THE COGNITIVE AND NEURAL UNDERPINNINGS OF LANGUAGE LEARNING AND PROCESSING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ethan Donald Jost

December 2018

© 2018 Ethan Donald Jost

THE COGNITIVE AND NEURAL UNDERPINNINGS OF LANGUAGE LEARNING AND PROCESSING

Ethan Donald Jost, PhD

Cornell University 2018

Language is at the epicenter of our existence, allowing us to communicate and think about complex ideas, feelings, and anything else we want. While language is often regarded as an isolated, unique phenomenon when looking across taxa, it is important to realize that this apparent detachment from the rest of cognition is illusory – our human language abilities are in fact highly dependent upon more basic cognitive processes. Much of this dissertation focuses on the ways in which the process of statistical learning both allows for and constrains language learning and processing. Statistical learning can be thought of as a cognitive process by which learners implicitly form associations between stimuli by tracking and storing the underlying statistical relationships between such elements. To provide insight into the relationship between statistical learning and language, two studies are reported here. I first demonstrate the reliability of paradigms that are frequently used to test this construct, while also showing how individual differences in statistical learning are correlated with biases in language processing. In the second study, I characterize how constraints on the input available to learners can affect their ability to acquire statistically learned grammatical regularities. I also establish that such knowledge is retained over time, by examining performance at a follow-up session two-weeks after training.

The third study puts long-held assumptions about the modularity of the brain's language network to the test by examining neuroplasticity in adult patients with brain tumors. The results of this study show that the right frontal lobe is capable of maintaining language function when there is damage to the left frontal lobe. Together, the findings reported within offer evidence for a language system that is highly sensitive to the distributional properties of the input, and is characterized by processes of entrenchment and plasticity.

BIOGRAPHICAL SKETCH

Ethan Donald Jost was born in St. Louis, Missouri to Lynette and Larry Jost. He has one younger brother, Trevor. Ethan's first academic experiences took place at Christ Memorial Lutheran Preschool, before he moved on to Green Park Lutheran School. After completing his secondary education at Lutheran High School South, he matriculated to Saint Louis University, eventually graduating summa cum laude with an Honors B.A. in Psychology, minoring in History and Philosophy along the way. At Saint Louis University, Ethan had the privilege of working in Dr. Chris Conway's Brain Learning and Cognition Lab for three years. There, he worked on numerous projects, and was able to publish his first study on the neurodevelopment of statistical learning abilities. Ethan then joined the Cognitive Neuroscience of Language Lab at Cornell University, working under Dr. Morten Christiansen beginning in the fall of 2011. He also spent time as a Visiting Student at Memorial Sloan Kettering Cancer Center, where he worked in the Functional MRI Lab with Dr. Andrei Holodny beginning in the fall of 2012.

ACKNOWLEDGMENTS

First, I would like to thank my family. It may sound trite, but I could not have done this without you. To my parents, Larry and Lynette, you have always been there to encourage my academic pursuits and worked hard to give me every opportunity to better myself and learn new things. Knowing that I always have a loving home to come back to has provided me comfort and security in my years away, and I will be forever grateful for that. To my brother Trevor, you are a continued source of inspiration. Your work-ethic and commitment to constantly improving yourself has always been something I admired within you and have sought to emulate myself. You are also pretty enjoyable to spend time with, and I look forward to many more adventures in the years to come. To my grandparents, aunts, uncles, and cousins, thanks for also being constant sources of fun and support. In particular, my two Grandmothers, who were always kind enough to spend time on the phone with me and just chat about life. Those moments (ok, hours) are something I will cherish.

To my friends, you have been equally responsible for my success. When I told people at Cornell about the friends I made in college, they generally thought that I was using hyperbole when describing the strength of the bonds we forged in our brief time together at SLU. However, the ones who were fortunate to meet you quickly realized that it would be impossible to overstate your impact on my life. Also, to the friends I refrain from telling other people about (hey there Rackpocalypse), you are ok, too. To my fellow Cornellians and Ithacans, I thank you for befriending me in spite of the inevitably transient nature of our relationships. I have loved exploring New York with you all. To the friends I have made elsewhere along the way, thanks. Whether we met when we were infants, in high school, or wherever, the number of people who have brought joy and meaning to my life is enumerable. Thank you, Adam (even though I did not make it into your acknowledgments, although this contrast does seem appropriate), for being an excellent roommate and better friend. Also, thank you to the students I had the honor of teaching over the years; time spent in the classroom with you was always refreshing.

To the office staff, wow. Pam, Cindy, Lisa, Linda, Keith, Liz, Julie, Fred, and Mary Lou, I appreciate all of the ways in which you have made my life easier over the last seven years. The number of times each of you has solved a problem for me is difficult to keep track of. I would have floundered without your guidance and help.

To my lab, you are wonderful. Erin, Jen M., Stewart, and Julia, thank you for being my lab siblings over the years. I have enjoyed collaborating, meeting, talking, writing, analyzing, emailing, chunking, etc. with you and literally needed all of the help you gave me to finish this off. Also, to our honorary lab member Rebecca, thanks for being a constant source of encouragement and commiseration via Snapchat. To all of the research assistants who helped me over the years, thank you. Data does not collect itself, at least for the kind of research I was doing, and your hours spent in the lab over the past seven years are the foundation upon which this dissertation was built. Also, to my other lab, the BABY lab, thanks for being a source of fun, baked goods, and insight over the years. Especially, though, thank you Jen S.; your kindness and generosity are matched only by your knowledge of statistics and the quality of your biscotti. Most importantly, you made Ithaca feel more like home.

To my collaborators, thank you for giving me the opportunity to do fun, interesting, and even at times, inspiring research. To Andrei, Kyung, and Nicole, thank you for working with me over the many years and welcoming me into your lab. My time in the hospital spent working with you led to some of the most rewarding experiences of my graduate career. To Kate and

Kara, thanks for letting me become part of your project, and I am continually enthralled by the work we are doing. To Aaron and Dan, I wish I had been able to complete our project, but am glad to have had the chance to work with you and hold out hope that something comes out of the work we have put in. And to Chris, thank you for getting me started in research and for continuing to support me throughout my graduate career. Your initial inspiration and persistent geniality got me interested in cognitive psychology, and I am beyond grateful for your guidance through the years.

To my committee, thank you for being the kinds of mentors that make graduate school fun and challenging in equal measure. Morten, I could not have hoped for a more understanding, helpful, and kind advisor. You gave me the freedom to explore interesting questions, and never ceased to support me in my academic endeavors. Without you, this dissertation would have ended with the preface. Graduate school may not have always been easy, but I do not think it is supposed to be, and will always remember how you aimed to understand rather than judge. Mike, thanks for being both a teacher and friend. Whether we were discussing sensitive periods, cameras, methods, or wood-chopping techniques, our conversations were always stimulating and full of insight. Thank you (and Jen) for letting me into your home, and sharing your cats with me. To Barb, I appreciate the fact that I could always wander into your office and learn something new. Your enthusiasm for understanding the brain helped me land at MSK thanks to the random meeting you had with Andrei many years ago, and without you that collaboration would have never been conceived. To Thom, thanks for being such a jovial dude. I know that I can count on you for a smile and a laugh in one moment, and a great question about my work in the next. I admire your balance, and thank you for adding breadth to my understanding of our field. And last but not least, thank you to the NIH (and Barb and BJ) for grant 5T32HD055177.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SL.. Statistical learning

AGL... Artificial grammar learning

vWM... Verbal working memory

B2.. Brocanto2

GJT………………………………………………………….. Grammaticality judgment task

GLMM……………………………………………….... Generalized linear mixed-effects model

fMRI………………………………………….....……... Functional magnetic resonance imaging

rTMS…………………………………………...... Repetitive transcranial magnetic stimulation

BNT…………………………………………………………………….... Boston naming task

ROI……………………………………………………………………….... Region of interest

LI……………………………………………………………………….…..... Laterality index

IFG………………………………………………………………...…... Inferior frontal gyrus

MFG……………………………………………………………........... Middle frontal gyrus

PREFACE

While we all use language every day, most do not think about their ability to do so except in the rarest of circumstances. Maybe part of the reason that the study of language is not exciting to everyone is that it seems inevitable – we all just do it. However, while this seems like a fact on its face, the truth is perhaps a bit murkier, and certainly more interesting.

As an applicant to the PhD program in Cornell's Psychology Department, I posed a series of questions within my statement of interest, and have excerpted a paragraph that I would like to revisit here:

> *Language is the commonality shared by all people, a nearly universal ability in a diverse world full of unique individuals and cultures. The fact that language is natural for our species makes it an interesting topic to investigate. How do such unique individuals come to an understanding of this abstract form of communication? What rules does language follow so that its users can communicate information accurately and consistently? Why are people, especially children, so adept at language acquisition? What neural functions do we rely on for this process of language learning and production?*

Fortunately for me, I found a group of people here at Cornell to help me both answer these questions and come to a better understanding of what I meant by them in the first place. I believe that you will find some satisfactory answers to each of the questions posed above within the contents of this dissertation, although the questions themselves have become more polished over time. What do individual differences in language processing abilities tell us about the language system itself? What kinds of linguistic information are learners sensitive to? In what ways do constraints on linguistic input affect language learning? How can our brain's language network change over time?

In the chapters that follow, I will first introduce the concept of a language system that relies upon learners' ability to extract distributional information from the input. Then I will explore the validity of common measures used to evaluate these abilities, and also look for individual differences within them. Next, I will focus on ways in which intentionally changing the input can affect how well learners absorb and retain this kind of distributional information. After that, I will examine how plastic the cortical structures underlying the language system are. Finally, I will attempt to pull these findings together in a description of how language is underpinned by a system that relies on distributionally-defined input and is characterized by both entrenchment and plasticity. Hopefully you find the answers posed within compelling, and maybe even think of some new questions of your own along the way.

CHAPTER ONE

General introduction

The studies presented in this dissertation revolve around one critical principle: that language is shaped by the brain (Christiansen & Chater, 2008). Building on existing support for this idea, my graduate research sought to refine our understanding of the cognitive and neural underpinnings that underlie language learning and processing. This work has also lent credence to the idea that learning and processing are inextricably linked (Christiansen & Chater, 2016), which fits nicely alongside theories positing that our communicative system is characterized by entrenchment (Schmid, 2016). In order to do so, the studies featured within utilize a range of statistical learning (SL), artificial grammar learning (AGL), and language processing paradigms. Combined, this set of studies attempts to elucidate the ways in which basic cognitive abilities interface with and support language learning and processing, while also determining the properties of the cortical network that subserves language function.

