & Newport, 1999; Tillman & McAdams, 2004).

While initially proposed as a domain general learning mechanism spanning all modalities, recent evidence suggests that statistical learning may be a set of domain-general processing mechanisms that are subject to modality- and stimulus-specific constraints (e.g., Mitchel & Weiss, 2011; Frost, Armstrong, Siegelman, & Christiansen, 2015). For example, SL studies consistently find that transfer of knowledge of shared structure across modalities is either limited or non-existent (e.g., Redington & Chater, 1996; Tunney & Altmann, 1999; Conway & Christiansen, 2005; Conway & Christiansen, 2006; Emberson et al., 2011), and suggest that integration takes place in devoted multimodal networks subject to their own perceptual constraints (Mitchel & Weiss, 2011; Romanski & Hwang, 2014). Furthermore, Misyak & Christiansen (2012) suggest there is a great deal of individual variation within any given statistical learning task; for example, Saffran et al. (1997) and Saffran et al., (1999) reported wide ranges of learning of adjacent dependencies in their SL tasks, from an average lower bound of 40% to an average upper bound of 89%. They also reported a performance range from 40% to 97.5% for adjacent dependencies and between 30% and 100% for non-adjacent dependencies (a range similar to that reported in Gomez (2002) for non-adjacent dependency learning). Further, although test-retest reliability is high within a single modality, the same individual may show significant variation in their performance on SL tasks across modalities (Hsu & Bishop, 2010; Thiessen, Kronstein, & Hufnagel, 2013). Taken together, these findings support the conclusion of Frost et al. (2015) that SL is not a deterministic learning mechanism, but instead, one that may respond and develop differently as a function of both specific properties of the input as well as a learner's previous experience.

Given this stance, one limitation of many experiments to date has been the lack of distributional variability presented to learners. Specifically, studies have often focused on the learning of a single (and often invariant) input, which does not capture the multilingual experience (and arguably is an oversimplification of monolingual input as well; e.g., see Qian, Jaeger, & Aslin, 2012). Recently, however, several studies have begun to explore how learners contend with multiple statistically distinct inputs. Gebhart et al. (2009), for instance, exposed adult learners sequentially to two artificial language streams with no pause between the streams, overlapping by 50% in their syllable inventory, and with transitional probabilities as the only cue to underlying structure. When the two streams were presented in the same voice, learners showed a primacy effect, in which there was above threshold learning for the first stream encountered and chance learning for the second. This primacy effect has been replicated in a number of subsequent studies, several of which employ similar stimuli and variants of the methodology used by Gebhart and colleagues (e.g, Bogulski, 2013; Bulgarelli & Weiss, 2016), but also in studies that have adopted different methods and occur in the visual modality (e.g. Jungé, Scholl & Chun, 2007; Heimbauer, Poepsel, Bulgarelli & Weiss, in prep).

There are several different proposals accounting for this primacy effect. It is possible the first structure may not have been acquired with a high degree of confidence (i.e., learners have not reached low estimation uncertainty) by the time the second language is introduced and therefore the variance associated with the second language fails to trigger the formation of a second representation as it is interpreted as statistical noise (Aslin, 2014; Qian et al., 2012). However, a more recent study suggests that learners actually reach low estimation uncertainty rather quickly and fail to learn the second structure due to overlearning. That is, increased exposure to the same structure past the point of robust learning may cause an entrenchment effect where the first structure is learned and the second is not (Bulgarelli & Weiss, 2016).

The change blindness literature offers additional insights with regard to the possible underpinnings of the primacy effect, with numerous findings confirming that learners are typically unaware of unanticipated events, and noting that the likelihood of detecting a change depends on the focus of a learner's attention. Thus, when observers are attending to one event, they are less likely to notice an unexpected event (e.g., Rensink, 2002; Simons & Levin, 1997; Grimes, 1996; Scholl 2000). Further, change detection seems to be modulated by the similarity of the changing events, such that changes between similar events are more difficult to detect than changes between dissimilar events, leading to the inference of a similarity based constraint for change detection (Scott-Brown, Baker, & Orbach, 2000; Scott-Brown & Orbach, 1999; Williams & Simons, 2000; Zelinsky, 1998). While changes between discrete events are not completely equivalent to shifting patterns, this insight may nonetheless provide insights as to why primacy effects emerge and suggest that manipulating the similarity in patterns may yield parallel effects. For instance, this is noted by Gebhart and colleagues (2009), who hypothesize that the similarity of their language pairs, which shared 50% of syllables, may have hindered learning of the second of two languages. Franco, Cleeremans, and Destrebecqz (2011) also hypothesize that language similarity may influence learning. Also in a speech segmentation paradigm, they familiarized learners to two languages that shared 100% of their syllables. The languages were differentiated by a strong contextual cue (i.e., speaker voice), a condition in which Weiss et al. (2009) and Gebhart et al. (2009) had found produced equivalent learning of both languages. Although both languages were learned above chance, the authors found a strong primacy effect, which they attributed to interference caused by the high degree of syllabic overlap. Importantly, one suggestion of these studies is an essential link between statistical incongruence and overlap, such that greater overlap tends to result in greater interference. Thus, while primacy effects disappeared at 50% overlap with the addition of a contextual cue in Gebhart et al., (2009), the relatively greater interference at 100% overlap resulted in primacy despite a contextual cue in Franco et al. (2011). Although language pairings that share syllables and maintain statistical congruence are possible (see Weiss et al., 2009), such relationships reflect a high degree of experimental control, and are unlikely to be representative of statistical patterns across natural languages (i.e., sound patterns that are systematic within a language but are divergent across languages).

SL studies involving multiple inputs have used artificial languages with a limited range of syllable overlap. For example, in auditory SL, the languages used in Gebhart et al. (2009), Bogulski (2013), Zinszer and Weiss (2013), and Bulgarelli and Weiss (2016) shared overlap of 50% in their syllabic inventories. These studies all reported a primacy effect in their baseline conditions, and subsequently investigated ways of mitigating the effect. In Weiss et al. (2009), congruent languages (which were learned without the addition of contextual cues) had 0% overlap while incongruent languages (which were unlearnable when interleaved in the absence of contextual cues) had 25% overlap in syllable inventory. In a follow up condition, the authors found that congruent overlap could support learning, thereby focusing their conclusions on the role of congruence across both languages. As noted, Franco et al. (2011) used languages that overlapped 100% in their syllable inventory, and found a primacy effect, although both languages were learned above chance. Finally, in the domain of visual statistical learning, Heimbauer, Poepsel, and Weiss (*in prep*) found a primacy effect when visual streams sharing 50% of their elements were combined (similar to Gebhart et al., 2009). While all of these studies are informative with respect to how multiple speech streams may (or may not) be acquired, there is a gap with respect to understanding the role that syllable overlap plays in influencing these learning effects, as none of these studies offer a systematic exploration of this factor.

The idea that syllable overlap may significantly impact learning has roots in real-world language learning as well as the laboratory research mentioned above. For example, there is significant structural variability across natural languages at all levels, from the phonetic inventory of language (i.e., differences in the number and types of sounds a language uses) to suprasegmental characteristics, such that languages may share more or less structure at each level (Jusczyk, 1997). Shared or congruent structure between languages has been shown to facilitate learning through transfer (e.g., Ard & Homburg, 1992; Odlin, 1999), while assuming too much shared structure, especially when the overlap is incongruent, can have negative effects on learning (e.g., Jarvis & Pavlenko, 2008; Ringbom, 2006).

As noted earlier, relatively few studies to date have explicitly examined how statistical incongruence between two artificial languages influences their learnability, and of these, none have performed a systematic investigation of this factor, so that only single data points arising from differently structured languages exist. In one SL study, Perruchet, Poulin-Charronnat, Tillmann, and Peereman (2014) examined how shared structure (both congruent and incongruent) influences the learning of multiple structures that are encountered sequentially. After initial exposure to an artificial language in which words consisted of two syllables, learners were familiarized to a three-syllable language that contained all of the bigram syllables (i.e., the languages shared these syllables). In the control condition, the trigram language featured the shared syllables in combinations learners had never encountered (i.e., if the sequence A-B represents a word of the bigram language, where each letter corresponds to a unique syllable, learners might encounter the sequence B-C-A D-E-F in the trigram language). In the experimental condition, trigram word boundaries were the words of the bigram language (i.e., if C-D were a bigram word, learners might encounter the sequence A-B-C D-E-F in the trigram language). In this way, while no interference from prior learning was expected in the control condition, the mismatching statistical information in the experimental condition was expected to impede learning of the trigram language. These expectations were matched by the results from adult learners as well as simulations, suggesting that overlapping structure influences learning when it diminishes the predictive power of transitional probabilities. As described above, Weiss and colleagues (2009) demonstrated that statistically incongruent languages, when presented without an indexical cue (such as a change in speaker voice) were unlearned, as the incongruency of the shared syllables caused the statistical noise floor to increase, resulting in a decrease in the predictive power of transitional probabilities. In a follow up condition, the authors examined the effects of exposing learners to languages in which the shared syllables were congruent, to further investigate whether the effect of first finding was related to the statistical incongruence or the shared syllables. The results showed that the congruent languages with shared syllables were learnable, suggesting that statistical congruence and incongruence factor strongly in the degree to which a system of languages is learnable.

While these studies point to the importance of the type of overlap across languages (i.e., either congruent or incongruent), the overall degree of overlap of the languages in SL studies with multiple inputs has not been systematically varied. Consequently, the extent to which this factor, insofar as it results in statistical incompatibility, can influence learning of multiple inputs is largely unknown. For instance, while Gebhart and colleagues (2009) reported a primacy effect for two successive languages with incompatible statistics, Weiss and colleagues (2009) found no learning (although these languages were also interleaved many times). Notwithstanding the differences in stimulus presentation between the studies (longer exposures and a single switch in Gebhart et al., 2009; shorter exposures with multiple switches in Weiss et al., 2009), an unexplored question remains regarding how incongruent overlap in the phonetic language inventory impacts learning. The purpose of the present study is to fill this gap in our understanding by comparing learners in an auditory SL task in which we systematically vary the degree of syllabic overlap (resulting in statistical incongruence) in five language pairings, from 0% overlap to 100% overlap (with 25% increments in between). We employ the methodology of Gebhart et al. (2009), as previous experiments using this methodology have largely reported primacy effects. We anticipate that a high degree of incongruent overlap in the syllable inventory of two languages will correlate with lower learning performance overall and possibly unequal learning of languages (i.e., primacy or recency effects), as within-word and between-word transitional probabilities will converge such that they are no longer reliable indicators of where structures begin and end. When the degree of syllabic overlap is low (and the statistics of the languages are accordingly more congruent), within and between-word transitional probabilities will remain highly predictive of structure, and learners should accordingly be better able to acquire both inputs (and the inventory itself may serve as a contextual cue for the change in structure).

4.1.2 Multiple Inputs and Bilingualism

A secondary goal of Experiment 2 was to compare monolingual and bilingual learners in an auditory SL task presenting multiple inputs. The results of Experiments 1 and 2 above suggest that bilingualism may facilitate the resolution of statistical ambiguity associated with the interaction of multiple inputs within a single learning environment. Specifically, bilingual learners more quickly resolved mapping ambiguity in Experiments 1 and 2 in cases where equal evidence was presented for both mappings. Taken together, these findings suggest that bilingual experience may impact auditory statistical learning, particularly when the familiarization contains multiple inputs.

Despite the differences observed in the CSSL tasks, several previous segmentation statistical learning studies have compared bilingual learners with monolinguals without finding clear evidence of learning differences. Bogulski (2013), for instance, compared adult English monolinguals and high proficiency Spanish-English bilinguals in the task employed by Gebhart et al. (2009) and found no difference in performance between the groups (each showed a primacy effect when exposed to sequential inputs). Yim and Rudy (2013) compared Spanish-English monolinguals and bilinguals between the ages of 5 and 13 in auditory and visual SL tasks presenting a single input, and also reported no differences in learning between groups. Wang and Saffran (2014) compared monolinguals, tonal-language bilinguals, and non-tonal language bilinguals in a task that presented an artificial language with tonal cues to word boundaries. Non-tonal bilinguals showed greater sensitivity to the tonal cues than monolinguals though tonal bilinguals did not. Nonetheless, the results suggest some impact of previous language experience on statistical learning, though as the authors note, this task may require inhibition of the tonal cues to follow the statistics. Finally, Bartolotti et al. (2011) compared adult monolinguals and bilinguals of varying proficiency in a task that presented Morse Code languages that contained incongruent cues to word boundaries. Bilinguals outperformed monolinguals in a low-interference condition (in which only a single language was presented). Following the low-interference condition, participants were familiarized to a second language in which cues to word boundaries interfered with those in the first language. In this high-interference condition, there was no bilingual advantage, although the authors did find that performance on an inhibitory control task was positively correlated with learning. This finding accords with that of Wang and Saffran (2014) in the suggestion that greater executive control abilities may facilitate learning when language processing demands are high (e.g., in cases where knowledge of incongruent patterns must be suppressed in order for learning to occur).

As noted above, Perruchet et al. (2013) and Weiss et al. (2009) also demonstrate that statistical incongruence factors importantly in the learnability of two languages in that incongruence seems to increase the difficulty of acquiring a system of languages. Although these studies did not compare groups of monolinguals and bilinguals, together with the results of Bartolotti et al. (2011), they suggest that differences in the way that learners manage interference between languages may predict performance in tasks that present multiple statistically incongruent languages. Success in SL tasks presenting multiple inputs to learners, especially when the statistics of the languages are incompatible, may depend not only on the basic ability to track statistics, but also the ability to engage in higher-level processing over those statistics in order to manage interference. Statistical learning encompasses many forms of learning, and so it is possible that different types of statistical learning involve different cognitive processes. One attempt to formalize this kind of distinction has been proposed by Thiessen, Kronstein, and Hufnagel (2013), who differentiate between statistical learning tasks involving extraction and those involving integration. Extraction entails sensitivity to conditional statistics, and so involves holding several elements in working memory and binding them into a basic chunk (Perruchet & Vintner, 1998), whereas integration operates across these stored chunks and results in sensitivity to distributional statistics, such as similarity of exemplars or central tendency (see Erickson & Thiessen, 2015). For instance, infant learners may use language-universal cues, such as transitional probability, to extract words from speech initially (Graf-Estes et al., 2007). Integrating across these words may provide sensitivity to language-specific phonological patterns, such as stress, which may then supplant language-universal cues in subsequent learning (e.g., Thiessen & Saffran, 2003). In this way, a learner can develop sensitivity to specific patterns that hold across an input, which may be part of a constellation of cues that can be used to facilitate learning and manage variability.

These distinctions may provide a useful framework for interpreting and predicting the results of statistical learning studies comparing monolinguals and bilinguals. In Experiments 1 and 2 above, we employed CSSL tasks that involved integration (as information about word mappings is stored across trials to deduce the correct associations). Thus, it is possible that the bilingual advantage in forming 2:1 mappings, or otherwise acquiring multiple inputs, may be evidenced in processes involving integration (i.e., the tracking and accruing of information across time) but not those involving extraction. Thiessen and Erickson (2013) suggests that speech segmentation is best described as an extraction task, and the results of Bogulski (2013) suggest that bilinguals may show the same primacy effect as monolinguals when asked to track sequential regularities across two artificial speech streams presented consecutively (and lacking cues to the structural change). Given these findings, it may be the case that the bilingual advantage noted in Experiments 1 and 2 does not extend to speech segmentation tasks, even when multiple inputs are available to the learner. Alternatively, acquiring multiple inputs, especially in the case of incongruent overlap, arguably involves remapping (i.e., a process similar to that in which bilinguals were advantaged in Experiments 1 and 2), since learners must recognize that the predictive relationships of the first language may not apply in the second. One possibility, then, is of an interaction between bilingual experience and the degree of interference in the statistics of two inputs. That is, as interference increases (i.e., with increasing levels of syllabic overlap) and the need to learn new relationships between syllables across languages increases, a bilingual advantage in speech segmentation with multiple inputs may become more evident.

4.2.1 Participants

59 Penn State undergraduates (11 male; M: 18.8 years, SD = 1.3) in Introductory

Psychology participated for course credit, and 8 participants who indicated that they were early learners of an L2 were excluded from the present analysis. None had previously participated in a statistical learning study. Participants self-rated their English proficiency at 9.74 out of 10 (SD = .54). While functionally monolingual, 24 of these participants did indicate exposure to a language other than English. Second languages reported include Spanish (N = 15), French (N = 3), Japanese (N = 3), German (N = 2), and Chinese (N = 2). For these participants, average age of first exposure to an L2 was 12.9 (SD = 1.9), and average self-rated proficiency was 2.6 out of 10 (SD = .91).

4.2.2 Stimuli

The stimuli for Experiment 3 consisted of six artificial languages (Languages A-F), each of which had an inventory of four unique trisyllabic words (see Table 2). The languages were composed of 12 syllables, and within a language, each syllable occurred in a single position within a single word. Syllables were created using the AT&T Natural Voices speech synthesizer (http://www.naturalvoices.att.com) in a female American English voice ("Crystal"). Syllables were CV in structure and followed American English phonological constraints. In total there were 60 unique syllables, each of which was instantiated in four forms: a basic form with no co-articulatory properties on the vowel, and three additional forms wherein the vowel was co-articulated with /p/, /t/ and /k/ respectively. These syllables were hand edited for quality, sampled to 22050 Hz, normalized to 70 dB intensity, and edited to a duration of 250 ms with a pitch contour beginning at 160 Hz and falling 5% across the duration to end at 152 Hz. Inter-syllable duration for all positions was 30ms (of silence) to prevent duration becoming a cue to word boundary.

The four unique words of each language were concatenated into streams in which each word occurred 96 times (for a total of 384 words and 1152 syllables). The end of each syllable was coarticulated to the onset of the following syllable, such that coarticulation could not be used to determine the location of word-boundaries. The duration of each stream was 5 minutes and 22 seconds. The only constraint governing order of concatenation was that no word was allowed to follow itself.

The syllable inventories for each of the six languages were designed to overlap by varying degrees. Language A was the language to which every other language (i.e., B-F) was related. Languages A and B had 0% overlap in their syllable inventories (i.e., any syllable occurring in one language did not occur in the other), languages A and C had 25% overlap (i.e., 25% of the syllables in one language also occurred in the other), A and D had 50% overlap, A and E had 75% overlap, and A and F had 100% overlap. Languages B-F were constructed so as to be statistically incongruent in their combination with the base language, and, to reduce artificiality and the difficulty of the learning tasks, were designed not to be mirror languages (i.e., onset syllables in one language of a pairing did not exclusively appear as offsets in the other, and vice versa). Although syllables were shared across languages, these syllables never predicted the same within-word transitions (i.e., shared syllables were never preceded or followed by the same syllables across languages). Furthermore, we minimized the instances in which a shared syllable occurred in the same within-word position in both the base language and its paired language. Out of the 30 shared syllables across languages B-F, only 3 occurred in the same position in the base language A (one in medial position in the 25% overlap pairing, one in final position in the 75% overlap pairing, and one in initial position in the 100% overlap pairing). Finally, we examined the instances in which a word onset in one paired language appeared as a word offset in the other. This never occurred in the 0% and 25% overlap pairings, once in the

50% overlap pairing, three times in the 75% pairing, and twice in the 100% pairing. Thus, although in isolation each language had fully predictive within-word TPs and chance-level between-word TPs, in combination, the within and between word TPs became less reliable indicators of word boundaries (see Table 4 for precise TP values for each language combination).

	Lang A	Lang B	Lang C	Lang D	Lang E	Lang F
Word 1	to_ga_tae	cha_pu_ti	po_tu_da	pae_ge_du	pae_ki_to	ki_bo_te
Word 2	be_do_ka	go_bi_taj	pae_te_go	te_ki_go	bo_ku_pi	ga_be_gaj
Word 3	gaj_te_bo	dae_ko_pa	ta_ge_du	bo_tae_da	do_be_ka	pae_ka_to
Word 4	bu_pae_ki	pi_du_ke	bi_gaj_ku	bi_gaj_po	pa_gaj_bu	bu_tae_do

Table 3 Trigram languages presented in Experiments 3a and 3b.

Overlap	Trigram within-word mean/range	Trigram between-word mean/range	Bigram within-word mean/range	Bigram between-word mean/ range
0% (A+B)	100% (100% - 100%)	33.3% (34% - 32%)	100% (100% - 100%)	33.9% (42% - 25%)
25% (A+C)	81.3% (100% - 50%)	33.3% (34% - 32%)	87.5% (100% - 50%)	29.7% (42% - 13%)
50% (A+D)	71.8% (100% - 50%)	27% (34% - 16%)	77.1% (100% - 50%)	27.6% (66% - 13%)
75% (A+E)	70.8% (100% - 50%)	24.9% (67% - 16%)	58.4% (100% - 50%)	26.7% (67% - 13%)
100% (A+F)	51.8% (67% - 50%)	20.3% (67% - 15%)	54.3% (68% - 50%)	24.1% (68% - 13%)

Table 4 Summary of the collapsed transitional probabilities for each of the five language pairings in Experiments 3b and 3d.

4.2.3 Procedure

Experiment 3a consisted of a single familiarization phase followed by a single test. Each participant was exposed to one of the six language streams while seated in a sound attenuated booth and listening over headphones. Following familiarization, participants completed a 16-item 2AFC test in which the four words of the language they had listened to were exhaustively paired with a set of four partwords. Two of these partwords were of the “231” type, such that their onset and second syllable were drawn from the middle and coda syllables of a statistically-defined word, while their coda syllable was drawn from the onset of another word. The other two partwords were of the “312” type, such that their onset syllable was drawn from the coda syllable of a statistically-defined word concatenated with the onset and middle syllables of another word. On each test trial, one word and one partword were played in random order, separated by a 1s silence. Participants were asked to indicate (by circling the number “1” or “2” on a piece of paper) which stimulus sounded “more familiar” to them based on the preceding familiarization phase, and had a 5s window in which to produce a response before the next trial began. The order of the test trials was pseudo-randomized for each participant such that no word or partword was presented in consecutive trials. After participants finished the experimental task, they completed a questionnaire that gathered detailed information about their experiences learning languages other than English along with a self-proficiency rating.

4.2.4 Results

Learning of all six languages significantly exceeded the level of chance (which was 50% or 8 out of 16; see Table 5). Additionally, there were no significant differences in learning across the languages (one-way ANOVA: $F(5,53) = 1.18$, $p = .33$), with post-hoc comparisons (Bonferroni-corrected) of each possible language pair also finding no significant differences. These results suggest that all six languages presented in Experiment 3a were learnable in isolation, and further that performance was statistically equivalent across the languages.

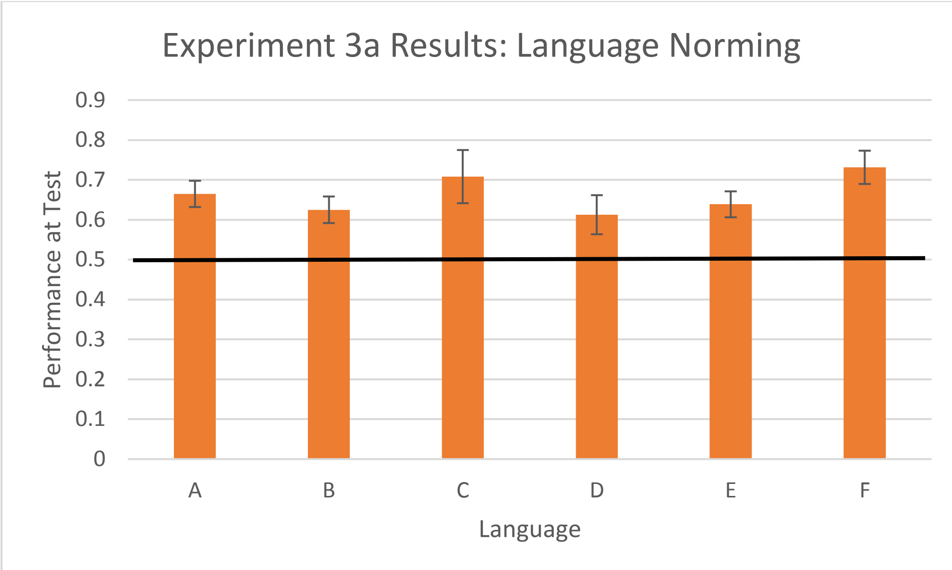

Figure 8 Results of normalization of the six trigram languages in Experiment 3a. All languages were learned above the level of chance (50%) in isolation.

Language	Mean	Comparison against chance
A	66% (SD = 11%)	$t(9) = 5.0$, $p < .01$
B	63% (SD = 11%)	$t(9) = 3.7$, $p < .01$
C	71% (SD = 20%)	$t(9) = 3.1$, $p < .05$
D	62% (SD = 15%)	$t(9) = 2.3$, $p < .05$
E	64% (SD = 10%)	$t(8) = 4.3$, $p < .01$
F	73% (SD = 13%)	$t(9) = 5.5$, $p < .001$

Table 5 Comparison of normalization data in Experiment 3a against chance (50%).

4.3.1 Experiment 3b

Having found that the six languages in Experiment 3a were learnable in isolation, Experiment 3b tested English monolinguals and English-Spanish bilinguals with consecutive languages that varied in the degree to which their syllable inventories overlapped (0%, 25%, 50%, 75%, 100%).

4.3.2 Participants

71 functionally monolingual Penn State undergraduates (11 male; M: 18.8 years, SD=.9) in Introductory Psych participated for course credit, with 8 additional participants who indicated early exposure to a second language excluded from the analysis. None had previously participated in Experiment 3a. These participants self-rated their English proficiency at an average of 9.8 out of 10 (SD = .63). Twenty-two of these participants also indicated a modest degree of late exposure to a second language. Second languages reported included Spanish (N = 14), French (N = 3), German (N = 2), and Chinese (N = 1), Hebrew (N = 1), and Italian (N = 1). For these participants, average age of first exposure to their L2 was 12.5 years (SD = 2.2), and their average self-rated proficiency was 2.7 (SD = .85).

68 English-Spanish bilinguals (23 male; M: 20.4 years, SD=2.5) also participated in Experiment 3b, and were compensated with course credit or payment, while an additional 8 participants who indicated early exposure to an L2 were removed prior to analysis. None had previously participated in Experiment 3a or any other statistical learning experiment. These participants self-rated their English proficiency at an average of 9.8 out of 10 (SD = .5). We also collected information about their L2. Average age of first instruction was 11.7 years (SD = 2.2), and their average self-rated proficiency was 6.43 (SD = 1.5).

4.3.3 Stimuli

In Experiment 3b, five language pairings were made by combining the six languages used in Experiment 3a. For each pairing, Language A and one of the other languages (i.e., B-F) were concatenated such that there was a single point of transition between the languages. There was no pause between the two languages aside from the standard 30ms of silence that occurred between all syllables and words within both languages. Additionally, for the purpose of counterbalancing, each pairing was created in both orders (i.e., A+B and B+A). The partwords used in the 2AFC test for each language pairing were controlled such that they did not inadvertently form words from either language.

4.3.4 Procedure

Participants in Experiment 3b completed a single familiarization phase followed by a single 2AFC test. As participants were exposed to two languages in Experiment 3b (in comparison to a single language in Experiment 3a), the duration of the familiarization phase was correspondingly doubled (from 5:22 seconds to 10:44) and the number of test trials was doubled, as well (from 16 to 32). Test trials for each language (i.e., A and B) were interleaved such that they alternated every trial. The order of this interleaving was counterbalanced across participants. The order of the test trials for each language was randomized. As a consequence of the design of the languages and constraints on presentation of test trials, no word or partword appeared in consecutive trials.

4.3.5 Results

The difference between the self-rated L2 proficiency for the functional monolinguals who indicated knowledge of an L2 and the English-Spanish bilinguals was significant ($t(91) = 10.9$, $p < .001$), indicating that bilinguals generally considered themselves more proficient L2 learners.

Participants in Experiment 3b were exposed to two languages in sequential order (hereafter L1 and L2). Performance on each of the languages at each level of overlap for the combined groups (i.e., monolinguals and bilinguals) is shown in Figure 9. Comparisons of performance on L1 and L2 at each level of overlap against chance (50%) are shown in Table 6. Paired sample t-tests showed a significant recency effect in the 25% overlap condition ($t(24) = 2.94$, $p < .01$), and no significant differences between performance on L1 and L2 in the other conditions (all *t*s < 1.54, all *p*s > .14).

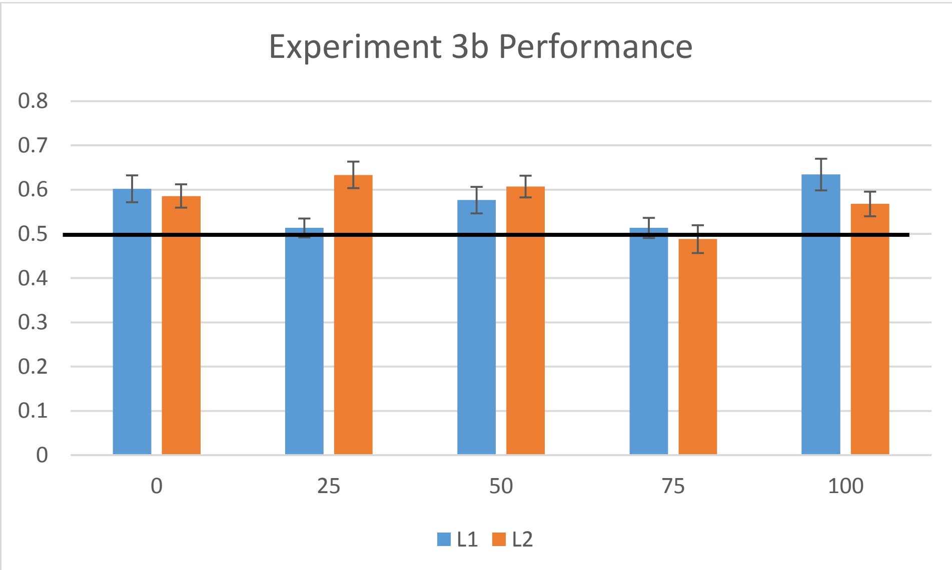

Figure 9 Results of Experiment 3b collapsed across all learners (monolingual and bilingual). Chance learning is 50%.