An Introduction to Statistical Learning

Language is notoriously complex, and language learning is arguably one of the most difficult challenges humans face. Yet, learners overcome this challenge with relative ease - thanks (at least in part) to the myriad cues contained within language itself. One such cue is the distributional nature of language; there is a large body of evidence to suggest that for countless elements of natural language, their structure can be described in terms of statistical or distributional relations (Lashley, 1951; Mandelbrot, 1953; Rubenstein, 1973). Accordingly, abilities relating to detecting and learning relations among linguistic elements have been suggested to play a critical role in language acquisition (e.g. Altmann, 2002; Conway, Bauernschmidt, Huang, & Pisoni, 2010; Conway & Christiansen, 2005; Conway & Pisoni, 2008; Gupta & Dell, 1999; Saffran, 2003). This includes SL; the implicit process of discerning and acquiring distributionally-defined structure through complex computations of item co-occurrence.

SL is the cognitive process that serves as the focus of the first two experimental chapters of this dissertation. The past 20 years have seen a wealth of research on humans' capacity for SL, particularly with relation to language; research has demonstrated that learners of all ages are sensitive to the distributional regularities contained within a stream of linguistic input (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), for both natural and artificial language (Pelucchi, Hay, & Saffran, 2009). Further, SL has been documented for linguistic structures of varying complexity – from detecting and extracting simple trisyllabic sequences denoting word-like units (Saffran, Aslin & Newport, 1996; Aslin, Saffran & Newport, 1998), to learning more complex grammatical structure (Frost & Monaghan, 2016; Gerken, 2006; Gómez & Gerken, 2000; Marcus et al., 1999). Together, these findings demonstrate humans' remarkable capacity for learning linguistic structure with the help of distributional cues. Crucially, given the distributional properties of natural language, they also indicate that SL could feasibly contribute to natural language acquisition.

CHAPTER 1: INTRODUCTION

This work builds on foundations laid in earlier seminal work on learning, perception, and cognition. In a landmark study on implicit learning, Reber (1967) provided early documentation of adults' sensitivity to distributional linguistic structure using an artificial grammar. Equally influential (though often overlooked) is Gibson and Gibson's (1955) theory of perceptual learning, which proposed that repeated experience with a percept enhances one's ability to discriminate between it and other percepts - paving the way for accounts of learning with a basis in sensory experience, as well as current ideas about linguistic entrenchment (e.g. Schmid, 2007). In other early work, Miller and Selfridge (1950) suggested that a reliance on transitional probabilities may be similar to the way in which grammar is learned.

Research informing these critical works demonstrated that frequent co-occurrence of items (due to underlying structure) improved participants' recall of letter sequences (Miller, 1958), and that learning the positional relationships between linguistic units (i.e., morphemes) occurs as an experiential process of familiarization with the temporal positions in which such units are frequently encountered (Braine, 1963). This gave rise to the future research investigating the close relationship between frequent co-occurrence and the strength and automaticity of recall at various levels of linguistic analysis which serves as the foundation for the work described in this dissertation.

From the beginning, research on implicit learning related to language has focused on the way(s) in which units of linguistic information are formed. Many of the early explanations for the ways in which this learning happened relied upon experience-based accounts, as described above. However, experience-independent theories of language acquisition quickly became the dominant perspective primarily due to the widespread acceptance of the "poverty of the stimulus" argument (Chomsky, 1965; Crain, 1991). Saffran, Aslin and Newport's (1996) seminal research gave the psychology of language an experience-dependent process by which at least one aspect of linguistic knowledge (words) could be learned, and demonstrated that this could be accomplished fairly rapidly even at an early stage in development; SL can thus be thought of as the acquisition of distributional information from perceptual input.

Since the exact nature of the distributional information learners are thought to be sensitive to varies across studies, this chapter aims to bring together research from multiple perspectives, in order to give a thorough overview of the field. The kinds of statistics that learners are using in each task and study will be highlighted and contrasted, particularly when such differences are theoretically important. With the uncovering of this ability, and the increased weight given to connectionist ideas about how the items and structure of language can emerge from the input (Elman, 1990), experience-dependent accounts of language learning and processing have again become central to the psychology of language. Building on these ideas, we define SL as the process by which learners uncover the structure of the input from its distributional properties (Frost, Armstrong, Siegelman, & Christiansen, 2015).

Implicit Learning Meets Statistical Learning

Since the resurgence of experience-dependent accounts of language, attempts have been made to synthesize classic implicit learning literature with contemporary research on SL (e.g., Conway & Christiansen, 2006; Perruchet & Pacton, 2006; Christiansen, in press). This endeavor has given rise to questions concerning the “implicitness” of SL, and the related AGL paradigms that are commonly employed by the implicit learning literature. This is particularly relevant to discussions of entrenchment processes, as automaticity – or unconscious activation – is usually considered a feature of entrenchment; the naming of an entrenched visual stimulus (e.g., an apple) does not require conscious processing in healthy adults (Schmid, 2007; see Hartsuiker & Moors, 2016, for more details). However, considering the manner in which most SL paradigms are designed, with explicit familiarity judgments used at test, the relative amount of conscious processing that learners rely upon has been debated.

Within most SL studies, self-report data (i.e., that participants cannot explicitly recall when asked critical features of the training input) and the mere fact that task instructions give no mention of what the participants are expected to learn are used as evidence for implicit processing. Recent work has put this to the test, with evidence both for (Kim, Seitz, Feenstra, & Shams, 2009) and against (Bertels, Franco, & Destrebecqz, 2012) implicit interpretations of SL. Further research has shown that access to the statistical relationships defined within two different artificial languages can be consciously controlled by the participant, demonstrating that at least some aspects of the learned relationships is available for explicit processing (Franco, Cleeremans, & Destrebecqz, 2011). Early AGL research pointed towards diminished performance when participants were given explicit instructions (Reber, 1976), although newer research suggests that the duration of stimulus presentation may modulate this relationship, with longer presentations leading to an improvement in learning when instructions are explicit, at least in the visual domain (Arciuli, Torkildsen, Stevens, & Simpson, 2014). There appears to be a strong argument for the implicit and incidental nature of SL, but some room for explicit processing should be built into accounts of SL. Some of the issues in understanding the implicit nature of SL are due to the lack of coherence between the implicit and SL literatures, but may be resolved in time as the two become more closely integrated.

Perruchet and Pacton (2006) have claimed that while the two literatures have grown increasingly similar in terms of methodology, implicit learning relies more on the process of chunking as an explanation of learning (see Gobet, 2016), while the SL literature is primarily interested in exploring the role of distributional information. However, these computations do not need to be interpreted as dichotomous; depending on the properties of the input they could both occur in what we think of as SL (Franco & Destrebecqz, 2012). Tracking conditional probabilities may lead to the formation of chunks at later stages of learning, which then become elements themselves between which conditional probabilities may be tracked. In fact, recent models of language acquisition have demonstrated the feasibility of such a process (Monaghan & Christiansen, 2010; McCauley & Christiansen, 2014; in press). Thinking of chunks as the outcome of SL provides a direct connection with entrenchment: Throughout learning, frequently co-occurring elements and structures become more deeply entrenched, strengthening such representations.

Statistical Learning as a Domain-General Cognitive Process

Importantly, SL falls into the broader category of domain-general cognitive abilities (Kirkham, Slemmer, & Johnson, 2002. This means that SL is not expressly *for* language, rather it is used both when performing linguistic tasks, and for learning across a range of other cognitive and perceptual domains. If SL was domain-specific and only related to the way in which language is learned and processed, then statistical relationships between non-linguistic elements should not be learnable. This appears not to be the case, as the ability to learn from the transitional probabilities in sequences of auditory tones has been well described in the literature. Saffran and colleagues (1999) first reported the sensitivity of adults and infants to the underlying statistical relationships between tones, using the same type of dependency previously investigated using syllables (Saffran et al., 1996; Saffran, Newport, & Aslin, 1996). The ability of participants to track adjacent dependencies between tones that are inherently non-linguistic indicates that SL is likely a domain-general ability.

Other kinds of acoustic information have also been used in SL studies, with varying results depending on the properties of the acoustic stimuli (Creel, Newport, & Aslin, 2004). Interestingly, certain aspects of the stimulus (e.g., pitch register and timbre) led to different patterns of sensitivity in learning non-adjacency vs. adjacency structure in the stimulus stream, suggesting that Gestalt-like properties of the stimulus may shape learning in different ways. Other reports of SL have relied on artificial grammars using musical stimuli, further demonstrating the domain-general nature of SL (e.g., Bly, Carrion, & Rasch, 2009). This domain-generality indicates that language is subserved by neural mechanisms that are used for processing a variety of input, and/or that the same general computational principle operates across perceptual and cognitive domains.

Auditory input is still somewhat language-like, though, and in isolation these effects are somewhat difficult to disentangle - particularly as linguistic and non-linguistic auditory stimuli requires the same sensory modality. Compelling evidence for the domain generality of SL would benefit from research showing that such learning exists within the visual domain; demonstrating that learners can compute distributional information pertaining to non-linguistic visual sequences would provide a strong indication that SL is neither language nor modality specific. Indeed, evidence of visual SL began with a study examining infant looking times to statistically determined patterns of shapes, finding differences in looking times between familiar and unfamiliar patterns (Fiser & Aslin, 2002; Kirkham et al., 2002; Raviv & Arnon, 2018; Frost, Monaghan, & Tatsumi, 2017). The statistical coherence between elements within these visual scenes led to their entrenchment as higher-order representations. The features of visual stimuli often consist of color, shape, and positional information with various types of biases existing between learning these features vs. objects (Turk-Browne, Isola, Scholl, & Treat, 2008), similar to the effect of the stimulus-level differences noted in auditory SL. For example, when two features, such as color and shape, perfectly co-vary within each object in a triplet, participants struggle to identify acceptable triplets when tested on only one of the two features (either color or shape). However, when shape and color are decoupled during training and vary across objects, the underlying pattern for each feature can be learned independently. In terms of development, adults and children seem to show similar underlying neural processes when learning sequential information in the visual domain, with stable P300 responses across age groups to visual stimuli that are highly predictive of a target stimulus (Jost, Conway, Purdy, Walk, & Hendricks, 2015).

Touch is another modality in which SL has been studied. Conway and Christiansen (2005) investigated whether or not statistical structure could be learned purely from tactile input. They found that performance with tactile input is similar to performance in the visual modality, though auditory learning was superior to both when the same artificial grammar was used in each modality. Further theories point towards the use of SL as a basis for social understanding (Lieberman, 2000; Ruffman, Taumoepeau, & Perkins, 2012) and motor skill learning (Robertson, 2007).