Degree of Overlap	L1 vs chance	L2 vs chance
0%	$t(26) = 3.3$, $p < .01$	$t(26) = 3.23$, $p < .01$
25%	$t(24) = .65$, $p = .53$	$t(24) = 4.44$, $p < .001$
50%	$t(28) = 2.56$, $p < .05$	$t(28) = 4.37$, $p < .001$
75%	$t(29) = .60$, $p = .55$	$t(29) = -.37$, $p = .71$
100%	$t(26) = 3.75$, $p < .01$	$t(26) = 2.43$, $p < .05$

Table 6 Comparison of performance on L1 and L2 at each level of overlap against chance (50%).

We used a 2 (Language) x 2 (Language Experience) x 5 (Overlap Degree) repeated measures ANOVA to investigate the factors that influenced learning in Experiment 3b. Language (L1 or L2) was a within-subjects factor, while Language Experience (English monolingual or English-Spanish bilingual) and Overlap Degree (0%, 25%, 50%, 75%, 100%) were between-subjects factors. The main effect of Language was not significant ($F(4,129) = .24$, $p = .63$), such that performance on L1 and L2 across all levels of overlap was equivalent (L1: M = .568, SD = .013; L2: M = .577, SD = .013). The two way interaction of Language x Overlap Degree was significant ($F(4,129) = 2.6$, $p < .05$). Planned contrasts exploring this effect showed that performance on L1 had a significant quadratic term ($F(1,134) = 7.33$, $p < .01$), so that performance was highest at the endpoints of the overlap spectrum and lower in the middle, and performance on L2 had a significant linear term ($F(1,134) = 3.95$, $p < .05$), indicating that performance decreased as overlap increased. The between-subjects factor of Overlap Degree was also significant ($F(4,129) = 5.15$, $p < .001$, $\eta^2 = .14$), indicating that performance collapsed across both languages varied significantly as an effect of degree of overlap. Planned contrasts here revealed a significant cubic term ($F(1,134) = 6.917$, $p < .05$), which reflects the movement of the combined performance scores for L1 and L2 at each level of overlap in Figure 12. The three-way interaction of Language x Language Experience x Overlap was marginally significant ($F(4,129) = 2.0$, $p = .09$, $\eta^2 = .06$), suggesting that monolinguals' and bilinguals' performance may have shown different relationships between language (L1 or L2) and overlap degree. Finally, the two-way interaction of Language x Language Experience, Language Experience x Overlap, as well as the between-subjects factor of Language Experience, did not reach significance, (all Fs < 1.43, all ps $> .23$).

An additional analysis examined the frequency with which participants at each level of overlap showed a primacy or recency bias, in order to determine the relationship between aggregate and individual performance. The performance of each participant was categorized either as primacy (learning of L1 > than learning of L2), recency (learning of L2 > L1) or equivalent (to account for all possible outcomes), and plotted in Figure 10. We used a two-sample z-test to compare the proportion of learners at each level of overlap showing a primacy or recency trend. This comparison was significant at 25% overlap ($z = 2.2$, $p < .05$), corroborating the finding above of a significant recency effect in this condition. The comparison did not reach significance at any other degree of overlap (all *z*s < 1.7, all *p*s > .1). Together, these results suggest that the performance of individuals mapped reliably onto aggregate performance, such that the results were not driven a subset of individuals with divergent learning patterns.

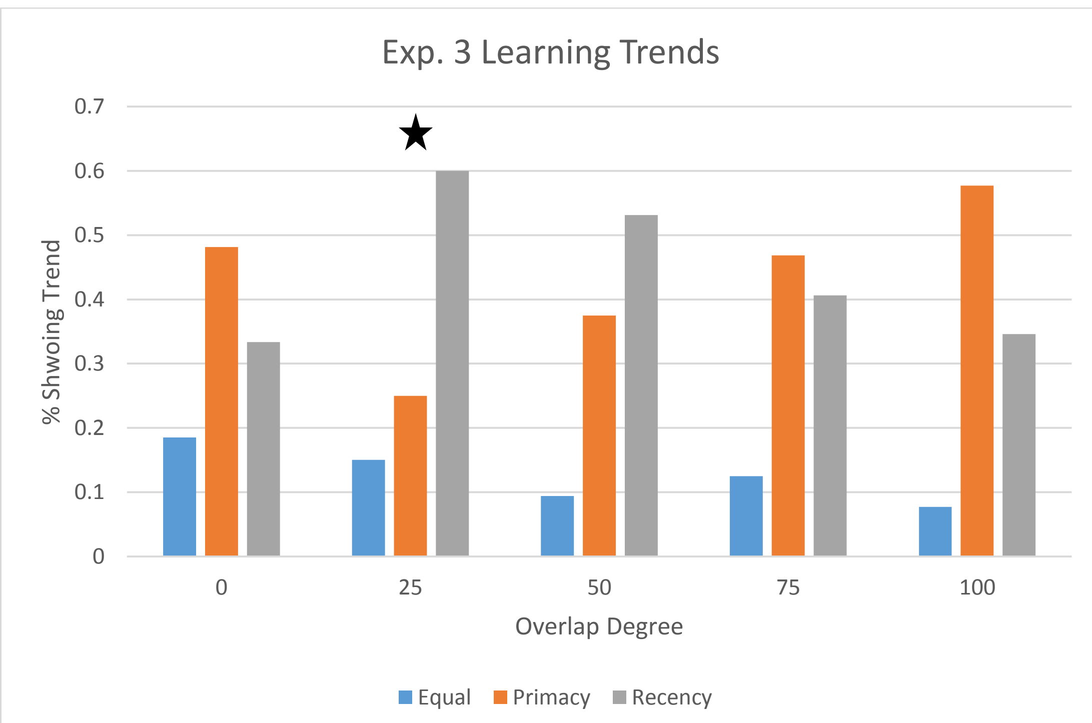

Figure 10 Proportion of learners at each degree of overlap showing a primacy effect, recency effect, or equivalent performance.

4.4.1 Experiment 3c

In Experiments 3a and 3b, learners were exposed to trigram languages that were difficult to learn (see Table 1). Consequently, in Experiment 3c, learners were exposed to bigram languages that were less statistically complex. This manipulation sought to further assess the influence of syllabic overlap by minimizing the difficulty of the statistical learning task itself. In Experiment 3c, we normalized 6 new bigram languages for later use in an experimental condition testing the effect of overlap in multi-language presentation.

4.4.2 Participants

Participants were 54 Penn State undergraduates (7 male; M: 19 years, SD = .97) who had not previously participated in any statistical learning experiments. These participants gave an average self-rated English proficiency was 9.75 (SD = .58). 37 indicated exposure to a language other than English, with a mean age at the onset of learning of 12.5 years (SD = 2.3) and an average self-rated L2 proficiency of 3.4 (SD = 1.2). Second languages reported included Spanish (N = 24), Italian (N = 4), French (N = 4), German (N = 3), Chinese (N =1) and Hebrew (N = 1). Five participants who indicated they were early learners of an L2 were excluded from the analysis.

4.4.3 Stimuli

Six new languages were created for Experiment 3c, each composed of four two-syllable words. Each language was constructed from eight unique syllables. The languages were related in the same way as those described in Experiments 3a-b: there was one base language that overlapped with the other five to varying degrees (from 0% to 100% in steps of 25%). As in Experiments 3a-b, languages were constructed to be statistically incongruent with respect to overlap. While syllables were shared across languages, these syllables never predicted the same transitions (i.e., within a word, shared syllables were never preceded or followed by the same syllables across languages). In the creation of the trigram languages, we attempted to minimize the instances in which a shared syllable occurred in the same within-word position across languages. Given the reduction of within-word positions in the bigram languages relative to the trigram languages, this constraint was more difficult to apply, such that in 7 out of 16 cases the shared syllable appeared in the same within-word position in language A and the paired language. Specifically, at 25% and 50% overlap, the shared syllables occurred in opposite within-word positions across the languages, while at 75% and 100% overlap, half of the shared syllables occurred in the same within-word position across languages, and half in opposite within-word positions. This feature of the languages occurred as a result of the principle design constraint of ensuring that no within-word syllable transitions were identical across languages, and the choice to reduce artificiality (and difficulty of learning) by not creating complete mirror languages (i.e., pairings in which all offsets in one language occur as onsets in the other, and vice versa). As in Experiments 3a-b, each word of a language appeared 96 times during familiarization, and languages were sequenced so that no word could follow itself. Aside from this constraint, the sequence of words for each language was generated randomly.

	Lang A	**Lang B (0%)**	**Lang C (25%)**	**Lang D (50%)**	**Lang E (75%)**	**Lang F (100%**
Word 1	gu_bo	da_po	mi_da	gaj_to	mi_du	du_to
Word 2	cha_du	ku_gaj	po_ki	be_gu	cha_gaj	gu_mi
Word 3	to_mi	pi_ju	gaj_be	mi_da	gu_to	cha_pae
Word 4	pae_ka	bae_ki	ta_gu	ka_pu	jo_pae	ka_bo

Table 7 Bigram languages presented in Experiment 3c and 3d.

4.4.4 Procedure

The procedure of Experiment 3c was identical to that of Experiment 3a, save for a decrease in the total duration of exposure to each language (i.e., from 5 minutes and 22 seconds in 3a to 3 minutes and 35 seconds in 3c) as a result of the use of bigrams instead of trigrams.

4.4.5 Norming Results

Average performance on each of the six languages created for Experiment 3c is shown in Figure 11. Learning significantly exceeded chance (50%) in all cases (Language A: $t(8) = 8.8$, $p < .001$; Language B: $t(9) = 8.4$, $p < .001$; Language C: $t(8) = 3.3$, $p < .05$; Language D: $t(8) = 3.5$, $p < .01$; Language E: $t(8) = 3.3$, $p < .01$; Language F: $t(7) = 6.7$, $p < .01$). A one-way ANOVA comparing performance across the six languages was not significant ($F(5,48) = 1.01$, $p = .42$), and post-hoc tests adjusted for multiple comparisons (Tukey's HSD) also showed no significant differences (all, qs < 2.42, all ps > .45), suggesting that there were no differences in learning of the individual languages in Experiment 3c.

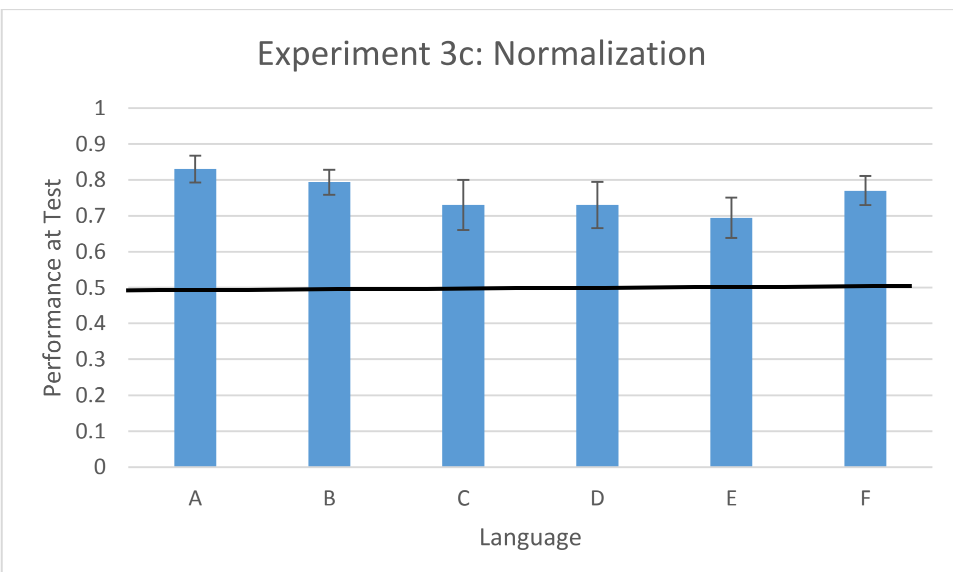

Figure 11 Average performance on each language presented in isolation in Experiment 3c.

4.5.1 Experiment 3d

In Experiment 3d, the languages normed in Experiment 3c were grouped into pairs that differed in the degree of overlap between them in five steps, from 0% to 100%.

4.5.2 Participants

Participants were 50 Penn State undergraduates (M: 19 years, SD = 1.3; 11 male) who had not previously participated in any statistical learning experiments. These participants gave an average self-rated English proficiency of 9.81 (SD = .66). 38 indicated classroom exposure to a language other than English, with a mean age at the onset of learning of 11.9 years (SD = 2.6) and an average self-rated L2 proficiency of 3.1 (SD = 1.2). Second languages reported included Spanish (N = 33), Italian (N = 2), French (N = 2) and Arabic (N = 1). Four participants who indicated they were early learners of an L2 were excluded from the present analysis.

4.5.3 Stimuli

Experiment 3d combined the languages normed in Experiment 3c. There were five language pairings that corresponded to the five degrees of overlap (A+B: 0%, A+C: 25%, A+D: 50%, A+E: 75%, A+F: 100%). As in Experiments 3a-b, the order in which participants encountered the paired languages was counterbalanced. The collapsed within and between-word TPs for the five language pairings, along with ranges, are presented in Table 4. Finally, the partwords used in the 2AFC test for each language pairing were controlled such that they did not inadvertently form words from either language.

4.5.4 Procedure

The procedure of Experiment 3d was identical to that of Experiment 3b, save for the replacement of the trigram languages with the bigram languages that were normed in Experiment 3c. Given the shorter duration of bigrams relative to trigrams, the total duration of the familiarization period was cut to approximately 6 minutes.

4.5.5 Results: Combined Bigram Languages

Average performance on L1 and L2 by degree of overlap is shown in Figure 12, and comparisons of performance on L1 and L2 at each degree of overlap against chance (50%) are shown in Table 8. Significant primacy effects were found at 25% overlap ($t(10) = 2.3$, $p < .05$) and 75% overlap ($t(9) = 2.4$, $p < .05$). Comparisons of performance on L1 and L2 at other levels of overlap were not significantly different (all *t*s < 1.4, all *p*s > .2).

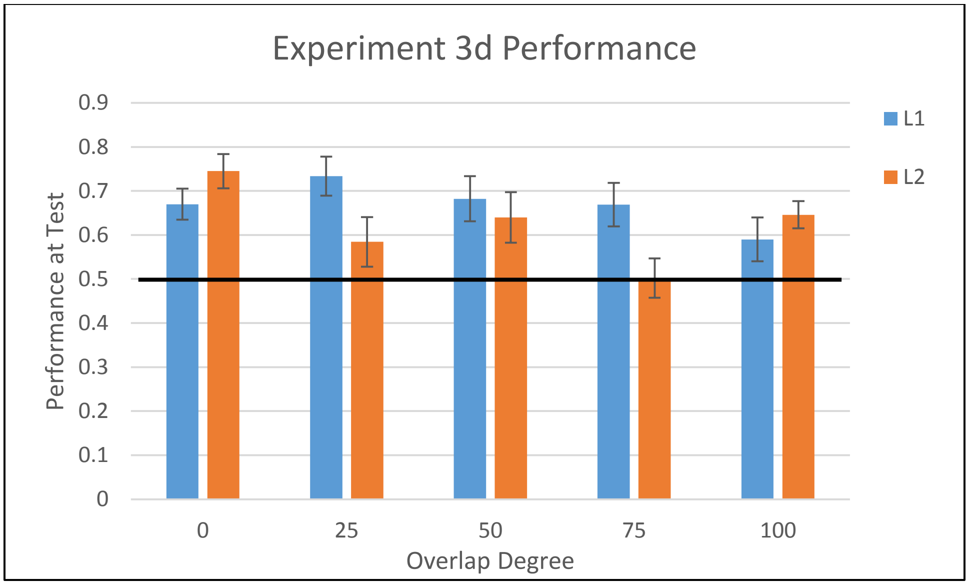

Figure 12 Learning by degree of overlap and language (L1 or L2) in Experiment 3d. Chance performance is 50%.

Degree of Overlap	L1	L2
0%	$t(9) = 4.8$, $p < .01$	$t(9) = 6.3$, $p < .001$
25%	$t(9) = 5.2$, $p < .001$	$t(9) = 1.5$, $p = .16$
50%	$t(10) = 3.6$, $p < .01$	$t(10) = 2.4$, $p < .05$
75%	$t(8) = 3.4$, $p < .01$	$t(8)$ .05, $p = .97$
100%	$t(9) = 1.8$, $p = .1$	$t(9) = 4.7$, $p < .01$

Table 8. Comparison of performance on L1 and L2 in Experiment 3d against chance (50%).

We used a 2 (Language) by 5 (Degree of Overlap) repeated measures ANOVA to determine how changes in the degree of overlap influenced learning in Experiment 3d. Language (L1 or L2) was a within-subjects factor, while Degree of Overlap was a between-subjects factor. The main effect of language was not significant ($F(1,45) = 2.2$, $p = .14$, $\eta^2 = .05$), but the interaction of Language and Degree of Overlap was significant ($F(4,45) = 2.86$, $p < .05$, $\eta^2 = .2$). Planned contrast analyses explored this significant interaction, showing that while there was no significant linear or higher-order terms for performance on L1 (all Fs < 2.4, all ps < .15), there was a significant quadratic term for the relationship between L2 and Degree of Overlap ($F(1,45) = 5.4$, $p < .05$), suggesting that performance followed a U-shaped trend with respect to overlap. Thus, while learning of L1 did not seem to be influenced by overlap, learning of L2 tended to decrease as the degree of overlap increased. The between-subjects factor of Degree of Overlap was marginally significant ($F(4,45) = 2.14$, $p = .09$, $\eta^2 = .16$). A planned contrast analysis found a significant linear term for overall performance by degree of overlap ($F(1,45) = 6.45$, $p < .05$), such that overall learning decreased as the degree of similarity of the languages increased.

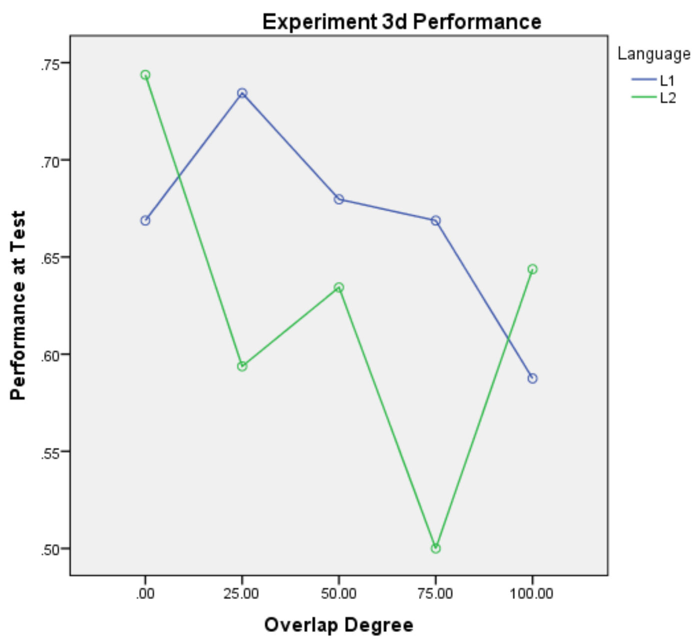

Figure 13 Performance in Experiment 3d as a function of overlap, split by language (L1/L2).

As in Experiment 3b, we examined the frequency with which participants at each level of showed a primacy or recency bias, as well as equivalent performance on L1 and L2 (see Figure 14). We used a two-sample z-test to compare the proportion of learners at each level of overlap showing a primacy (L1 > L2) or recency (L2 > L1) trend. This comparison was significant at 0% overlap (z = 2.2, p < .05) and 75% overlap (z = 2.7, p < .01), and marginally significant at 25% overlap (z = 1.8, p = .07), providing additional support for the significant primacy effects noted above at 25% and 75% overlap, and further suggesting that a majority of learners performed better on L2 than L1 at 0% overlap (despite the lack of a significant recency effect when the aggregated scores were compared in this condition). The comparison of learning trends was not significant at 50% or 100% overlap (zs < 1.4, ps > .15).

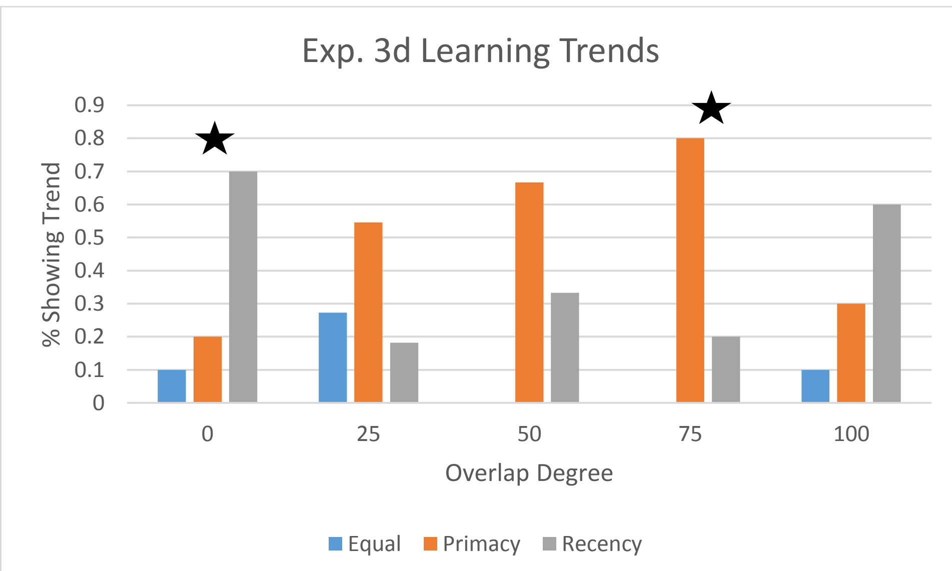

Figure 14 Proportion of participants at each degree of overlap in Experiment 3d showing a primacy effect, recency effect, or equivalent performance on L1 and L2.

4.5.6 Comparison of Performance on Trigram and Bigram Languages

A final analysis compared the performance of participants in Experiments 3b and 3d to determine whether differences in the structure of the languages (i.e., trigram vs bigram) learning. For this analysis, we used a 2 (Language) x 5 (Overlap Degree) by 2 (Language Structure) repeated measures ANOVA. Overlap Degree and Language Structure (Bigram/Trigram) were between-subjects factors, and Language (L1/L2) was a within-subjects factor. The between-subject factor of Language Structure was significant ($F(1,179) = 19.15$, $p < .001$, $\eta^2 = .1$), reflecting the generally higher learning for bigram languages compared to trigram languages (Bigram: M = .65, SD = .01; Trigram: M = .57, SD = .01). The between-subjects factor of Overlap Degree was also significant ($F(4,179) = 4.74$, $p < .01$, $\eta^2 = .1$), and a significant linear contrast term ($F(1,184) = 5.0$, $p < .05$) again indicated that performance tended to decrease as overlap increased, here across both bigram and trigram languages. The three-way interaction of Language x Overlap Degree x Language Structure was also significant ($F(4,179) = 4.04$, $p < .01$, $\eta^2 = .08$), such that the trends of performance on L1 and L2 by overlap degree were reversed across the bigram and trigram languages (see Figures 13 and 15). The main effect of Language, and the two-way interactions of Language x Overlap Degree, Language x Language Structure, and Overlap Degree x Language structure did not reach significance (all Fs < 2.2, all ps < .14). The results of this analysis suggest that while overlap between inputs resulting in statistical interference is, in general, negatively correlated with performance, the specific consequences of overlap for performance on L1 and L2 may be affected by language structure and degree of difficulty. This will be explored in greater detail in the Discussion.

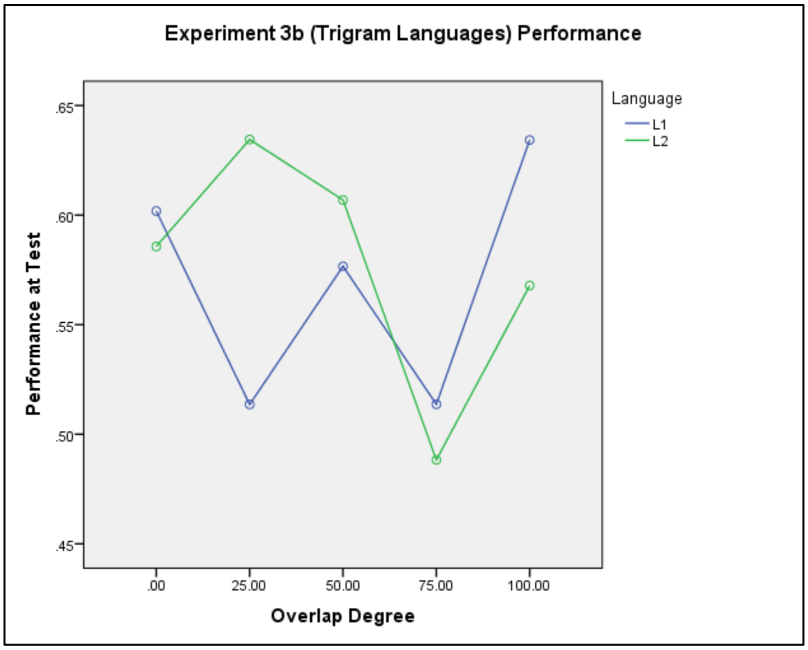

Figure 15 Performance in Experiment 3d as a function of overlap, and split by language (L1/L2).

4.5.7 Experiment 3 Discussion

In Experiments 3a-b, functionally monolingual adults and late-learning English-Spanish bilinguals were familiarized sequentially to two artificial trigram (i.e., each word was composed of three syllables) languages, L1 and L2, that varied in the degree to which their syllable inventories overlapped (i.e., 0%, 25%, 50%, 75%, 100%). Given our conjecture that overlap and congruence are highly correlated features of naturalistic language pairings, the languages were designed so that as the degree of overlap between languages and their statistical incompatibility increased in tandem. That is, as overlap between two languages increased, the collapsed transitional probabilities became increasingly less informative with respect to the location of word boundaries. After familiarization, participants completed a 2AFC test in which the words from each language were pitted against part-words. Across all participants, performance on L1 followed a U-shaped trajectory across the five levels of overlap, so that learning was highest at the endpoints (i.e., 0% and 100%), and lower at the intermediate values. Performance on L2 followed a linear trend, so that learning was highest when the degree of overlap was low (0%), and lowest when the degree of overlap was high (100%). We also investigated the specific learning patterns at each level of overlap across all participants and found that learning was mixed in a rather unprincipled way. For some degrees of overlap, both languages were acquired above chance (0%, 50%, and 100% overlap), whereas no languages were learned at 75% overlap. At 25% overlap a recency effect was found where the second language was learned and the first was not. Further analysis of the learning trends for individuals on L1 and L2 at each degree of overlap mapped reliably onto aggregate performance, suggesting that the results were not driven by small groups of individuals with divergent learning patterns. Unlike the CSSL studies above, we did not find significant evidence of a difference between monolinguals and late learning bilinguals.

In Experiments 3c-d, functionally monolingual learners were familiarized to bigram language pairings as a means of assessing whether simplifying the learning task would yield a different pattern of results. As in Experiment 3b, these pairings varied in the degree to which their syllable inventories overlapped in 5 steps (from 0% to 100%). Learning was again assessed with 2AFC tests which pitted the words of each language against part-words. Across all levels of overlap, learning of both L1 and L2 was significantly above the level of chance and equivalent (i.e., not significantly different). We found a significant effect of degree of overlap, so that overall learning (i.e., collapsed across L1 and L2) decreased as the degree of overlap increased. We also found a significant interaction between the language tested (L1 or L2) and overlap. Further analysis of this interaction showed that while there was no effect of overlap on learning of L1, learning of L2 was U-shaped in relation to overlap degree, as learning tended to be higher at the extremes of overlap and lower in intermediate levels. We again examined the specific patterns of learning at each level of overlap, finding primacy effects at 25% and 75% overlap, learning of both structures at 0% and 50% overlap, and a mixed effect at 100% overlap, in which performance on L2 but not L1 exceeded chance, despite no significant difference in performance in a direct comparison. Individual performance in Experiment 3d was generally consistent with aggregate performance, but not to the same degree as in Experiment 3b. This was likely an effect of the smaller sample size at each degree of overlap in Experiment 3d.

We also compared performance across Experiments 3b and 3d to determine if learning was influenced by differences in the difficulty of learning these languages in isolation, as Experiment 3b used trigram languages while Experiment 3d used bigram languages. Overall learning was significantly higher for bigram languages compared to trigram languages. Across both studies, we found that overlap degree was significantly negatively correlated with performance, suggesting that increasing statistical interference impeded learning. Furthermore, we found a significant interaction between the difficulty of learning the languages in isolation (i.e., bigram vs trigram) and the pattern of learning for L1 and L2 at a given level of overlap, such that the patterns were fully reversed as a factor of the difficulty of learning the languages in isolation. This finding parallels other results in speech segmentation studies with multiple inputs. For instance, the languages of Gebhart et al. (2009) were learned well in isolation (approximately 80%), and the authors found robust primacy effects when these languages were combined without contextual cues (or additional exposure to the second language). In our bigram languages (which were learned to a higher degree in isolation that the trigrams), we noted several primacy effects. One explanation of this effect is that the initial language is overlearned, causing learners to reduce their attention to the input and fail to accommodate the statistics of subsequent inputs. Bulgarelli and Weiss (2016) demonstrates this principle: when learners were exposed to five minutes of one language followed by five minutes of another, primacy effects resulted. When learners received less exposure to the initial language (and so experienced less entrenchment in the statistics of that language) learning of a subsequent input was facilitated. In Weiss et al. (2009), the incongruent languages were relatively more difficult to learn in isolation (a range from 62% - 73%), and the authors found catastrophic interference (no learning of either languages) when these languages were paired. In our trigram languages, which were learned less well than the bigrams, we found at least one interference effect and a recency effect, which may reflect destructive interference between the statistics of a first and second language (such that knowledge of the first language is eliminated).