These findings lead to interesting questions about what kinds of constraints are placed on learning due to the nature of stimuli in different sensory modalities. For example, auditory information is usually encountered in rapid succession and is quite transient in nature. Thus, basic sensory processing mechanisms for auditory input are tuned to this bias in presentation.

Visual input varies across time as well, but is much more stable and thus SL studies incorporating visual stimuli require longer inter-stimulus intervals to achieve the same levels of learning as in audition (Emberson, Conway, & Christiansen, 2011). One possible explanation for the patterns of similarity and differences in SL across domains is the existence of multiple modality-specific processes, each using the same underlying computational principles, but subject to different modality-specific constraints (Frost et al., 2015).

The evidence of SL across different modalities and domains suggests that such entrenchment might not be a language-specific phenomenon. Examples such as the incidental categorization of single tones into triplets due to frequent co-occurrence in a continuous stream (e.g. Saffran et al., 1999) and the extraction of statistical structure from visual scenes (e.g. Fiser & Aslin, 2002, Kirkham et al., 2002) provide compelling arguments for SL as a domain-general process of entrenchment. The construction of holistic units out of basic elements is a hallmark of entrenchment, and the wide range of research within the literature on SL captures the basic properties of a process which, as described above, may operate at various levels as a foundation for the formation of learned associations that underpin language learning and processing.

Brief Introductions to the Remaining Chapters

Research on SL, as introduced above, heavily relies upon using AGL paradigms to implicitly measure learners' sensitivity to various kinds of distributionally-defined input, with different kinds of associative relationships. In psycholinguistics, these studies are conducted with the ultimate aim of using AGL (and the assessment thereof) as a window into the processes that underlie language acquisition. This inspired the study described in Chapter 2, which set out to i) demonstrate that such AGL paradigms demonstrate test-retest reliability, and ii) examine the relationship between individual differences in actual language processing biases and sensitivity to a specific type of probabilistic dependency within an artificial grammar. Considering the widespread use of such paradigms, and the ongoing debate about the reliability of other commonly used SL paradigms (e.g., Siegelman, Bogaerts, & Frost, 2017), investigating their reliability fills a significant gap within the literature. Moreover, combining this with an examination of the relationship between individual differences in domain-general learning sensitivity and language processing biases serves to undergird one of the major claims of Christiansen and Chater (2008) – that language must be built upon a foundation of more basic cognitive mechanisms.

Chapter 3 is an extension of this work, and can be characterized as an attempt to better understand the ways in which learners actually use and store the different kinds of information embedded within AGL paradigms, like those evaluated for reliability in Chapter 2. While learners seem to be able to extract the abstract grammatical regularities that exist between items in SL and AGL paradigms (Reber, 1967), they are also clearly sensitive to the surface level fragment information to which they are exposed during training (Knowlton & Squire, 1994). Determining the extent to which learners are able to extract these higher-level relationships and the conditions that allow them to do so best gives us a window into the relationship between general learning and memory abilities and language learning and processing. The effects of extensive training on simple, rather than complex, items are evaluated in terms of how this kind of constrained input may improve learning outcomes, inspired by insights from the "starting small" literature (Elman, 1993). Furthermore, there is very little research into the retention of what is learned within AGL paradigms, a problem that this study seeks to address with a unique design that incorporates a follow-up test which takes place two-weeks after training. The importance of showing retention within such tasks is borne out of past research that suggests experience with specific kinds of linguistic constructions facilitates the later processing of the same kind of construction (Reali & Christiansen, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). If SL is indeed involved in this aspect of language processing, learned associations in a task like that reported in Chapter 3 should persist over time.

Chapter 4 of this dissertation reports a study seeking to uncover what happens when the neural network that language learning and processing relies upon is perturbed. In conjunction with Chapters 2 and 3, we can think of this project as an attempt to determine what happens when the cortical components of the network underlying the system studied within each of those chapters falls apart. While most research into the cognitive neuroscience of language has focused on the left-hemisphere and its involvement in language processing (e.g., Hagoort, 2014), a growing literature has identified that the right-hemisphere may play a complementary role (Beeman & Chiarello, 1998). However, this role has been mostly confined to higher-level, non-syntactic aspects of linguistic processing, such as discourse (Beeman, 1993) and prosody (Snow, 2000). The right-hemisphere is usually afforded little attention, particularly when it comes to discussions of "core" linguistic systems involving the frontal lobe (e.g., van der Lely & Pinker, 2014; Berwick, Friederici, Chomsky & Bolhuis, 2013).

Many of these claims flow out of the argument that humans evolved a new, highly specified language acquisition device, championed by Pinker and Bloom (1990) but first specified by Chomsky (1965). While these theories claim that language processing is subserved by specialized, uniquely adapted neural circuitry, other theories posit that humans rely on domain-general cognitive mechanisms for processing linguistic input, which would require a neural network that is not dedicated for language (e.g., Christiansen & Chater, 2008). Past research has indicated that the right-frontal lobe may be able to take over language function when homologous structures in the left-hemisphere are damaged, however, this work has been mostly confined to case studies (e.g. Holodny, Schulder, Ybasco, & Liu, 2002) and work in much younger populations (e.g., Thal et al., 1991; Vicari et al., 2000). Thus, the research reported in Chapter 4 attempted to determine the extent of plasticity in the mature brain's language network, notably the ability of the right-frontal lobe to take over what is usually the left-frontal lobe's typical role in language processing. Evidence of plasticity and contralateral reorganization (with maintenance of function) would support the idea that the left-frontal lobe is not uniquely adapted for language and undermine the idea that language relies on a unique, highly-specified cortical network.

To conclude, the final chapter will explore the ways in which these studies tie together. It will focus on how this combination of research provides evidence that the cognitive and neural underpinnings of language learning and processing are characterized by individual differences, entrenchment, and plasticity by relating them back to the existing literature and expanding upon it.

References

Altmann, G. T. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15250-15251.

Arciuli, J., Torkildsen, J. V. K., Stevens, D. J., & Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*, *5*(July), 747.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science*, *9*(4), 321-324.

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain and Language*, *44*(1), 80-120.

Beeman, M. J., & Chiarello, C. (1998). Complementary right-and left-hemisphere language comprehension. *Current Directions in Psychological Science*, *7*(1), 2-8.

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences*, *17*(2), 89–98.

Bly, B. M., Carrión, R. E., & Rasch, B. (2009). Domain-specific learning of grammatical structure in musical and phonological sequences. *Memory & Cognition*, *37*(1), 10–20.

Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychological Review*, *70*(4), 323–348.

Christiansen, M. H. (in press). Implicit statistical learning: a tale of two literatures. *Topics in Cognitive Science*. DOI: 10.1111/tops.12332.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The Behavioral and Brain Sciences*, *31*(5), 489-508; discussion 509–58.

https://doi.org/10.1017/S0140525X08004998

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *The Behavioral and Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525X1500031X

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

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356–371.

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*, 905–912.

Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences*, *1145*(1), 113-131.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, *14*(4), 597–612.

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.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, *48*(1), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4

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.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

Franco, A., & Destrebecqz, A. (2012). Chunking or not chunking? How do we find words in artificial language learning? *Advances in Cognitive Psychology*, *8*(2), 144–54.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: how conscious? *Frontiers in Psychology*, *2*(September), 1–12.

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

Frost, R. L., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70-74.

Frost, R. L., Monaghan, P., & Tatsumi, T. (2017). Domain-general mechanisms for speech segmentation: The role of duration information in language learning. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 466.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67-B74.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32.

Gobet, F. (2016). Entrenchment, Gestalt Formation, and Chunking. In H.J. Schmid (ed.) *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge, 245-268*. Washington, D.C.: De Gruyter Mouton.

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

Gupta, P., & Dell, G. S. (1999). The emergence of language from serial order and procedural memory. *The emergence of language*, 447-481.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, *28C*, 136–141.

Hartsuiker, R. J., & Moors, A. (2016). On the automaticity of language processing. In H.J. Schmid (ed.) *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*, 201.

Holodny, A. I., Schulder, M., Ybasco, A., & Liu, W.C. (2002). Translocation of Broca's Area to the Contralateral Hemisphere as the Result of the Growth of a Left Inferior Frontal Glioma. *Journal of Computer Assisted Tomography*, *26*(6), 941–943. https://doi.org/10.1097/00004728-200211000-00014

Jost, E., Conway, C., Purdy, J. D., Walk, A., & Hendricks, M. (2015). Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Research*, *1597*, 95-107.

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.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 79–91. https://doi.org/10.1037/0278-7393.20.1.79

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–149.

Lashley, K. S. (1951). *The problem of serial order in behavior* (Vol. 21). Bobbs-Merrill.

Lieberman, M. D. (2000). Intuition: A social cognitive neuroscience approach. *Psychological Bulletin*, *126*, 109–137.

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication theory*, *84*, 486-502.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77-80.

McCauley, S. M. & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *Mental Lexicon*, *9*, 419–436.

McCauley, S.M. & Christiansen, M.H. (in press). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*. DOI: 10.1037/rev0000126.

Miller, G. A. (1958). Free recall of redundant letter strings. *Journal of Experimental Psychology*, *56*(6), 485–491.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *The American Journal of Psychology*, 176–185.

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*(3), 545–564.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707–784.

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Developmental science*, *21*(4), e12593.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1–23.

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

Reber, A. S. (1976). Implicit learning of synthetic languages: the role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 88–94.

Robertson, E. M. (2007). The serial reaction time task: Implicit motor skill learning? *Journal of Neuroscience*, *27,* 10073–10075.

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, *30*(1), 87-104.

Rubenstein, H. (1973). Language and probability. In G.A. Miller (Ed.) *Communication, language, and meaning: Psychological perspectives*, 185-195. New York: Basic Books, Inc.

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 (New York, N.Y.), 274*(5294), 1926–8.

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. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.

Schmid, H. J. (2007). Entrenchment, salience, and basic levels. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of cognitive linguistics* (pp. 117-138). Oxford, England: Oxford University Press.

Schmid, H. J. (2016). A framework for understanding linguistic entrenchment and its psychological foundations. In H.J. Schmid (ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*, 9-36.

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning : Current pitfalls and possible solutions. *Behavior Research Methods*, *49*, 418–432. https://doi.org/10.3758/s13428-016-0719-z

Snow, D. (2000). The emotional basis of linguistic and nonlinguistic intonation: Implications for hemispheric specialization. *Developmental Neuropsychology*, *17*(1), 1-28.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC neuroscience*, *10*(1), 21.

Thal, D. J., Marchman, V., Stiles, J., Aram, D., Trauner, D., Nass, R., & Bates, E. (1991). Early Lexical Development in Children with Focal Brain Injury, *527*, 491–527.

van der Lely, H. K., & Pinker, S. (2014). The biological basis of language: insight from developmental grammatical impairments. *Trends in Cognitive Sciences*, *18*(11), 586-595.

Vicari, S., Albertoni, A., Chilosi, A. M., Cipriani, P., Cioni, G., & Bates, E. (2000). Plasticity and reorganization during language development in children with early brain injury. *Cortex*, *36*, 31–46.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271.

CHAPTER TWO

A psychometric evaluation of statistical learning paradigms[1]

Abstract

Artificial grammar learning (AGL) paradigms have been used to study our ability to learn complex sequential structures such as those thought to underlie language for decades, yet the psychometric properties of commonly used paradigms have not been examined until recently. This study compares the reliability of both the standard AGL and a more recent experimental measure that embeds AGL within a serial response time task (Misyak & Christiansen, 2010). Additionally, each paradigm's correlation with language processing is examined to determine how well they tap into the cognitive processes thought to underlie language. Analyses include test-retest reliability as well as an examination of the relationship to natural language processing while taking working memory into account. The results suggest that performance on AGL paradigms is relatively stable over time - but not for all kinds of measures - and that some measures of learning correlate with individual differences in language processing abilities while others do not.

Introduction

The idea of 'statistical learning' has become an important part of the literature on language processing, development, and evolution (see Armstrong, Frost, & Christiansen, 2017

[1] Co-authored with Morten H. Christiansen; currently in preparation for publication.

for an overview). We can define statistical learning (SL) as the process by which learners uncover the structure of the input from its distributional properties (Frost, Armstrong, Siegelman, & Christiansen, 2015). This type of learning has been said to play an important role in language acquisition (Saffran, Aslin, & Newport, 1996), and contribute to individual differences in adult language processing (Misyak, Christiansen, & Tomblin, 2010). However, for all of its influence, there is a profound lack of psychometric testing on this construct.

As SL has become a such a large part of the literature on language acquisition and processing, a deeper investigation into the methods used to measure this ability has become an obvious necessity. Typically, such learning is measured by examining how well participants perform on a post-test immediately after a period of exposure to items that are generated by an underlying set of rules. This is most frequently done either in the form of a two-alternative forced choice task, wherein participants must correctly select the item that could have been generated by the rules they had been exposed to rather than an alternative item, or in a familiarity judgment task that requires participants to say whether or not an item that they see in the post-test seems like an item that followed the same rules that generated the training set using only their gut instinct. While there is a dearth of literature on the psychometric properties of such paradigms, Siegelman and Frost (2015) have reported reliability for a variety of SL tasks across both visual and auditory modalities. Four of their five tasks only utilized offline judgment tasks like those described above, although their serial response time task did also demonstrate reliability over time. Interestingly, these tasks did not correlate with a battery of other general cognitive abilities, such as verbal working memory or intelligence, and also tended not to correlate with one another, although there were a few exceptions.

Siegelman, Bogaerts, and Frost (2017) also reported test-retest reliability for a SL task that was specifically designed to have psychometrically valid properties. However, neither of these in-depth looks at the psychometric properties of SL paradigms included any corresponding sentence processing tasks, so as to examine their hypothesized relation to language. While others have begun to examine the reliability of a variety of SL measures, there is still a major gap when it comes to artificial grammar learning (AGL) paradigms, especially those used most frequently within the literature.

Reber's (1967) initial studies on implicit learning using an AGL paradigm in many ways formed the basis for the modern study of SL. This initial contribution demonstrated that extensive training on strings of letters generated by an underlying grammar facilitated recall of grammatical strings relative to random ones. The separate literatures on implicit learning and SL have continued along parallel paths in many ways (Perruchet & Pacton, 2006), while maintaining important differences in terms of the way they focus on stored units (chunks) versus probabilistic cues, respectively (Christiansen, in press).

Yet the few studies that do query the reliability of SL paradigms do not yet extend to those most commonly used in implicit learning studies, namely AGL paradigms. The present study aims to correct this oversight, by examining the test-retest reliability of both the standard AGL paradigm, with its offline measures of putative implicitly learned words, bigrams, and trigrams, and a version that incorporates online reaction time measures throughout learning, in addition to traditional offline measures of learning (Misyak & Christiansen, 2010). When constructing AGL tasks, we must think about how they reflect the demands of actual language learning and processing. Language production and comprehension for the most part takes place unconsciously, generally with little effort, and as such, online measures of learning, such as the

RT based metrics used in the AGL-SRT task, may better correlate with language processing data, as has been found in past research (Misyak & Christiansen, 2010, 2012).

In fact, Christiansen (in press) suggested characterizing tasks with this kind of online testing as "processing-based," as opposed to the more common, offline "reflection-based" measures. This dichotomy reflects the degree to which participant's responses are or are not the product of an explicit judgment, a growing concern among those studying how task demands can affect outcome measures in both SL and AGL studies (Arciuli, Torkildsen, Stevens, & Simpson, 2014; Bertels, Franco, & Destrebecqz, 2012; Franco, Cleeremans, & Destrebecqz, 2011; Franco & Destrebecqz, 2012; Isbilen, McCauley, & Christiansen, 2017; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012). Removing the "consciousness filter" found within standard AGL paradigms could get us closer to measuring important changes in the way that participants process input as it is happening, throughout learning. This feature meshes nicely with the idea that learning and processing may be one and the same within the context of SL and language (Christiansen, in press; Christiansen & Chater, 2016) that also motivates this work, as the AGL-SRT task in particular seeks to measure learning as it is happening. In this way, the measurement of learning takes place in the moment that the input is at the critical "Now-or-Never" bottleneck of processing, rather than merely looking for the effects after the fact.

By looking closely at these various methods for measuring learning, this study aims to ensure that commonly used AGL paradigms provide reliable data on individual differences in learning abilities. In order to accurately determine the relationship between language outcomes and learning abilities, researchers require a good and reliable measure (Armstrong et al., 2017; West, Vadillo, Shanks, & Hulme, 2017). While some previous research has examined the relationship between AGL abilities and language learning within the same study, this research has assumed that such measures are accurately measuring a stable trait. Knowing that we are indeed measuring individual differences in statistical learning abilities is a topic of great concern to the field, as a full appreciation of individual differences in SL abilities would serve to highlight the impact of experience and processing on language abilities (Kidd, Donelly, & Christiansen, 2018).

As such, the present study also seeks to establish the validity of this new class of online measures of learning by pitting them against standard offline learning measures in terms of their ability to predict language processing abilities. By doing so, we aim to ensure that the construct we are measuring is not only stable across time within individuals, but actually relates to language learning and processing. Again, we want to examine whether or not the outcome measures from the AGL-SRT will outperform those provided by the Standard AGL paradigm in terms of their relationship to processing relevant linguistic constructions.

Past research, suggests that individual differences in the ability to process dependencies within an artificial grammar are linked to biases in natural language processing (Conway, Bauernschmidt, Huang, & Pisoni, 2010; Misyak & Christiansen, 2010, 2012; Misyak et al., 2010). The current study, in addition to examining the reliability of these AGL paradigms, also seeks to extend the literature on this relationship, as it would provide additional evidence for the argument that language learning and processing rely on contributions from more basic, domain-general cognitive processes (e.g., Christiansen & Chater, 2008). This study can be considered an extension of prior work that demonstrated a relationship between the learning of adjacent dependencies using the AGL-SRT paradigm and increased interference when processing local noun-verb mismatches in a sentence processing task (Misyak & Christiansen, 2010). Therefore, one goal of the present study is to demonstrate a similar relationship while also verifying the reliability of the paradigms used.

Method

Participants

In order to test these hypotheses, 48 participants were recruited from the undergraduate population of Cornell University (33 female; mean age: 19.44, range: 18-22). All participants were native English speakers. Participants were excluded from this study if they got under 75% of the comprehension questions within the sentence processing task correct, if they responded inaccurately on over 15% of trials with the AGL-SRT task, or failed to show-up for their second session.

Statistical Learning Tasks

Materials

Both statistical learning tasks incorporated the same underlying artificial grammar featuring adjacent dependencies, and the same amount of exposure. Four different lists of nonce words were used to ensure that each task at each session had a unique set of items (see **Table 1**). Between participants, the order of items within these sets was randomized to control for any bias in terms of the learnability of specific associations.

Nonce word lists for AGL paradigms

List A	***List B***	***List C***	***List D***
cav	biff	bix	dak
dupp	fis	gens	hep
hes	vot	jic	jux
kif	klor	leb	tiz
loke	lum	meep	mib
neb	nib	tood	pell
pilk	rem	rud	rauk
sep	sig	tam	tash

Table 1. The four sets of nonce words used in this task.

Training consisted of 7 blocks of 64 trials. Each trial contained a sequence of four nonce words that appeared on the screen simultaneously. Within each block, no items were repeated, so that the participant saw all possible grammatical sequences (8x2x2x2) within each block. As described above, each nonce word could only be followed by two others; the first word within each trial is unpredictable, meaning there were no relevant statistics across trials (see **Figure 1** for an overview of the dependencies within the grammar). Each of the eight nonce words could only be followed by two of the other words. For example, if a trial started with nonce word G, it could only be followed by either H or A. If A were the next word, it could only be followed by either B or C, and so on. This grammar was designed to be a window into each participant's sensitivity to proximal, adjacent relationships between elements in a sequence, as each item directly predicts the appearance of the subsequent item.

Item	A	B	C	D	E	F	G	H
A	--	.5	.5	--	--	--	--	--
B	--	--	.5	.5	--	--	--	--
C	--	--	--	.5	.5	--	--	--
D	--	--	--	--	.5	.5	--	--
E	--	--	--	--	--	.5	.5	--
F	--	--	--	--	--	--	.5	.5
G	.5	--	--	--	--	--	--	.5
H	.5	.5	--	--	--	--	--	--

Figure 1. Chart depicting the probabilistic dependencies within the grammar shared by the two AGL paradigms.

Procedure

Participants underwent two sessions of testing, with each session involving the completion of two statistical learning tasks (both the Standard AGL and the AGL-SRT), along with the verbal working memory task (Session 1) or the sentence processing task (Session 2) taking place between them. The order in which participants completed the statistical learning tasks was balanced, both within and between participants. Half of the participants did the AGL-SRT first in Session 1 and second in Session 2, while the other half completed the Standard AGL first in Session 1 and second in Session 2. The two sessions took place approximately two weeks apart.