One goal of Experiment 3 was to provide an investigation of how the syllabic overlap between two languages influences learning. Although a number of previous auditory SL studies on the learning of multiple inputs have employed languages that overlapped to greater or lesser degrees (e.g., Perruchet et al., 2014; Weiss et al., 2010), and whose collapsed statistics were correspondingly more or less reliable cues to word boundaries, none have undertaken a systematic investigation of how this factor influences learning. That is, these previous studies have used languages that vary significantly in structure (e.g., languages in Gebhart et al., 2009 contained 16 words created from specific vowel or consonant frames, while those in Weiss et al., 2009 consisted of 4 words with no such consonant frames), so that generalizing conclusions across studies may be difficult. Across two experiments using both bigram and trigram languages whose basic structural and phonological properties were highly similar, we found evidence that learning of two languages is significantly affected by language similarity that results in statistical incongruence. More specifically, we found that overall learning of two languages tends to decrease as the degree of incongruent overlap increases. The specific patterns of performance on

L1 and L2 varied between bigram and trigram languages, but significant contrasts were found for both the L1 and L2 of trigram languages and the L2 of bigram languages, suggesting that the relationship between performance and incongruent overlap tended to be both robust and predictable. As noted, the difference in the patterns of learning between bigram and trigram languages may reflect the degree of difficulty of learning bigram and trigram languages, as we found significantly higher performance on bigram languages in isolation. These findings were likely influenced by the phonological complexity of the languages as well as the degree of statistical interference caused by the syllabic overlap. For example, there were relatively more instances in the bigram languages of shared syllables occurring in the same position across a language pairing (7/16) than in the trigram languages (3/30), and bigram languages also had more instances of word offsets functioning as word onsets (and vice versa) across paired languages (13 across all bigram pairings vs 6 across trigram pairings). Nonetheless, a consistent finding here is that incongruent overlap significantly modulates the learning of multiple inputs.

Given that acoustic patterns across languages are highly variable (e.g., Christophe, Dupoux, Bertoncini, & Mehler, 1994), it has been proposed that learners must rely on language-universal rather than language-specific cues (e.g., Thiessen & Saffran, 2003). One real world example is the incongruent rhythmic structure of English and French, which has been shown to disrupt cross-language speech segmentation tasks in infants (Polka & Sundara, 2003). As such, our results afford a deeper understanding of how learners accommodate gradient levels of variability, and so contribute to a growing number of studies that have begun to investigate the effects of distributional non-uniformity reflective of natural patterns of language variability on learning in SL tasks (e.g., Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Ellis & O'Donnel, 2012; Kurumada, Meylan, & Frank, 2013).

An additional contribution of this work is the suggestion of greater complexity in the factors that influence the learning of multiple inputs in contingency learning studies, and specifically speech segmentation paradigms. The majority of previous work on this issue interprets findings with a binary criterion of whether learners showed significant learning of a first or second input or structure. In the domain of auditory SL, for instance, Gebhart et al. (2009) present evidence for a primacy effect in statistical learning that impedes learning of a second structure in the absence of adequate contextual cues (such as a pause and explicit instruction between inputs) or additional familiarization to the second structure. Using the same stimuli, this effect has been replicated by Zinszer and Weiss (2013), Bogulski (2013), and Bulgarelli and Weiss, 2016. A growing number of studies, however, have begun finding evidence that many factors may abate primacy effects in learning in environments that contain multiple inputs (in speech segmentation: Zinszer & Weiss, 2013; Bulgarelli & Weiss, 2016; Bartolotti et al., 2011; in CSSL: Yurovsky, Yu, & Smith, 2013; Yurovsky, Fricker, Yu, & Smith, 2015; Poepsel & Weiss, 2016). In the present experiment, we find evidence that performance in an environment containing multiple inputs does not always result in a primacy effect when contextual cues are not available, and may be related to the degree of statistical interference between two languages. For example, while the amount of variability in learning for L1 and L2 at any single level of overlap is high, our experiment captures the general trend that statistical incongruence is correlated with interference in learning (e.g., Franco et al., 2011; Perruchet et al., 2014). Although primacy effects were evidenced at several levels of overlap, we also found that learning of both languages is possible, and even that recency effects result from at least one language pairing. Much importance has recently been placed on explicit contextual cues as indicators of change in an input (e.g., Gebhart et al., 2009; Weiss et al., 2009; Poepsel et al., 2012), although implicit cues to change may factor just as importantly in acquisition (as not every change is explicitly cued – Qian et al., 2012). We provide evidence here that learners may be able to determine structural change by attending to the statistics of the input alone, and further that this ability may be independent of previous language experience or an increased prior expectation for a change (as no feature of the present experiment, aside from the structural change at midstream, would have served to indicate that an increased prior expectation was warranted or useful). Thus, our results suggests a complex relationship between the properties of a system of languages and how learning of that system proceeds.

A second goal of Experiment 3 was to compare the performance of adult monolinguals to late learning English-Spanish bilinguals. Critically, we included a novel and systematic manipulation of the degree to which multiple languages overlapped, so that Experiment 3 provided the first opportunity among speech segmentation studies to investigate how language overlap interacts with previous language experience. As noted in the Introduction, among the studies that have compared monolinguals and bilinguals, the findings have been mixed with respect to differences in the basic task of tracking statistics (e.g., Yim & Rudoy, 2013; Bartolotti et al, 2011; Nation & McLaughlin, 1986), and have largely supported the idea that bilingual experience does not affect the learning of multiple inputs (e.g., Bogulski, 2013; Bartolotti et al, 2011). Here, again, we find no significant evidence to support the assertion that bilingual experience facilitates the acquisition of multiple inputs in a speech segmentation paradigm. This finding accords with the general distinction posited by posited by Poepsel & Weiss (2016) that differences between bilinguals and monolinguals in SL tasks with multiple inputs may be affected by the nature of the task itself. That is, bilingualism may differentially affect SL of multiple inputs that is mediated by extraction processes (such as segmentation) versus those that involve integration (such as CSSL). While monolinguals and bilinguals do not differ in their basic ability to track statistics and form chunks from local input, these groups may differ in the extent to which they are able to integrate statistical information (i.e., determine similarity and central tendency) across time particularly when there are multiple underlying generating models (i.e., multiple languages).

In sum, the results of Experiment 3 demonstrate a relationship between statistical incongruence and the learning of multiple inputs, as well as the difficulty of learning an initial language and the emergence of primacy or interference effects. In investigating these issues systematically across several degrees of incongruent overlap, we find that primacy effects, in which learning of a first input impedes learning of a second, do not always result, and may be tied to specific properties of the stimuli. We found no significant effects of bilingual experience on learning, in accordance with several previous studies (e.g., Yim & Rudoy, 2011; Bogulski, 2013), and take this as additional evidence that a bilingual advantage in learning multiple inputs may be most likely to emerge in tasks that require a learner to integrate statistical information across time or learning environments (as in Experiments 1 and 2). One limitation of the present work, and indeed all studies to date employing multiple inputs, is that in all language pairings, the phonetic/phonological properties of the stimuli are highly similar (see Polka & Sundara, 2002, 2012; Sundara, 2011 for investigations of how these properties affect learning of a single input as a factor of language experience). A number of studies suggest that bilinguals may have improved language and word learning abilities relative to monolinguals as a result of enhanced phonological working memory (e.g., Service et al., 2002; Majerus et al., 2008; Adesope et al., 2010), such that identical properties across languages may bias null results in the comparison of monolinguals and bilinguals. One possible route for future work comparing these groups, then, is an investigation of how changes in the phonological or phonetic properties of stimuli across inputs influences learning (as one might expect bilinguals to more quickly accommodate such changes).

Chapter 5: General Discussion

More than half of the world's population is bilingual. For these learners, language acquisition entails developing sensitivity to the relationships and patterns that exist across several languages, and crucially, learning rules and regularities specific to each language (Werker, 2012; Weiss, Poepsel, Gerfen, 2015). Although a growing body of SL research has begun to investigate how learners perform in environments that contain multiple statistically distinct inputs, relatively few have asked how the experience of bilingualism influences learning in such environments. Given that bilingualism engenders a range of differences in performance on cognitive tasks across the lifespan (Bialystok, 2004; Prior & MacWhinney, 2010; Pettito et al, 2012; for review, see Li, Legault, & Litcofsky, 2014), an important question for SL research is whether bilingualism influences the way learners approach the task of tracking statistics. This is particularly apt for statistical patterns arising from the interaction of multiple inputs (i.e., environments that contain "bilingual-like" distributional information). To this end, we compared adult English monolinguals and late English-Spanish bilinguals in three statistical learning tasks (two cross-situational word learning tasks and one auditory statistical learning task), each of which presented the opportunity for learners to form multiple mappings.

5.1.2 Chapter 2 Summary

In Chapter 2, we tested whether bilingualism influences the acquisition of many-to-one word-object mappings in a cross-situational statistical learning (CSSL) task. In Experiment 1a, we assessed the ability of adult English monolinguals, late English-Spanish and Chinese-English bilinguals to engage in CSSL in a task presenting only one-to-one mappings. All groups succeeded in learning one-to-one mappings, and there were no significant differences between the groups, suggesting that the ability to successfully engage in the basic task of cross-situational learning did not differ as a function of language experience. In Experiment 2, learners were asked to form both one-to-one and two-to-one mappings, and were tested at three points during familiarization. Overall, monolinguals and bilinguals did not differ in their learning of one-to-one mappings, again suggesting that language experience does not affect the basic ability to track co-occurrence statistics across trials. However, bilinguals more quickly acquired sensitivity to two-to-one mappings, forming more true many-to-one mappings, and also exhibiting greater proficiency on these mappings than monolinguals. Based on these results, we conclude that bilinguals, when tracking novel statistical inputs over time, may be more flexible than monolinguals in their ability to posit multiple mappings in an input.

5.1.3 Chapter 3 Summary

In Chapter 3, we further investigated the finding that bilinguals may more readily accommodate multiple mappings by adding variability in the amount of evidence that learners encountered for each side of a 2:1 mapping. Thus, in Experiment 2, we used a CSSL paradigm in which learners (adult English monolinguals and late English-Spanish bilinguals) encountered both 1:1 and 2:1 mappings that varied in mapping strength (i.e., seven of one mapping versus one of the other, six of one versus two of the other, etc.) within one familiarization phase. Overall mapping performance for both groups was equivalent across all mapping types (1:1 and 2:1) and positively correlated with mapping strength for 2:1 mappings, suggesting that learners were sensitive to fine-grained statistical information within the task. An analysis of performance at each level of mapping strength for 2:1 mappings indicated that bilinguals first achieved above chance performance at a lower level of mapping strength relative to monolinguals. Additionally, bilinguals formed significantly more true 2:1 mappings (i.e., learned both sides of the mapping) than monolinguals when the evidence for both sides was equal (but not when evidence was unequal). These results corroborate the findings of Experiment 1 (in that bilinguals learned significantly more 2:1 mappings than monolinguals when evidence for each side was equal), and extend them, suggesting that bilinguals may be able to posit and acquire 2:1 mappings on the basis of weaker evidence, overcoming a statistically noisy learning environment to do so. These results also offer the first evidence that learners may show gradient sensitivity to mappings in CSSL tasks, and so bear on a currently lively debate on the nature of learning in CSSL tasks.

5.1.4 Chapter 4 Summary

In Chapter 4, we extended our investigation of bilingualism and the learning of multiple inputs to an auditory speech segmentation task, with the goal of examining how the degree of similarity of sequentially presented and statistically incompatible artificial languages influences learning, as well as assessing the influence of bilingual experience. In Experiment 3a, we tested adult English monolinguals of learning of six artificial languages in isolation. In Experiment 3b, we created five pairings of these languages that varied in the degree to which their syllable inventories overlapped (i.e., 0%, 25%, 50%, 75%, 100%), so that as the degree of overlap increased, interference in the collapsed statistics also increased. We tested adult English monolinguals and late English-Spanish bilinguals and found that across all participants performance was significantly influenced by incongruent overlap. Further, we demonstrated that primacy effects do not always arise in the case of sequential inputs that lack contextual cues. There were no significant differences between monolinguals and bilinguals in the learning of multiple inputs, however. Experiments 3c and 3d sought to further investigate how the degree of overlap influences learning by presenting participants (monolinguals) with easier languages. In Experiment 3c, we normed a new set of six bigram languages. In Experiment 3d, these languages were combined into pairings that shared the same properties of overlap as those in Experiment

3b. Results again showed that learners' performance was significantly influenced by overlap resulting in statistical interference, so that learning tended to decrease as overlap increased. We also found some evidence for primacy effects that varied with the amount of overlap.

5.2 Discussion

The results of this series of experiments address a number of gaps in the current SL literature regarding how learners contend with variability in the input as well as the influence of previous language experience. In two cross-situational learning tasks, we found consistent evidence of a bilingual advantage in the learning of many-to-one word-object mappings. Experiments 1 and 2 both represent novel investigations of the influence of bilingualism on statistical word learning, and contribute a new understanding regarding how bilingual experience influences the constraints that typically govern word learning (e.g., Werker & Byers-Heinlein, 2009; Houston-Price et al., 2010). By contrast inn Experiment 3, we compared monolinguals and bilinguals in a speech segmentation task with multiple inputs that varied in their degree of overlap. Consistent with previous studies (e.g., Bogulski, 2013; Bulgarelli & Weiss, 2016), we found no differences in learning across groups. This suggests that the bilingual advantage in learning multiple inputs noted in Experiments 1 and 2 does not extend to all SL tasks.

This discrepancy in findings between monolinguals and bilinguals supports the notion that SL may not be a unitary mechanism that operates independent of modality and task (e.g., Frost et al., 2015). Rather, SL may be closely tied to specific properties of both task and modality, and supported by a number of cognitive processes with sensitivities to different types of statistics. The results of Experiments 1 and 2 provide further evidence for such a claim. Specifically, learners showed no differences in the acquisition of 1:1 mappings, but bilinguals' learning was significantly higher for 2:1 mappings, indicating that certain processes underlying

SL may be differentially affected by bilingual experience. There have been several recent attempts to more precisely characterize the processes that underlie SL, which may provide some explanation of the effects noted in the experiments of this dissertation. For instance, Frost el al. (2015) theorizes that SL is composed of many different modality specific mechanisms for learning that are sensitive to many different types of statistics throughout an input. These different mechanisms are potentially supported by different neural networks, and so the output of statistical learning in different domains may partially depend on the function and connectivity of these networks. For instance, a number of neuroimaging studies find that visual statistical learning engages higher-level visual networks (such as the lateral occipital cortex and inferior temporal gyrus: e.g., Turk-Browne et al, 2009; Bischoff-Grethe et al., 2000), while auditory statistical learning engages a set of auditory networks (such as the left temporal and inferior parietal cortices, frontotemporal networks, and motor areas associated with speech production; e.g., McNealy et al., 2006). A number of brain regions and networks that are always active in SL processing have also been identified (e.g., the hippocampus and medial temporal lobe; Schapiro et al., 2014; as well as the basal ganglia and thalamus; McNealy et al., 2006; Poldrack et al., 2005), which serve to encode, bind, and consolidate representations that have been generated in modality or stimulus-specific networks (Turk-Browne et al., 2009; Cohen & Eichenbaum, 1993; Shohamy & Turk-Browne, 2013).