There were major differences between the designs of the Standard AGL and the AGL-SRT. The Standard AGL paradigm attempted to replicate the traditional task demands used in past research on the topic, while maintaining the same kind of audiovisual exposure that was used in the AGL-SRT. This meant that the presentation of the stimuli was not dependent on participant engagement in the task. Trials advanced at a predetermined rate, and the exposure to the audiovisual items was passive in nature. In the Standard AGL paradigm, while the stimuli were shown on the screen, they were also presented aurally for 550ms with an inter-stimulus interval of 150ms. See **Figure 2** for an illustration of a single trial within the task. Because of this passive training, the only learning measures that could be obtained came after training and required an explicit decision to be made.

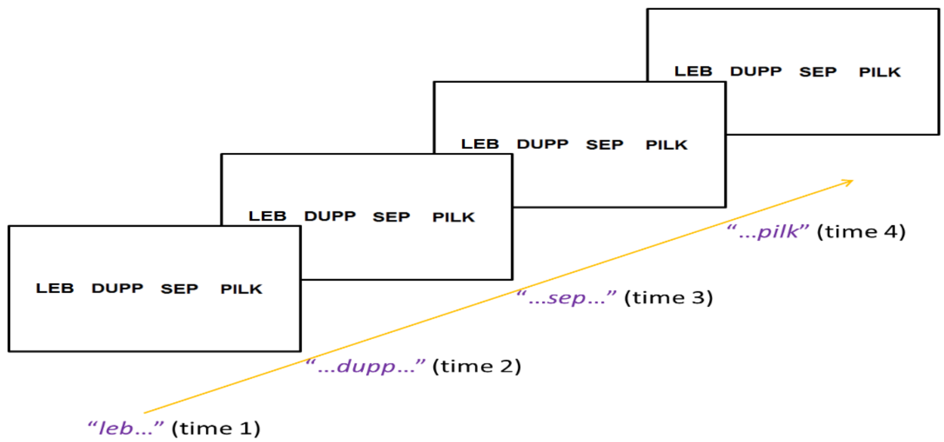


Figure 2. Visual representation of a single trial in the Standard AGL paradigm. A set of four nonce words would appear on the screen, before being presented aurally via headphones in sequence.

The first of these offline tasks consisted of a bi-/tri-gram grammaticality judgment. In this task, participants were instructed to judge whether or not a sequence of two or three nonce words seemed familiar by responding either “Yes” or “No” to each trial. Each participant underwent sixty-four trials of this type, half of which were bigrams, and half of which were trigrams. Half of the items were grammatical, while half were ungrammatical.

The other offline task consisted of a traditional two-alternative forced choice (2AFC) task. For this, participants were presented with two different four-word (just like in training) items one after the other and asked to choose whether the first or the second followed the rules of training. Half of the initially presented items were grammatical, while the latter item was grammatical the other half of the time. This task consisted of sixteen trials.

The AGL-SRT paradigm also utilized traditional offline measures of learning, including the exact same bi-/tri-gram grammaticality judgment task as the Standard AGL. However, to better mirror training, a slightly different prediction task replaced the traditional 2AFC found within the Standard AGL paradigm (described below). In this task, participants would click through the first three words as they did throughout training, with an aural presentation of the next word in the sequence following their click. After clicking on the third column, however, there was no cue for the final nonce word. Participants were instead instructed to use their gut and to guess on the final item, allowing us to test their accuracy in predicting the final word in the sequence.

Most importantly, the AGL-SRT featured an online ‘cover task’ throughout training (see **Figure 3** for an illustration of a single trial in this paradigm). In this paradigm, the first word was presented aurally after the visual overlay had been on-screen for 250ms, while the presentation of the second word did not occur until the participant clicked on an item in the first column. The active nature of the training blocks engaged participants by requiring them to click on each nonce word after an aural presentation. This allowed for a comparison of RTs between the predictable latter and unpredictable initial elements within each trial sequence. Trials in which participants took greater than 2500ms to respond were discarded from the analyses. This resulted in the removal of under 1% of the data at either session.

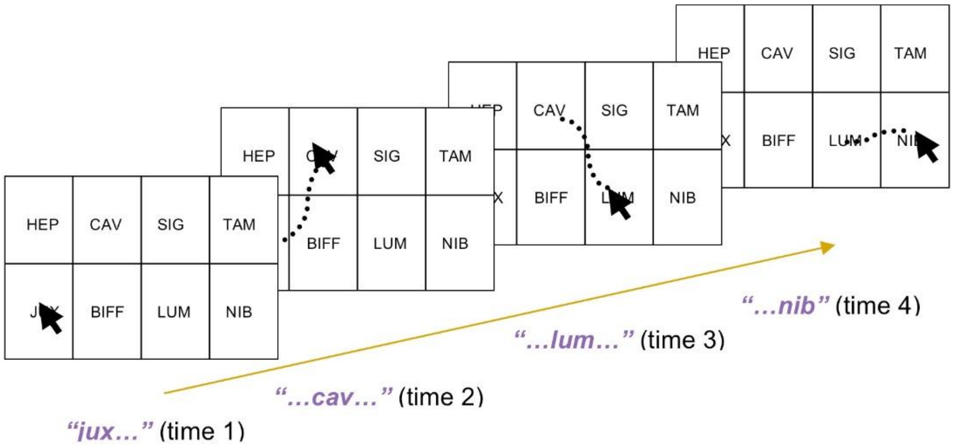


Figure 3. Visual representation of a single trial in the AGL-SRT paradigm. The initial nonce word (e.g. *jux)* in the sequence would be heard after the grid appeared on-screen. After clicking on that word, the next would be presented aurally via headphones (e.g. *cav)*, at which point the participant would click where it was located on the grid, and so on.

This active online task was designed to be potentially more sensitive to individual differences than traditional offline measures of learning. The ability to measure learning throughout training rather than merely at the end of the task should allow for easier identification of learners. Additionally, the AGL-SRT paradigm itself is more similar in terms of task demands to standard sentence processing paradigms, like the one used in this study.

<u>Sentence Processing Task</u>

In order to determine how well each paradigm taps into the same cognitive mechanism that underlies language processing, participants also engaged in a typical sentence processing task at the second session between the two AGL paradigms. This took the form of a self-paced reading task in which participants pressed a designated button on the keyboard to progress word-by-word through each sentence, before answering a simple yes/no comprehension question (Just, Carpenter & Woolley, 1982). The sentence materials consisted of a variety of sentence types, but the construction of interest in this study was contained within sentences that had a local noun-verb number agreement match or mismatch. In a match-sentence (*The key to the cabinet was rusty*), both nouns within the sentence were singular, so there was no conflicting adjacent information. In a mismatch-sentence (*The key to the cabinets was rusty*), the noun adjacent to the verb was instead plural. This was expected to lead to increased reading times at the main verb relative to the match condition as a result of interference upon processing the local mismatch (Pearlmutter, Garnsey & Bock, 1999). Note that both sentence types were perfectly grammatical. Importantly, this kind of local relationship is what participants needed to be attuned to in order to learn the grammar in the AGL tasks. Thus, we could examine participants' relative reading times as difference scores between the mismatch and the match sentences, of which there were ten apiece, as a window into their sensitivity towards this kind of adjacent dependency and a potential natural language processing correlate with AGL performance.

Verbal Working Memory Task

As reported in various other studies (for a review, see Farmer, Fine, Misyak & Christiansen, 2017), sentence processing abilities have been linked to individual differences in verbal working memory. In order to assess how well performance on our AGL tasks can explain individual differences in sentence processing skill relative to the contributions from individuals' verbal working memory abilities, we utilized the Waters and Caplan (1996) reading span task as an assessment of our participants' verbal working memory. Participants answered yes/no semantic plausibility judgments for several sentence sets of varying length, with each sentence presented one at a time. At the end of each set, participants were prompted to recall and say out loud all sentence-final words in that set, in order. The total number of sentences contained within each set increased incrementally from 2 to 6, with a total of three trials at each level. Reading span was defined as the maximum level at which a participant correctly recalled all sentence-final words in 2 out of 3 trials, with no more than one failed trial at the preceding levels and with half-a-point added if one trial had been correct at the next highest level.

Results

Learning Outcomes

Participants showed above-chance performance on each of the offline statistical learning tasks at both sessions within each paradigm. All of the statistical tests below were two-tailed one-sample t-tests with a test value of 50. Means and standard deviations for the relevant tasks can be found in **Table 2**.

Summary of participant performance on SL tasks across sessions

	Mean (% correct)		Standard Deviation		Minimum (% correct)		Maximum (% correct)	
Learning Measure	S1	S2	S1	S2	S1	S2	S1	S2
Standard AGL bi-/tri-gram	59.08	56.64	8.58	8.46	42.19	45.31	87.50	79.69
Standard AGL 2AFC	59.77	63.41	12.02	12.70	43.75	37.5	87.50	93.75
AGL-SRT bi-/tri-gram	54.43	55.60	6.06	6.96	43.75	43.75	68.75	73.44
AGL-SRT prediction	56.51	56.38	12.83	11.35	18.75	31.25	75.00	81.25

Table 2. Descriptive statistics for participant performance on each of the offline SL tasks in both paradigms, across sessions. Session 1 denoted by "S1" and Session 2 denoted by "S2".

In the Standard AGL, participants performed significantly above-chance at Session 1 on both the bi-/tri-gram familiarity judgments ($t(47) = 7.33$, $p < .001$) and the 2AFC task ($t(47) =$

5.63, $p < .001$). At Session 2, participants demonstrated learning on both the bi-/tri-gram familiarity judgments ($t(47) = 5.44$, $p < .001$) and 2AFC task ($t(47) = 7.32$, $p < .001$). For the AGL-SRT paradigm, significant learning effects were seen for both the bi-/tri-gram familiarity judgments ($t(47) = 5.06$, $p < .001$) and the prediction task ($t(47) = 3.52$, $p = .001$) at Session 1. At Session 2, participants again performed significantly above chance on the bi-/tri-gram familiarity judgments ($t(47) = 5.57$, $p < .001$) and the prediction task ($t(47) = 3.90$, $p < .001$). Charts depicting the group-level learning trajectories for the online component of the AGL-SRT can be seen in **Figures 4** and **5**, for Sessions 1 and 2, respectively.

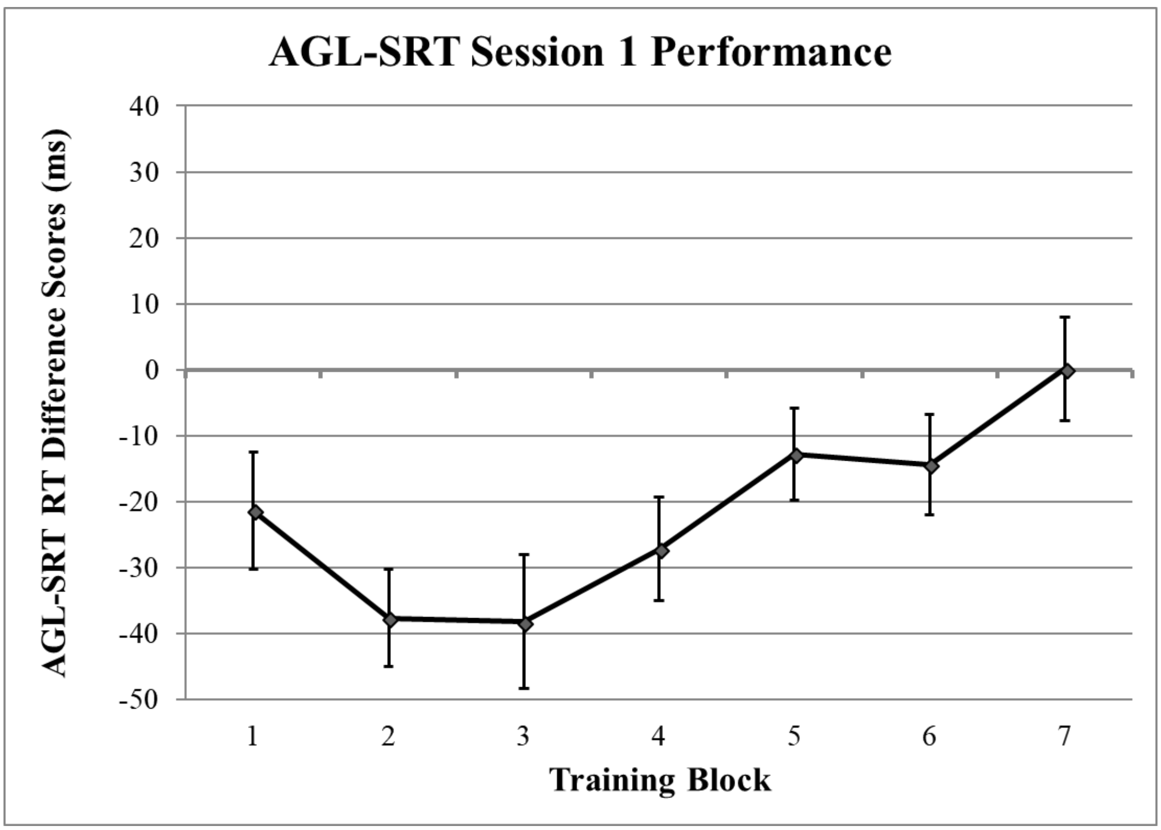


Figure 4. Group level performance across each block of trials in the AGL-SRT paradigm at Session 1. Error bars represent SE.

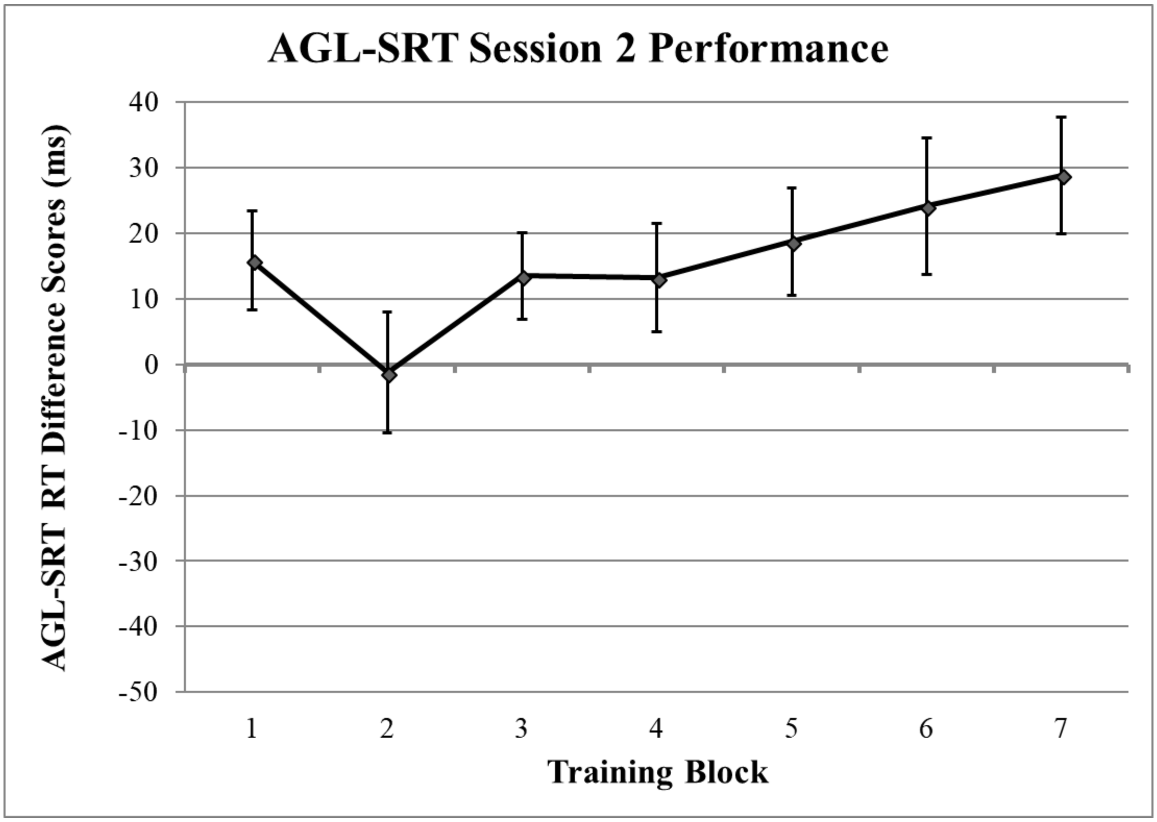


Figure 5. Group level performance across each block of trials in the AGL-SRT paradigm at Session 2. Error bars represent SE.

Test-Retest Reliability

The reliability of both statistical learning paradigms was evaluated by examining the correlation between first and second session task performance for each participant. This was done for each of the various outcome measures that were collected for each paradigm.

For the Standard AGL paradigm, test-retest performance on the bi-/tri-gram familiarity judgments was found to be reliable ($r = .483$, $p = .001$; see **Figure 6**). However, performance between sessions for the 2AFC task was not correlated ($r = .191$, $p = .194$).

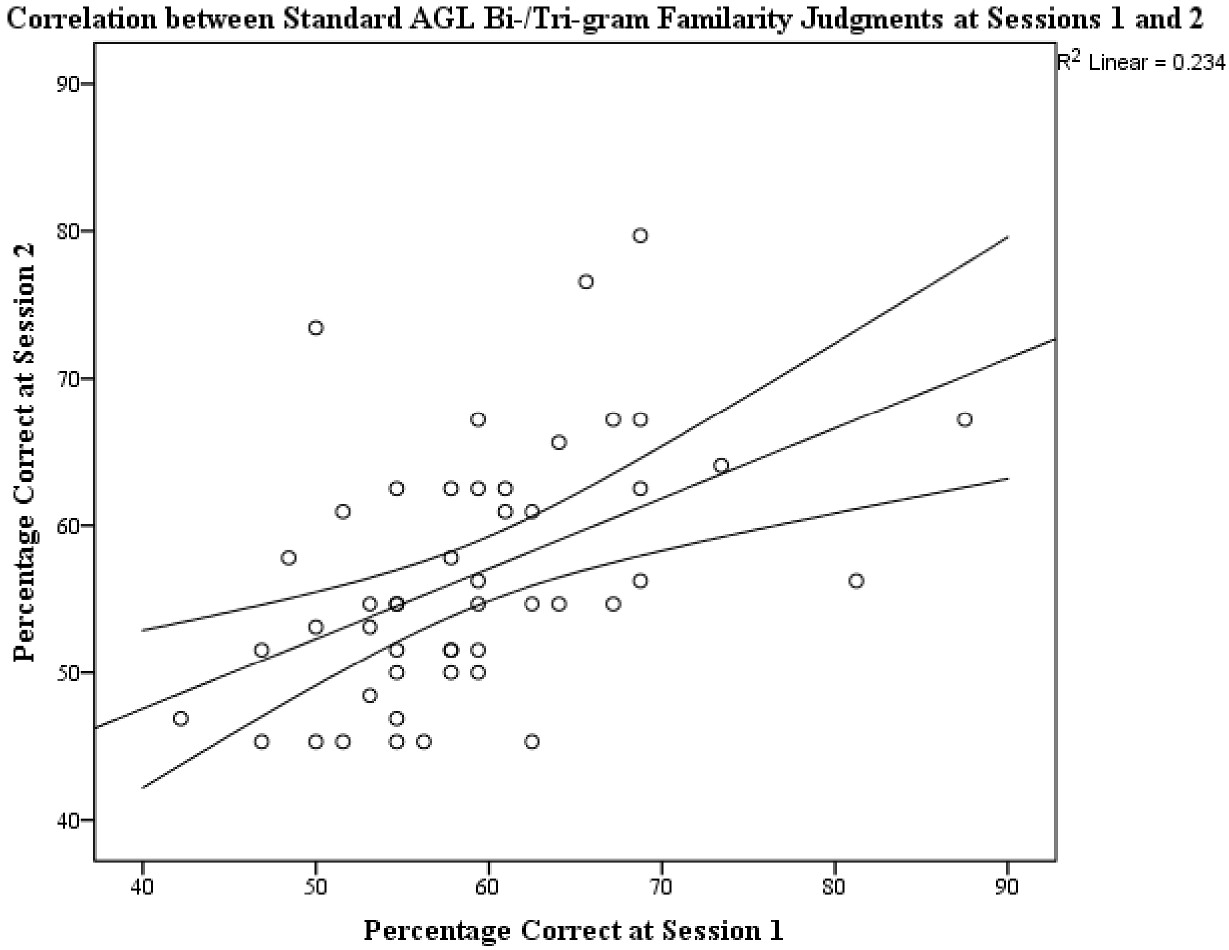


Figure 6. Test-retest performance on the bi-/tri-gram familiarity judgment task within the Standard AGL paradigm. Lines represent the fit and 95% CI.