These findings suggest that SL may be composed of several sets of processes, one of which may be task or modality specific processing, which is followed by a more general integration and consolidation process. Indeed, several recent studies demonstrate that cross-modal information can be integrated in SL tasks to create a unified percept, providing further evidence that SL may be characterized as a set of modality-specific sub-systems that are components of an interactive network capable of performing operations across modality specific representations (Mitchel & Weiss, 2011; Emberson, Conway, & Christiansen, 2011; Mitchel, Christiansen, & Weiss, 2014). The extraction and integration model of statistical learning (Thiessen et al., 2013; Thiessen & Erickson, 2013; Erickson & Thiessen, 2015) similarly suggests that SL is composed of distinct, but partially overlapping, processes, through which learners are hypothesized to show sensitivity to different types of statistical information. One sensitivity is to conditional statistics, such as relationships between syllables that define words in a speech stream, which learners engage with in the process of extraction. Another sensitivity is to distributional statistics, such as determining the central tendency of many extracted exemplars to infer higher-level patterns and similarities, which learners engage with via the process of integration. Previous studies investigating statistical learning with bilinguals have typically focused on sequence learning, which is proposed to be an extraction process (Erickson & Thiessen, 2015). As noted, with the exception of segmentation studies employing stimuli with multiple sets of cues that require inhibition (e.g., Wang & Saffran, 2014; Bartolotti et al., 2011), the results largely show no differences between monolinguals and bilinguals in learning (e.g., Yim & Rudoy, 2011; Bogulski, 2013; Bulgarelli & Weiss, 2016). Experiment 3 of this dissertation provides further evidence of this, as we noted no significant effects on the segmentation of multiple inputs as a function of bilingualism. Experiments 1 and 2 of the present work, however, were cross-situational learning tasks, in which learning entails the accumulation of information across trials, which is best characterized as an integration process. Given that we found several learning advantages for bilinguals across Experiments 1 and 2, and no differences in Experiment 3, one possibility is that the bilingual advantage may be more likely to emerge in processes involving integration, but not those involving extraction (see Poepsel & Weiss, 2016; Chapter 2 of this dissertation).

There are a number of possible explanations for how bilingualism causes the differences we see in the learning of single and multiple mappings in CSSL tasks. As noted, one possibility is linked to the type of input that bilinguals frequently receive. That is, bilinguals may have more frequent exposure to learning environments that contain multiple distributions, and so may be less likely than monolinguals to make restrictive assumptions about how many causal models give rise to the surface statistics. This idea finds support from a number of recent results. For instance, as discussed in the Introduction, two fast-mapping studies (in which learners encounter a single paired word and object on each trial) suggest that bilinguals are less bound by the constraint of mutual exclusivity than monolinguals, and further, that greater experience learning languages (i.e., trilingualism vs bilingualism) correlates with even less reliance on mutual exclusivity (Byers-Heinlein & Werker, 2009; Houston-Price et al., 2010). Although mutual exclusivity is typically discussed as a hard-wired learning constraint, these results, along with our own, imply that its manifestation is more closely tied to a learner's specific environment, so that learners who less frequently encounter mutually exclusive relationships, or environments that contain single inputs, are more likely to assume that variability in the statistics of an environment may be indicative of the presence of multiple inputs (e.g., McMurray et al., 2012).

This fits well with a number of other results suggesting that learners who are expecting a change in a learning environment are more capable of accommodating that change. Gebhart et al. (2009), for instance, explicitly informed learners that they would encounter two languages during familiarization in one of their conditions (in addition to a 30s pause between languages) – here, learners were able to acquire both inputs above chance (compared to a baseline condition in which, without an explicit cue or pause, learners showed a primacy effect). Gallistel et al. (2001), familiarized learners (i.e., rats) within either a frequently changing environment (with respect to reward contingencies) or a stable environment, finding that learners in the frequently changing environment more quickly adapted to changes compared to learners in the stable environment. A similar result was found in Zinszer & Weiss (2013), who exposed learners to multiple languages that contained multiple switches (as opposed to a single switch in the baseline condition of Gebhart et al., 2009), and found improved learning of the second language as a result. The suggestion here is that the increased frequency of switching between inputs may have functioned to decrease learners' threshold (e.g., of variability in expected patterns – see Qian et al., 2012) for positing the existence of multiple inputs, and thereby facilitated learning. The extent to which increased switching between inputs facilitates learning is likely to be modulated by a number of factors, among them the similarity of the basic units (i.e., syllables and words in a speech segmentation task), such that more similar inputs may be more difficult to learn than dissimilar inputs (e.g., Gebhart et al., 2009; Franco et al., 2011), even with increased switching frequency. This idea of an interaction between frequency of switching and similarity has yet to be explored in an SL paradigm, but is a logical extension of both Zinszer and Weiss (2013) and Experiment 3 of this dissertation, with consequences for understanding the extent to which variability in these properties functions as a cue to structure.

Together, these findings support the claim above that bilingualism may influence processes that require integrating information arising from multiple inputs across time or learning environments by increasing a learner's expectation for non-uniformity across inputs. That is, learners who are more likely to assume that variability across time correlates with structural changes (e.g., as a result of bilingual experience) may be better able to accommodate that variability, compared to learners with a greater expectation of structural stability. Accordingly, when learning requires attention to local and stable statistical relationships, or integration in the absence of multiple inputs, monolinguals' and bilinguals' performance may be similar. However, when learning requires attention to variable relationships that span time or multiple learning environments, bilingual learners may have an advantage, more quickly attributing such variability to the presence of additional structure rather than assuming it arises as local variability within a single input. Future SL experiments may more explicitly pursue this link between prior expectation and acquisition of multiple inputs in a number of ways. For instance, in the Gebhart et al. (2009) paradigm, the explicit cue (i.e., telling learners that there were multiple inputs) and the pause were presented together, such that the true contribution of either to the finding of improved learning is unclear. Pulling this apart in a condition where learners receive only explicit instruction (and no pause) may more clearly associate prior expectation with improved learning. Learners rarely encounter such explicit instruction of upcoming changes, however (Qian et al., 2012). Implicitly increasing a learner's prior expectation for changes in structure before exposure to a CSSL (or speech segmentation task) with multiple inputs may provide a more rigorous test of the link between expectation and performance. For instance, exposing learners to a non-linguistic sequence learning task containing a single pattern may increase learners' expectation of stability in an input, and so bias performance in accordance with mutual exclusivity in a subsequent CSSL task with multiple inputs. Exposing learners to such a task that contains multiple patterns prior to completing a CSSL task with multiple inputs may facilitate learning of non-mutually exclusive mappings.

An alternative hypothesis to the one posed above is that bilinguals are simply better statistical learners than monolinguals, and so, in more complex statistical learning tasks (such as those containing multiple inputs) outperform monolinguals. There are several reasons to doubt this alternative, however. Primarily, if this were the case, we might have expected to find a bilingual advantage for both 1:1 and 2:1 mappings in Experiments 1 and 2, as well as differences in Experiment 3 where multiple inputs were presented sequentially. In Experiments 1 and 2, we found no advantage for the learning of 1:1 mappings, and furthermore, performance on 1:1 mappings in Experiments 1 (after the first familiarization) and 2 did not approach a high ceiling, suggesting that the lack of a difference was not due to a lack of power. In Experiment 3, the performance of monolinguals and bilinguals was not significantly different. Other comparisons of monolinguals and bilinguals in SL tasks (typically speech segmentation, but also visual SL) have offered mixed effects: in some cases, there is no evidence of a bilingual advantage in the basic task of extracting units (e.g., triplets of syllables in speech segmentation, or triplets of shapes in visual SL) from an input (e.g., Yim and Rudoy, 2013; for explicit learning conditions in Nation & McLaughlin, 1986), while in others, bilinguals show advantages (Bartolotti et al., 2011; Wang & Saffran, 2014; for multilinguals as compared to both monolinguals and bilinguals in implicit learning conditions in Nation & McLaughlin, 1986). As noted above, the specific stimuli used in Bartolotti et al. (2011) and Wang and Saffran (2014), which may have engaged inhibitory or other executive function processes, might have favored bilinguals whose experience tends to confer advantages to executive function (e.g., Bialystok, 2004, 2008; Prior & MacWhinney, 2010; Abutalebi et al., 2012) and so these results may not actually bear on the question of differences in the cores statistical learning abilities. With the exception of these studies, most results, including those presented for the first time here, seem to converge on the idea that monolinguals and bilinguals do not differ in the acquisition of single distributions of statistics, suggesting that the bilingual advantage is not caused by general enhancement of statistical learning abilities, but rather enhancements to specific sensitivities within SL.

Notably, in our experiments, we compared groups of monolinguals to late-learning bilinguals. In contrast, a majority of the extant research comparing monolinguals and bilinguals has focused on early, highly proficient bilinguals. Given that we found a consistent pattern in Experiments 1 and 2, in which bilinguals outperformed monolinguals in the acquisition of multiple inputs, our research suggests that the benefits of bilingual experience are not limited to early bilinguals (although, as noted, they may extend unequally across different types of SL tasks). Further, inasmuch as our studies compare groups of learners whose early experiences with language were similar (i.e., groups that were both functionally monolingual until adolescence), our results provide a more rigorous demonstration of the link between bilingual experience and differences in performance on cognitive tasks. Specifically, there is evidence in support of the assertion that early bilinguals process and develop in different ways compared to their monolingual counterparts, with much attention given to the processing of native and non-native phonemic contrasts (e.g., Pettito et al., 2012; Byers-Heinlein & Fennell, 2014). Additionally, a number of studies demonstrate that differences in brain structure and function in early bilinguals are not always equivalent to those seen in late bilinguals, so that in general, there is a negative relationship between age of acquisition/proficiency in a second language and the extent of changes in brain structure and function (Mechelli et al., 2004; Klein et al., 2013; Mohades et al., 2012; Kim, Relkin, Lee, & Hirsch, 1997; see Li, Legault, & Litcofksy, 2014 for review). One possible consequence of this differential development is that early bilinguals, in comparison to both late learners and monolinguals, may rely on partially or wholly different processing strategies and brain-networks while engaged in language-related tasks. This may complicate the comparison of early bilinguals and monolinguals.

Our significant findings of differences between late bilinguals and monolinguals in Experiments 1 and 2 suggest a close link between the process of becoming bilingual and the emergence of differences in processing. One possibility, broadly consistent with the arguments of the desirable difficulty literature, is that becoming bilingual later in life imposes considerable difficulty on a learner. Examples of these difficulties are an increased need to inhibit the dominant L1 and manage interference between languages through conflict resolution, which are hypothesized to tune the brain networks related to executive function and cognitive control (e.g., Abutalebi et al., 2012; Gold, Kim, Johnson, Kriscio, & Smith, 2013; see Bjork & Kroll, 2015). Early bilinguals frequently show advantages across a range of cognitive tasks drawing on executive function (such as inhibition, task switching, ambiguity resolution and conflict monitoring; e.g., Bialystok, 2004, 2008; Prior & MacWhinney, 2010; Costa et al., 2009), and several recent studies have found that these advantages extend to late bilinguals, as well (e.g., Pelham & Abrams, 2014; Tao et al., 2011). As cross-situational learning arguable draws on components of executive function (see Yu & Smith, 2007; McMurray et al., 2012; Vlach & Sandhofer, 2014), the success of late bilinguals in our tasks provides further evidence that late learning bilinguals may see benefits to executive function from their experience in ways similar to early bilinguals. This conclusion adds to a growing literature addressing the effects of age of acquisition on the emergence and degree of the bilingual advantage, and corroborates several findings indicating that bilingual advantages may be more a function of a learner's habitual use and proficiency in an L2 (e.g., Pelham & Abrams, 2014), rather than age at the onset of learning (despite evidence that development may differ as a result of AoA). Thus, the age-related benefits of bilingualism, such as improved memory and executive function, as well as delayed onset of Alzheimer's symptoms which have typically been associated with early or highly proficient bilinguals (e.g., Bialystok et al., 2012; Craik, Bialystok, & Friedman, 2010; Schroeder & Marian, 2012) may also be conferred on late-learning or lower proficiency bilinguals. Additionally, a number of recent neuroimaging studies show that even limited exposure to and instruction in a second language can have effects on brain structure and function (e.g., after 3 months of training in Mårtensson et al., 2012; after 14 hours of training in McLaughlin, Osterhout, & Kim, 2004) and increase neural plasticity, again suggesting that late bilingual experience may confer differences in processing that are at least related to those seen in early bilinguals (Luk, De Sa, & Bialystok, 2011; Bialystok et al., 2008; Tao et al., 2011; Steinhauer, White, & Drury, 2009; Foucart & Frenck-Mestre, 2012; Pelham & Abrams, 2014). The results of our studies thus contribute to a growing literature on the differences that bilingualism engenders in a learner, and suggest that late-learning bilinguals with mostly classroom exposure are nonetheless affected in measurable ways by their experience.

As noted in the Introduction, one limitation of SL studies to date has been a lack of variability in the input to learners, especially variability of the sort that learners face in real-world language input. A majority of studies thus present learners with a single input distribution that is invariant in the relationships it presents. For example, in CSSL tasks, learners frequently encounter inputs that contain only 1:1 mappings, where all mappings are presented an equal number of times, while in speech segmentation tasks, learners frequently encounter only a single language. Although the insights from studies of simplified and homogenous input have contributed significantly to our understanding of the kinds of statistical sensitivities that learners have, it is only recently that SL experiments have begun to test whether the theoretical models of

SL scale up to real-world language input (e.g., Pelucchi, Hay, & Saffran, 2003; Kurumada, et al., 2013; Frank et al., 2010; Frank, Tenenbaum, & Gibson, 2013; Meylan, Kurumada, Borschinger, Johnson, & Frank, 2012). Many studies have begun to examine how learners deal with multiple mappings, as well as variation in properties such as word length, speaker voice, or duration of familiarization, with findings across the board suggesting that variability may have profound effects on learning, and often a facilitatory effect in the sense that it functions as a cue to the location of structure or changes in an input (e.g., Frank et al., 2013; Gebhart et al., 2009; Weiss et al., 2009; Poepsel et al., 2012; Zinszer & Weiss, 2013; Bulgarelli & Weiss, 2016). Thus, an additional contribution of the present work is in refining our understanding of the kinds of variability that statistical learners can accommodate, as well as how different types of variability influence the outcome of statistical learning tasks. Experiments 1 and 2 used the CSSL paradigm, in which experiments examining the acquisition of 2:1 mappings have been relatively rare. One contribution of these studies, then, is further evidence that learners (both monolinguals and bilinguals) are able to violate mutual exclusivity when necessary, suggesting that, for adult learners, this constraint and the strength of its application may be closely tied to the statistics of a given learning environment (rather than an immutable constraint governing word learning). This is similar to the idea expressed in Markman and Wachtel (1988) that mutual exclusivity likely extends into adulthood (e.g., Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Halberda, 2006) but weakens as learners experience more variability and overlap in terms of mappings. Our results, especially in Experiment 2, further demonstrate that learners maintain the ability to closely track word-object co-occurrence statistics and maintain several mapping hypotheses even when the learning environment contains high levels of mapping ambiguity. At least two previous studies theorize that mapping ambiguity and the ability to engage in a fully associative learning strategy are correlated (e.g., Vouloumanos, 2008; Yurovsky & Frank, 2015). Experiment 2 contributes additional evidence of this relationship, and further suggests that it may be modulated by bilingual experience (as bilinguals showed sensitivity to weaker mappings than monolinguals).

Although such disparities in the frequency with which certain mappings appear might be expected to hinder learning, at least one finding from the speech segmentation literature seems to support the opposite idea. Kurumada et al. (2013), for instance, exposed learners to an artificial language in which the words appeared according to a Zipfian frequency distribution. That is, only a few words appeared with high frequency, while the majority of other words appeared with low frequency. The authors found that, in contrast to a flat distribution of words, the Zipfian distribution actually facilitated learning, as quickly learning the high frequency words allowed learners to locate additional structure (i.e., certainty about where high frequency words began and ended highlighted the beginnings and ends of other, lower frequency words). A convergent finding from the CSSL paradigm is offered in Yurovsky, Fricker, Yu, & Smith (2013), in which the authors report that partial knowledge of word-object mappings facilitates learning of other mappings. In their paradigm, learners were exposed to two cross-situational learning blocks. Half of the words in the second block were words that participants had failed to learn in the first block, and re-exposure to these words facilitated learning of entirely novel word-object mappings (and the previously unlearned words themselves). The authors hypothesize that partial knowledge of some mappings reduces the complexity of the learning problem for unpaired words and objects by reducing the number of possible pairings. Taken together, one possibility is that the non-uniform distribution of evidence for word-object mappings in Experiments 1 and 2 may have been beneficial for some learners, an idea that finds some support in a significant positive correlation between performance on 1:1 mappings and performance on 2:1 mappings in both experiments (although without an investigation of the time course of learning, it may also be the case that some learners were simply better at the CSSL task than others). Speculatively, the results of Experiments 1 and 2 extend the finding of Kurumada et al. (2013) to a new SL task (i.e., from speech segmentation to CSSL), suggesting that variability may be an important component of the statistical learning process, and further that this feature of SL (the ability to leverage knowledge of stronger mappings to acquire unknown or less frequently encountered mappings) may be supported in part by modality or task-general cognitive processes.

In Experiment 3, we manipulated syllable overlap in a speech segmentation task, and present the first systematic investigation of how variability in the degree of incongruent overlap between inputs has significant effects on learning (with the general trend being one of decreased learning as statistical interference increases). While we did find evidence of primacy effects at several degrees of incongruence in accordance with previous studies (e.g., Gebhart et al., 2009), we also found several instances in which both languages were learned above chance, and even a recency effect, which is seldom reported in speech segmentation studies. Although predicting precisely how a certain degree of incongruent overlap will influence learning is not obvious from the results of this study, our findings more generally suggest a complex relationship between the structure of an input and how learning of the system proceeds, so that the expectation of primacy in all cases of sequentially presented inputs may not be warranted. This echoes the findings of Experiments 1 and 2, in which we also showed that a primacy effect in statistical word learning (mutual exclusivity) may be modulated by structural or statistical properties of the input, such as the amount of variability in co-occurrence of words and objects present in the environment. This contrasts with contextual cues, which may be another important source of information about how uniform the statistics of a given environment are (e.g., Gebhart et al., 2009; Weiss et al., 2009; Poepsel et al., 2012), but which learners may not always have access to (i.e., structural changes are not always marked explicitly – Qian et al., 2011) or may not be able to utilize correctly in the absence of specific experience with a cue or set of cues (e.g., Houston & Jusczyk, 2000). Importantly, this suggests that statistical learning may scale quite well to naturalistic language input. That is, although context is useful, learners' sensitivity to the statistics of an input alone may be enough in many cases to enable the extraction of structure and the formation of generalizations about that structure.

All the same, it is clear that our findings do not extend equally across both CSSL and speech segmentation tasks, specifically with respect to the performance of bilinguals. As noted, one suggestion is that these tasks rely to different degrees on sensitivities to different types of statistics (i.e., conditional vs distributional statistics). For example, CSSL may require the tracking and integration of information across time more so than speech segmentation, which may rely to a greater extent on sensitivity to local conditional statistics (Thiessen & Erickson, 2013; Erickson & Thiessen, 2015). We hypothesize that bilingualism may facilitate integration in tasks containing multiple inputs, and accordingly find a bilingual advantage in Experiments 1 and 2 (which used CSSL tasks) but not Experiment 3. One means of increasing the parity of these tasks is adding an integrative component to speech segmentation. In the context of speech segmentation, integration entails developing sensitivity to patterns at a level above the basic statistics of co-occurrence. One example of this is the shift in infant learners from segmenting via statistics (i.e., universal cues to structure) to segmenting via prosodic cues (i.e., language specific cues to structure such as consistent stress patterns across words - see Thiessen & Erickson, 2013), once these more salient cues have been revealed via integration across many exemplars.

The input in speech segmentation is commonly impoverished in comparison to natural language, containing very few words (e.g., four in each language in Experiment 3, as well as Weiss et al., 2009), and few cues beyond transitional probabilities. The addition of a second layer of patterns beyond TPs (such as stress), but congruent with them, to the input in speech segmentation tasks with multiple inputs (i.e., adding the possibility of an integrative task) may thus increase the sensitivity of future experiments comparing monolinguals and bilinguals.

5.3 Conclusions

To date, research on language acquisition in the statistical learning paradigm has largely focused on the performance of monolingual learners in environments that contain a single and invariant input distribution. In this thesis, we have presented evidence that statistical learning is able to accommodate variability arising from the interaction of multiple inputs in SL tasks (and which scales to variability found in natural languages), and further that bilingualism may facilitate the acquisition of multiple inputs in tasks that require the integration of information across time and learning environments, even in late-learning bilinguals. This research contributes to a growing body of work attempting to scale the input of SL tasks to that encountered in real-world learning environments, as well as work suggesting that experience learning multiple languages has significant effects on cognition across many types of tasks.

6 References

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R. & Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*(2), 207-245.

Ard, J. & Homburg, T. (1992). Verification of language transfer. In *S. Gass & L. Selinker (eds.), Language Transfer in Language Learning* (pp. 47–70). Amsterdam: Benjamins.

Aslin, R. N. (2014). Phonetic Category Learning and Its Influence on Speech Production. *Ecological Psychology*, *26*(1-2), 4-15.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, *32*(8), 1369-1378.

Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, *5*(11), 1210-1216.

Barac, R., & Bialystok, E. (2011). Cognitive development of bilingual children. *Language Teaching*, *44*(01), 36-54.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Barrett, M. D. (1978). Lexical development and overextension in child language. *Journal of Child Language*, *5*(02), 205-219.

Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology, 2*.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development*, 498-510.

Bialystok E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, 70, 636–644.

Bialystok, E. (2005). Consequences of bilingualism for cognitive development. *Handbook of Bilingualism*, 417-432.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(01), 3-11.

Bialystok E., Craik F. I. M., Klein R., Viswanathan M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychological Aging*, 19, 290–303.

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, memory, and Cognition*, *34*(4), 859.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(04), 525-531.

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, *24*(01), 27-44.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325-339.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences*, *7*(02), 173-188.

Bischoff-Grethe, A., Proper, S. M., Mao, H., Daniels, K. A., & Berns, G. S. (2000). Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *The Journal of Neuroscience*, *20*(5), 1975-1981.

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology*, *128*(2), 241-252.

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633-660.

Bogulski, C. (2013). Are bilinguals better learners? A neurocognitive investigation of the bilingual advantage. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1511655851).

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, *109*(4), 383-396.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1), 93-125.

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, *58*(1), 49-67.

Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners. *Journal of Child Language*, *22*(02), 307-324.

Bulgarelli, F., & Weiss, D. J. (2016). Anchors Aweigh: The Impact of Overlearning on Entrenchment Effects in Statistical Learning.

Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: insights from bilingual infants. *Developmental Psychobiology*, *56*(2), 274-291.

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, *12*(5), 815-823.

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407-416.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*(2), 282-298.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 484.

Chang, C. H. (1993). Corpus-based adaptation mechanisms for Chinese homophone disambiguation. In *Proceedings of the Workshop on Very Large Corpora, Ohio State University, USA*.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*(2-3), 221-268.

Cenoz, J., & Valencia, J. F. (1994). Additive trilingualism: Evidence from the Basque country. *Applied Psycholinguistics*, *15*(02), 195-207.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354.

Chomsky, N. (1955). Logical syntax and semantics: Their linguistic relevance. *Language*, 36-45.

Chomsky, N. (1965). Aspects of the theory of syntax Cambridge. *Multilingual Matters: MIT Press*.'

Clark, E. V. (1983). Convention and contrast in acquiring the lexicon. In *Concept Development and the Development of Word Meaning* (pp. 67-89). Springer Berlin Heidelberg.

Clark, E.V. (1987). The Principle of Contrast: A constraint on language acquisition. In *B. MacWinney (Ed.), Mechanisms of Language Acquisition* (pp. 1–33). Hillsdale, NJ: Erlbaum.

Cohen, N. J., Eichenbaum, H., & Memory, A. (1993). The Hippocampal System.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1980). Segmenting speech into words. *The Journal of the Acoustical Society of America*, *67*(4), 1323-1332.

Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., & Hommel, B. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 302.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905-912.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491-511.

Costa A., Hernández M., Sebastián-Gallés N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition*, 106, 59–86.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience*, *15*(5), 336-345.

Craik, F. I., Bialystok, E., & Freedman, M. (2010). Delaying the onset of Alzheimer disease Bilingualism as a form of cognitive reserve. *Neurology*, *75*(19), 1726-1729.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2011). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633-664.

Davidson, D., & Tell, D. (2005). Monolingual and bilingual children’s use of mutual exclusivity in the naming of whole objects. *Journal of Experimental Child Psychology*, *92*(1), 25-45.

De Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in a bilingual and a monolingual task: Support for nonselective access to bilingual memory. *Quarterly Journal of Experimental Psychology*, *53*, 397-428.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. *Memory*, *10*, 317-344.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: a pragmatic account. *Developmental Psychology*, *37*(5), 630.

Dijkstra, T. (2003). Lexical processing in bilinguals and multilinguals: The word selection problem. In *The Multilingual Lexicon* (pp. 11-26). Springer Netherlands.

Dufort, R. H., Guttman, N., & Kimble, G. A. (1954). One-trial discrimination reversal in the white rat. *Journal of Comparative and Physiological Psychology*, *47*(3), 248.

Ellis, N. C., & O'Donnell, M. B. (2012). Statistical construction learning: Does a Zipfian problem space ensure robust language learning. *Statistical learning and language acquisition*, 265-304.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *The Quarterly Journal of Experimental Psychology*, *64*(5), 1021-1040.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19-23.

Eramudugolla, R., Irvine, D. R., McAnally, K. I., Martin, R. L., & Mattingley, J. B. (2005). Directed attention eliminates 'change deafness' in complex auditory scenes. *Current Biology*, *15*(12), 1108-1113.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66-108.

Eskes, G. A., Klein, R. M., Dove, M. B., Coolican, J., & Shore, D. I. (2007). Comparing temporal order judgments and choice reaction time tasks as indices of exogenous spatial cuing. *Journal of Neuroscience Methods*, *166*(2), 259-265.

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*(3), 254-260.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104*(2), 301.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 6, 1017–1063.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS One*, *8*(10), e77661.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499-504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, 35, 367–380.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226-248.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: how conscious? *Frontiers in Psychology, 2*, 229.

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107-125.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578-585.

Frank, M. C., Tenenbaum, J. B., & Gibson, E. (2013). Learning and long-term retention of large-scale artificial languages. *PloS One*, *8*(1), e52500.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, *3*(4), 128-135.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117-125.

Galambos, S. J., & Goldin-Meadow, S. (1990). The effects of learning two languages on levels of metalinguistic awareness. *Cognition*, *34*(1), 1-56.

Gallistel, C. R., Mark, T. A., King, A. P., & Latham, P. E. (2001). The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(4), 354.

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science, 33*(6), 1087-1116.

Gerken, L. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, *115*(2), 362-366.

Gilovich, T. D. (1991). The 'hot hand' and other illusions of everyday life. *The Wilson Quarterly*, 52-59.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295-314.

Gleitman, L. R., & Liberman, M. (1995). An invitation to cognitive sciences, Vol. 1: Language.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95-112.

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience*, *33*(2), 387-396.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, *9*(5), 234-241.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, 28(1), 99-108.

Gollan, T. H., & Kroll, J. F. (2001). Bilingual lexical access.

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

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178-186.

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, *20*(02), 229-252.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades.

Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.

Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.

Grosjean, F., & Li, P. (2012). *The psycholinguistics of bilingualism*. John Wiley & Sons.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*(4), 310-344.

Harris, Z. S. (1951). Methods in structural linguistics.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53-B64.

Hayes, J. R., & Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. *Cognition and the Development of Language*, 221-234.

Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastian-Galles, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition*, *13*(03), 315-325.

Hernandez, A. E., Hofmann, J., & Kotz, S. A. (2007). Age of acquisition modulates neural activity for both regular and irregular syntactic functions. *NeuroImage*, *36*(3), 912-923.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, *18*(4), 625-658.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570.

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, *15*(2), 125-150.

Hsu, H. J., & Bishop, D. V. M (2010). Grammatical difficulties in children with specific language impairment: Is learning deficient. *Human Development*, *53*, 264-277.

Hull, R., & Vaid, J. (2007). Bilingual language lateralization: A meta-analytic tale of two hemispheres. *Neuropsychologia*, *45*(9), 1987-2008.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658.

Hwang, J., & Romanski, L. M. (2015). Prefrontal neuronal responses during audiovisual mnemonic processing. *The Journal of Neuroscience*, *35*(3), 960-971.

Ichinco, D., Frank, M. C., & Saxe, R. (2009). Cross-situational word learning respects mutual exclusivity. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (Vol. 31).

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, *18*(3), 879-919.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, *44*(1), 2-31.

Jarvis, S., & Pavlenko, A. (2008). *Cross-linguistic Influence in Language and Cognition*. Routledge.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339-345.

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears constraints on bilingual lexical activation. *Psychological Science*, *15*(5), 314-318.

Jungé, J. A., Scholl, B. J., & Chun, M. M. (2007). How is spatial context learning integrated over signal versus noise? A primacy effect in contextual cueing. *Visual Cognition*, *15*(1), 1-11.