Within the AGL-SRT paradigm, Session 1 and 2 performance on the bi-/tri-gram familiarity judgments was found to be marginally reliable (r = .25, p = .086). Performance on the prediction task across sessions was not reliable (r = .234, p = .110).

Over the course of the seven training blocks, we could also examine whether or not each participant exhibited learning within each block, by comparing the difference between their RTs to unpredictable vs predictable elements against the group's mean difference score within each block, similar to the calculation used in Kaufman et al. (2010). This allowed for the identification of the best learners in the group within each block, with those who showed a stronger learning score earning a point for each block in which they outperformed their fellow participants. Points could be accrued in blocks 3 – 7, as this is the point in the task where evidence of learning has been found in past studies using the same paradigm. A learner who responded more quickly to the predictable vs unpredictable items than the group mean in every block would thus earn a score of “5” on this measure, while a participant who never showed a stronger learning effect in any of the blocks would earn a score of “0”.

There are some minor differences between this computation and the one originally used by Kaufman et al. (2010), as in this study the participants’ average difference score within each block was compared to their peers’, whereas they had compared the learning score within each block to the overall group mean across all blocks. Our modification should do a better of identifying those who learned the task well, as examining individual performance within block 7, for example, against a group mean that includes block 1 in its calculation would tell us whether or not that participant showed signs of learning, but would not do as good a job differentiating their learning ability from that of other participants – most participants would be expected to get a score of “1” in that block using the original Kaufman et al. (2010) calculation, whereas we would have a 50/50 split. In this way, our measure is a better measure of individual learning abilities, while their measure may do a better job of identifying group-level performance on the task. Another slight difference is that in the present study, difference scores were extracted for each trial individually and averaged for each block, whereas in the Kaufman et al. (2010) implementation, scores for predictable and unpredictable elements were extracted separately before computing the participant’s average within each block.

Critically, performance on the active, online RT difference score task within the AGL-SRT was found to be stable across time, with a significant correlation between scores at each session ($\tau = .286$, $p = .013$; see **Figure 7**). Kendall's tau was used, as the Kaufman et al. (2010) calculation described above generates a rank-order for participants.

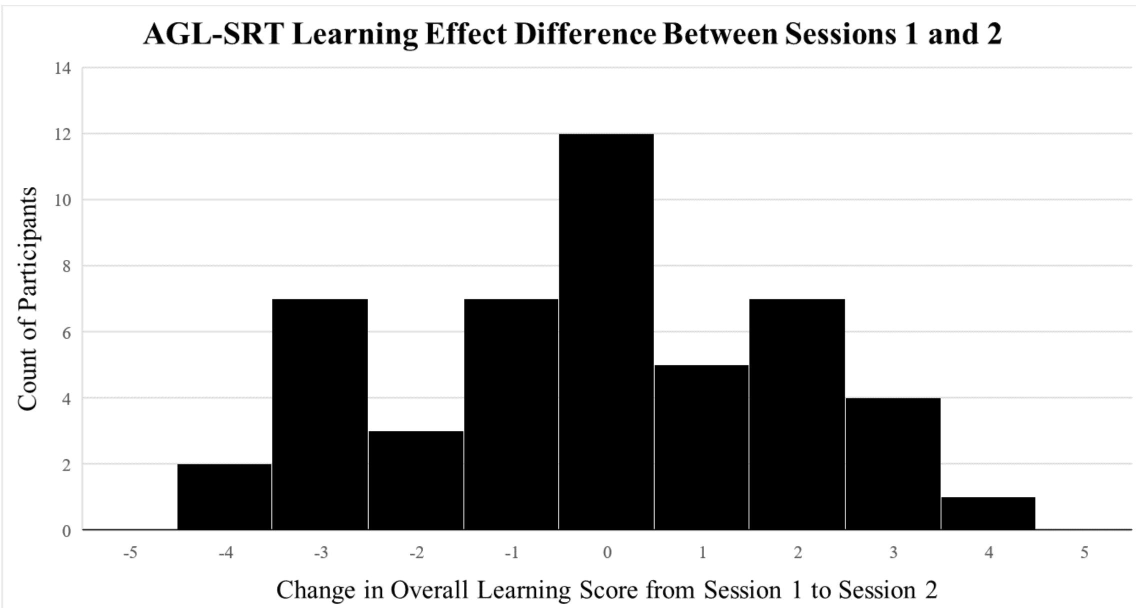


Figure 7. Participants could accrue a total of 0-5 points at each session for showing stronger signs of learning than their peers in each block of training within the AGL-SRT paradigm. This chart depicts the number of participants whose learning score changed by each possible amount from Session 1 to Session 2. A score of -2 on this chart could represent, for example, a participant who showed better learning than the group in two blocks at Session 1, and then in zero blocks at Session 2. Overall, we see that most participants did not show dramatic swings in terms of where they fell as learners relative to their peers between sessions.

AGL, vWM, and Sentence Processing

Participants were found to take significantly longer to process sentences with a local noun-verb mismatch relative to those that contained a match, with an effect found at the main verb (mean RT difference: 22.47ms; range: min = -80.4ms, max = 194.3ms; standard deviation = 59.86ms; $t = -2.731$, $p = 009$; see **Figure 8**); all RTs within the sentence processing task were length-adjusted based on the number of letters in each word (Ferreira & Clifton, 1986). The only significant correlation (see **Table 3** for all correlations) between AGL performance and sentence processing abilities was found for the bi-/tri-gram familiarity judgment task within the Standard AGL paradigm at Session 2, the same session at which the sentence processing task was administered, and the difference score between RTs at the main verb in match versus mismatch sentences ($r = .331$, $p = .022$; see **Figure 9**).

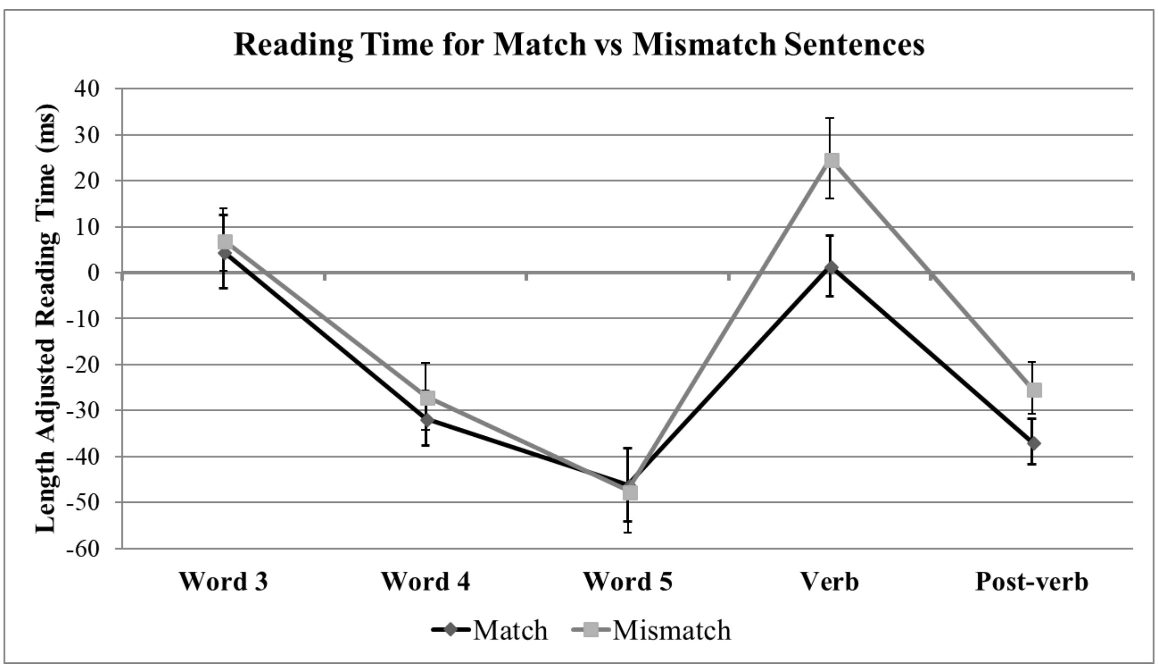


Figure 8. Length adjusted reading time for each word in the Match and Mismatch conditions of the Sentence Processing task. Error bars represent SE.

Correlation table for test-retest performance

	Correlation coefficient (p-value)	
Task	Session 1	Session 2
Standard AGL 2AFC	-.048 (.745)	.162 (.271)
Standard AGL bi-/tri-gram	.198 (.178)	.331 (.022)
AGL-SRT Prediction	-.171 (.245)	.194 (.187)
AGL-SRT bi-/tri-gram	-.244 (.094)	.132 (.372)
AGL-SRT RT Difference Score	.073 (.496)	-.085 (.426)

Table 3. Correlations between difference in RT between match and mismatch sentences at the main verb within the sentence processing paradigm and performance on the each of the learning measures from both AGL paradigms. All correlation coefficients are Pearson's r, except for correlations reported for the AGL-SRT RT Difference Score, which were calculated using Kendall's tau.

Verbal working memory abilities were significantly correlated with performance on the bi-/tri-gram familiarity judgments within the Standard AGL at both sessions (Session 1: $r = .518$, $p < .001$; Session 2: $r = .358$, $p = .013$), and with performance on the Standard AGL paradigm's 2AFC task at Session 2 ($r = .325$, $p = .024$), but not with performance on any of the tasks within the AGL-SRT paradigm. Performance on the verbal working memory task also showed a marginal correlation with sentence processing ability ($r = .234$, $p = .109$). Notably, when controlling for working memory ability, the correlation between performance on the sentence processing task and the bi-/tri-gram familiarity judgement task in the Standard AGL paradigm becomes marginal ($r = .272$, $p = .064$).