Jusczyk, P. W. (1997). Finding and remembering words: Some beginnings by English-learning infants. *Current Directions in Psychological Science*, 170-174.

Kachergis, G., Yu, C., & Shiffrin, R. (2009). Frequency and contextual diversity effects in cross-situational word learning. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (Vol. 31, pp. 2220–2225). Austin, TX: Cognitive Science Society.

Kahneman, D., & Tversky, A. (1974). Subjective probability: A judgment of representativeness. In *The Concept of Probability in Psychological Experiments* (pp. 25-48). Springer Netherlands.

Kalia, V., Wilbourn, M. P., & Ghio, K. (2014). Better early or late? Examining the influence of age of exposure and language proficiency on executive function in early and late bilinguals. *Journal of Cognitive Psychology*, *26*(7), 699-713.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*(3), 263-269.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966-968.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, *16*(4), 705-710.

Keshavarz, M. H., & Astaneh, H. (2004). The impact of bilinguality on the learning of English vocabulary as a foreign language (L3). *International Journal of Bilingual Education and Bilingualism*, *7*(4), 295-302.

Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and Motor Skills*.

Kim, K. H., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*(6638), 171-174.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Klein, E. C. (1995). Second versus third language acquisition: Is there a difference? *Language Learning*, *45*(3), 419-466.

Klein, D., Mok, K., Chen, J. K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: a cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, *131*, 20-24.

Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*(5940), 611-612.

Krechevsky, I. (1932). "Hypotheses" versus "chance" in the pre-solution period in sensory discrimination-learning. *University of California Publications in Psychology*.

Kroll, J. F. (2008). Juggling two languages in one mind. *Psychological Science Agenda, American Psychological Association*, *22*(1).

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497-514.

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two Languages in Mind Bilingualism as a Tool to Investigate Language, Cognition, and the Brain. *Current Directions in Psychological Science*, *23*(3), 159-163.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*(02), 119-135.

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Valdes Kroff, J. R. (2012). 7 Juggling Two Languages in One Mind: What Bilinguals Tell Us About Language Processing and its Consequences for Cognition. *Psychology of Learning and Motivation-Advances in Research and Theory*, *56*, 229.

Kuo, M. L. A., & Hooper, S. (2004). The effects of visual and verbal coding mnemonics on learning Chinese characters in computer-based instruction. *Educational Technology Research and Development*, *52*(3), 23-34.

Kurumada, C., Meylan, S. C., & Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition*, *127*(3), 439-453.

Lenneberg, E. H., Chomsky, N., & Marx, O. (1967). *Biological foundations of language* (Vol. 68). New York: Wiley.

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex*, *58*, 301-324.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507-1515.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474-496.

Luk, G., De Sa, E. R. I. C., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*(04), 588-595.

Majerus S., Poncelet M., van der Linden M., Weekes B. S. (2008). Lexical learning in bilingual adults: the relative importance of short-term memory for serial order and phonological knowledge. *Cognition*, 107, 395–419

Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. *Perspectives on Language and Thought: Interrelations in Development*, 72-106.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121-157.

Marian, V., & Shook, A. (2012, September). The cognitive benefits of being bilingual. In *Cerebrum: The Dana Forum on Brain Science* (Vol. 2012). Dana Foundation.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition*, *6*(02), 97-115.

Marsh, J. K., & Ahn, W. K. (2006). Order effects in contingency learning: The role of task complexity. *Memory & Cognition*, *34*(3), 568-576.

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240-244.

Mayberry, R. I., & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language*, *87*(3), 369-384.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122-134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

McMurray, B., Dennhardt, J. L., & Struck-Marcell, A. (2008). Context Effects on Musical Chord Categorization: Different Forms of Top-Down Feedback in Speech and Music. *Cognitive Science, 32*(5), 893-920.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119, 4, 831-877.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *The Journal of Neuroscience*, *26*(29), 7629-7639.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Neurolinguistics: structural plasticity in the bilingual brain. *Nature*, *431*(7010), 757-757.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*(22), 9014-9019.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596-606.

Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, i-129.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*(2), 201-213.

Meylan, S., Kurumada, C., Börschinger, B., Johnson, M., & Frank, M. C. (2012). Modeling online word segmentation performance in structured artificial languages.

Misyak, J. B., & Christiansen, M. H. (2007). Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language. In *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1307-1312). Austin, TX: Cognitive Science Society.

Mitchel, A. D., Christiansen, M. H., & Weiss, D. J. (2014). Multimodal integration in statistical learning: evidence from the McGurk illusion. *Frontiers in Psychology*, *5*.

Mitchel, A. D., & Weiss, D. J. (2010). What's in a face? Visual contributions to speech segmentation. *Language and Cognitive Processes*, *25*(4), 456-482.

Mitchel, A. D., & Weiss, D. J. (2011). Learning across senses: cross-modal effects in multisensory statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1081.

Mohades, S. G., Struys, E., Van Schuerbeek, P., Mondt, K., Van De Craen, P., & Luypaert, R. (2012). DTI reveals structural differences in white matter tracts between bilingual and monolingual children. *Brain Research*, *1435*, 72-80.

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology*, *25*(5), 531-546.

Myers-Scotton, C. (2008). Language contact: Why outsider system morphemes resist transfer. *Journal of Language Contact*, *2*(1), 21-41.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience*, *30*(37), 12366-12378.

Nation, R., & McLaughlin, B. (1986). Novices and experts: An information processing approach to the "good language learner" problem. *Applied Psycholinguistics*, *7*(01), 41-55.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127-162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85-117.

Newport, E., & Supalla, T. (1987). A critical period effect in the acquisition of a primary language. *Science*.

Odlin, T. (1999). An army, a navy, and Ebonics. *Language Contact, Variation, and Change*, 195-233.

Olson, C. L. (1976). Some apparent violations of the representativeness heuristic in human judgment. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 599.

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development*, *76*(2), 324-339.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology*, *48*(1), 98-107.

Pashler, H. (1989). Dissociations and dependencies between speed and accuracy: Evidence for a two-component theory of divided attention in simple tasks. *Cognitive Psychology*, *21*(4), 469-514.

Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 313.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*(3), 674-685.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 101.

Perfors, A., Tenenbaum, J., & Regier, T. (2006). Poverty of the stimulus? A rational approach. In *Annual Conference*.

Perruchet, P., Poulin-Charronnat, B., Tillmann, B., & Peereman, R. (2014). New evidence for chunk-based models in word segmentation. *Acta Psychologica*, *149*, 1-8.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246-263.

Petitto, L. A., Berens, M. S., Kovelman, I., Dubins, M. H., Jasinska, K., & Shalinsky, M. (2012). The “Perceptual Wedge Hypothesis” as the basis for bilingual babies’ phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and Language*, *121*(2), 130-143.

Piaget, J., & Play, D. (1962). Imitation in Childhood. *NY: Norton*.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, *92*, 377-410.

Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. *The Reality of Linguistic Rules, 321*, 51.

Poepsel, T., Gerfen, C., & Weiss, D.J. (2012). Context, mutual exclusivity, and the challenge of multiple mappings in word learning. In *Proceedings of the 36th Annual Boston Conference on Language Development*, (pp. 474-486).

Poepsel, T. J., & Weiss, D. J. (2014). Context influences conscious appraisal of cross situational statistical learning. *Frontiers in Psychology*, *5*.

Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., & Knowlton, B. J. (2005). The neural correlates of motor skill automaticity. *The Journal of Neuroscience*, *25*(22), 5356-5364.

Polka, L., Sundara, M., & Blue, S. (2002). The role of language experience in word segmentation: A comparison of English, French, and bilingual infants. *The Journal of the Acoustical Society of America*, *111*(5), 2455-2455.

Polka, L., & Sundara, M. (2003, August). Word segmentation in monolingual and bilingual infant learners of English and French. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1021-1024). Barcelona: Causal Productions.

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: native language, cross-dialect, and cross-language comparisons. *Infancy*, *17*(2), 198-232.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(02), 253-262.

Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: more than detecting changes. *Frontiers in Psychology*, *3*.

Quine, W. V. O. (1960). Translation and meaning. *Word and object*, 26-79.

Rawson, K. A., & Kintsch, W. (2005). Rereading Effects Depend on Time of Test. *Journal of Educational Psychology*, *97*(1), 70.

Reber, A. S. (1965). Implicit learning of artificial grammars. Unpublished master's thesis, Brown University

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855-863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, *125*(2), 123.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, *66*(1), 30-54.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. MIT Press.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*(1), 245-277.

Ringbom, H. (2006). The importance of different types of similarity in transfer studies. *Cross-Linguistic Influences in the Second Language Lexicon*, 36-45.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110-114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 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*(1), 27-52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101-105.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606-621.

Salvatierra, J. L., & Rosselli, M. (2010). The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism*.

Schlegel, A. A., Rudelson, J. J., & Peter, U. T. (2012). White matter structure changes as adults learn a second language. *Journal of Cognitive Neuroscience*, *24*(8), 1664-1670.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299-309.

Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology*, *24*(5), 591-601.

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, *22*(1), 106-129.

Scott-Brown, K. C., Baker, M. R., & Orbach, H. S. (2000). Comparison blindness. *Visual Cognition*, *7*(1-3), 253-267.

Scott-Brown, K. C., & Orbach, H. S. (1998). Contrast discrimination, non–uniform patterns and change blindness. *Proceedings of the Royal Society of London B: Biological Sciences*, *265*(1411), 2159-2166.

Service E., Simola M., Metsänheimo O., Maury S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, 14, 383–408. 10.1080/09541440143000140

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261-267.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1), 39-91.

Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General*, *142*(4), 1159.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558-1568.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning Versus Performance An Integrative Review. *Perspectives on Psychological Science*, *10*(2), 176-199.

Soveri, A., Laine, M., Hamalainen, H., & Hugdahl, K. (2011). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism: Language and Cognition*, *14*(03), 371-378.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, *25*(1), 13-41.

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395-411.

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*, *39*(4), 505-513.

Tang, C., & Van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, *119*(5), 709-732.

Tao, L., Marzecová, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: the role of age of acquisition. *Bilingualism and Cognitive Control*, 83.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology*, *39*(4), 706.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1131.

Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, *30*(5), 639.

Toro, J. M., Sebastián-Gallés, N., & Mattys, S. L. (2009). The role of perceptual salience during the segmentation of connected speech. *European Journal of Cognitive Psychology, 21*(5), 786-800.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25-B34.

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, *27*(7-8), 979-1001.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126-156.

Tunney, R. J., & Altmann, G. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1322.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934-1945.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty : heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, *126*(1), 367-376.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399-433.

Umemoto, A., Scolari, M., Vogel, E. K., & Awh, E. (2010). Statistical learning induces discrete shifts in the allocation of working memory resources. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1419.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780-789.

Van Hell, J. G., & Mahn, A. C. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, *47*(3), 507-546.

Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition*, *137*, 40-46.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375-382.

Vlach, H. A., & Sandhofer, C. M. (2014). Retrieval Dynamics and Retention in Cross-Situational Statistical Word Learning. *Cognitive Science*, *38*(4), 757-774.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729-742.

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*.

Weber, A., & Paris, G. (2004). The origin of the linguistic gender effect in spoken-word recognition: Evidence from non-native listening. In *Proceedings of the 26th Annual*

Meeting of the Cognitive Science Society (Vol. 26, pp. 1446-1451). (Mahwah, NJ: Erlbaum).

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development*, *5*(1), 30-49.

Weiss, D. J., Poepsel, T., & Gerfen, C. (2015). Tracking multiple inputs. *Implicit and Explicit Learning of Languages*, *48*, 167.

Wexler, K., & Manzini, M. R. (1987). Parameters and learnability in binding theory. In *Parameter Setting* (pp. 41-76). Springer Netherlands.

Williams, B. A. (1988). Reinforcement, choice, and response strength.

Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition*, *7*(1-3), 297-322.

Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two-and three-year-olds track a single meaning during word learning: Evidence for Propose-but-verify. *Language Learning and Development*, *12*(3), 252-261.

Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*(3), 288-297.

Yim, D., & Rudoy, J. (2013). Implicit statistical learning and language skills in bilingual children. *Journal of Speech, Language, and Hearing Research*, *56*(1), 310-322.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414-420.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165-180.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word–referent learning: Prior questions. *Psychological Review*, *119*(1), 21.

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Development*, *4*(1), 32-62.

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, *145*, 53-62.

Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, *21*(1), 1-22.

Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 715-720).

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive Processes in Cross-Situational Word Learning. *Cognitive Science*, *37*(5), 891-921.

Zelinsky, G. J., & Loschky, L. (1998). Toward a more realistic assessment of visual short-term memory. *Investigative Ophthalmology & Visual Science*, *39*, S224.

Zelinsky, G. J. (1998). Detecting changes between scenes: A similarity-based theory using iconic representations. (Tech. Rep. No. CNS-98-01). Champaign, IL: University of Illinois, Cognitive Neuroscience Group

Zinszer, B. D., & Weiss, D. J. (2013). When to Hold and When to Fold: Detecting Structural Changes in Statistical Learning. *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3858-3863)

VITA
Timothy J. Poepsel

Education

2016 (Expected)	Ph.D in Psychology and Language Science *The Pennsylvania State University*
2013	M.S. in Psychology *The Pennsylvania State University*
2009	B.A. in Linguistics *Northwestern University*

Selected Honors and Awards

2016	Outstanding Student Publication Award
2012	Confucius Scholarship
2010	University Graduate Fellow
2009	Phi Beta Kappa

Selected Publications

Poepsel, T.J., Weiss, D.J. (2016). “The influence of bilingualism on statistical word learning.” *Cognition, 152*: 9-19

Weiss, D. J., Poepsel, T.J, & Gerfen, C. (2015). Tracking multiple inputs. *Implicit and Explicit Learning of Languages*, *48*, 167.

Poepsel, T.J., Weiss, D.J. (2014). "Context influences conscious appraisal of cross-situational statistical learning.” *Frontiers in Psychology*.

Poepsel, T.J., Gerfen, C., Weiss, D.J. (2012). "Context, mutual exclusivity, and the challenge of multiple mappings in word learning.” *Proceedings of the 36th annual Boston Conference on Language Development.* Somerville, MA: Cascadilla Press.

Selected Presentations

Poepsel, T.J, Weiss, D.J. (2016). “The Influence of Bilingualism on Ambiguity Resolution in Cross-Situational Learning”. *Poster to be presented at the annual meeting of The Psychonomic Society*. Boston, MA

Poepsel, T.J, Weiss, D.J. (2014). “The influence of bilingualism on statistical word learning”. *Paper presented at the 122nd annual meeting of the American Psychological Association.* Washington, D.C.

Poepsel, T.J., Gerfen, C., Carlson, M. (2012). “On the perception of gay-sounding male speech by non-native speakers of English”. *Paper presented at the 31st Second Language Research Forum (SLRF*). Carnegie Mellon University, Pittsburgh, PA.

Poepsel, T.J., Gerfen, C., Weiss, D.J. (2011). “Context, mutual exclusivity, and the challenge of multiple mappings in word learning.” *Paper presented at the 36th Boston University Conference on Language Development (BUCLD).* Boston University, Boston, MA.