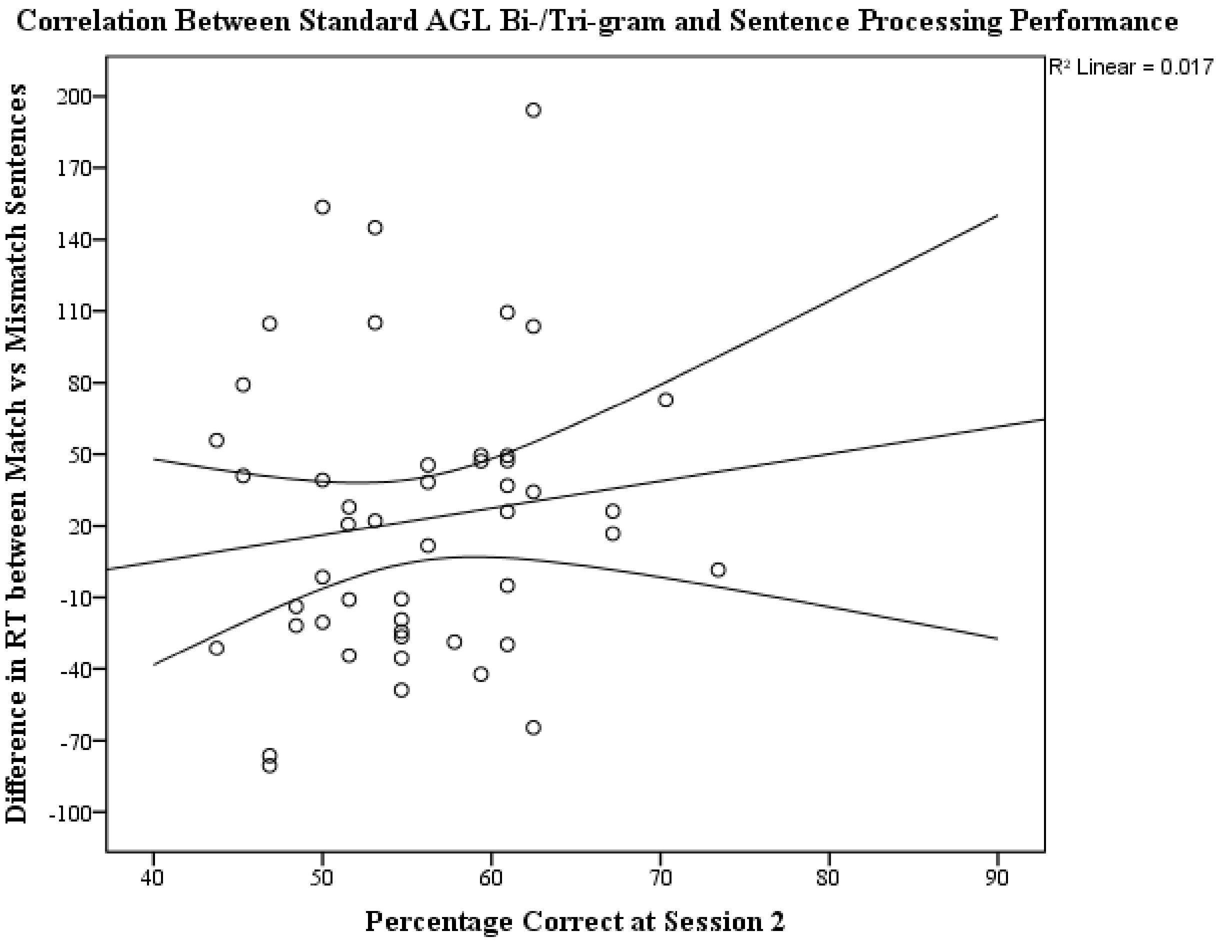


Figure 9. Relationship between learning on the Standard AGL as measured by the bi-/tri-gram familiarity judgment task and language processing ability, as measured in the sentence processing task. Participants who were better at learning the underlying grammar showed increased sensitivity to the mismatch condition when reading the main verb in such sentences.

Discussion

Performance on the critical tasks within both the traditional, offline Standard AGL and newer, online AGL-SRT paradigms showed moderate test-retest reliability. It is encouraging to find that these tasks which have formed the basis for much of the literature on this cognitive ability demonstrate stability over time within individuals. However, it is also somewhat surprising that the correlation coefficients were not found to be stronger. Importantly, the moderate-to-weak correlations found for test-retest reliability within this study suggest that finding correlations between tasks, such as between sentence processing measures and AGL performance, that are thought to tap into the same underlying cognitive processes were always going to be hard to find.

Interestingly, the weakest test-retest performance was found for the Standard AGL paradigm's 2AFC task, which has long been the gold-standard within the literature. This echoes other recent work from the SL literature that suggests the 2AFC may not always be the most effective task and is particularly unsuited for examining individual differences (Siegelman et al., 2017). The small number of trials within the 2AFC task, nearly chance performance across much of the sample, and a small range of difficulty between items have all been suggested to be potential limitations, and these were all present within the current study's implementation of the classic task. The set of findings in the present study suggests that n-gram familiarity judgments may be a better measure of learning when one is confined to using offline tests.

Merely marginal performance on the offline tasks within the AGL-SRT paradigm is also worth noting. Part of the issue here may have been due to task demands, as the test phase – particularly for the bi-/tri-gram familiarity judgments – was quite different from the training. Slightly modifying this particular task to more look more like training would be wise in the future. However, the critical outcome measure within the AGL-SRT was found to be reliable, which bodes well for future active, online, process-based measures of learning. It is also worth noting that overall learning performance was stronger in the Standard AGL, perhaps due to the fact that in this task participants always received the visually presented sequence of nonwords as they heard them, possibly strengthening learning, particularly of bi-/tri-gram information. Although subjects do see the nonce words in the AGL-SRT task, they also see the distractors and thus may not get the same benefit.

Additional reasons for the weaker than desired test-retest reliability could also be related to factors pertaining to the testing conditions that participants faced within this study. Participants engaged in three tasks at each session, and the total duration of each session from start to finish was approximately 90 minutes. This likely caused fatigue, which would have been an extraneous variable that affected performance and could have done so differently at each session. Additionally, while a two-week interval between sessions is typical for test-retest reliability studies, some recent studies have extended this much further, with months between sessions (e.g. Siegelman et al., 2017).

Beyond the psychometric findings within this study, the relationship uncovered between the learning of an AGL defined by adjacent dependencies and the processing of sentences that contain local information which may interfere with such processing is an interesting addition to the literature. Most past studies which have found similar relationships between AGL or statistical learning and language processing have relied on offline measures to do so (Misyak & Christiansen, 2010, 2012) as was the case in the present study, in spite of good reasons to have hypothesized that the new online measure put forward in this study might provide a stronger correlate.

This should not take away from the fact that performance on the bi-/tri-gram familiarity judgments correlated with sentence processing. Participants who were most sensitive to interference from an adjacent mismatched noun when processing the main verb were better at learning the adjacent dependencies in the AGL paradigm. This suggests that there is potentially a shared cognitive mechanism that underlies the processing of adjacent sequential input in both the AGL paradigm and the sentence processing task, as evidenced by the shared individual differences in language processing and SL abilities. Such a result mirrors the findings of Misyak, Christiansen, and Tomblin (2010), who found a similar trend between a variation of the AGL-SRT that featured nonadjacent dependencies, and sentences containing relative clause constructions with long-distance dependencies, and that of Misyak and Christiansen's (2010) findings for an artificial grammar featuring adjacent dependencies.

When controlling for working memory abilities, the relationship became marginally significant, suggesting that performance on the familiarity judgment task within Standard AGL may rely in part on more general memory processes. This possibility was part of the motivation for designing the online, active measures embedded within the AGL-SRT. Moving forward, it would be wise to consider using a between-subjects design to verify reliability while also examining potential language processing correlates, in order to reduce the burden on participants. Another recent study is also noteworthy, as it has brought into question the reliability of self-paced reading time paradigms commonly used in sentence processing tasks, such as the one in the present study (James, Fraundorf, Lee, & Watson, 2018). This finding suggests that future research in this area might need to consider the increased variability inherent to online measures of both language processing and SL in their design, and instead utilize some combination of online and offline measures to assess the relationship between domain-general cognitive abilities and individual differences in language processing.

Importantly, this relationship between AGL and language processing is not due merely to individual differences in speed of processing or general aptitude on RT based tasks due to the fact that difference scores are used for both measures. This demonstrates an important link between the learning abilities involved in the AGL paradigm, and the processing abilities tested in the language processing task, buttressing the claim that language skills are built upon a bedrock of domain-general cognitive abilities (Christiansen & Chater, 2008). It also highlights the tight link between language learning and processing and is perhaps additional evidence that learning and processing are in fact two sides of the same coin (Christiansen & Chater, 2016). The ability to utilize relational information similarly within both the sentence processing task and the AGL paradigm can perhaps be best described if we consider learning and processing as such, due to the fact these biases are shared.

Conclusion

This study is the first to measure the reliability of both traditional and new AGL paradigms, in an attempt to parallel the progress made in this regard within the SL literature. We demonstrated reliability for key measures within each task, while also calling into question that of other frequently used procedures. Moreover, this study further underscores the link between individual differences in domain-general learning and language processing abilities.

References

Arciuli, J., Torkildsen, J. V. K., Stevens, D. J., & Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*, *5*(July), 747. https://doi.org/10.3389/fpsyg.2014.00747

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: Past, present and future. *Philosophical Transactions of the Royal Society B*, *372*(1711), 20160047. https://doi.org/10.1098/rstb.2016.0047

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How Implicit Is Visual Statistical Learning ? *Journal of Experimental Psychology. Learning, Memory, and Cognition*. https://doi.org/10.1037/a0027210

Christiansen, M. H. (in press). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12332

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The Behavioral and Brain Sciences*, *31*(5), 489-508; discussion 509–58. https://doi.org/10.1017/S0140525X08004998

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *The Behavioral and Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525X1500031X

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition*, *114*(3), 356–71. https://doi.org/10.1016/j.cognition.2009.10.009

Farmer, T.A., Fine, A.B., Misyak, J.B. & Christiansen, M.H. (2017). The inter-relationship between reading span task performance, linguistic experience, and the processing of unexpected syntactic events. *Quarterly Journal of Experimental Psychology, 70,* 413-433.

Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical Learning of Two Artificial Languages Presented Successively: How Conscious? *Frontiers in Psychology*, *2*(September), 1–12. https://doi.org/10.3389/fpsyg.2011.00229

Franco, A., & Destrebecqz, A. (2012). Chunking or not chunking? How do we find words in artificial language learning? *Advances in Cognitive Psychology*, *8*(2), 144–54. https://doi.org/10.2478/v10053-008-0111-3

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. https://doi.org/10.1016/j.tics.2014.12.010

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2017). Testing Statistical Learning Implicitly : A Novel Chunk-based Measure of Statistical Learning. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, (May), 564–569.

James, A. N., Fraundorf, S. H., Lee, E., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, *102*(August 2017), 155–181. https://doi.org/10.1016/j.jml.2018.05.006

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 3,* 228-238.

Kaufman, S. B., Deyoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–40. https://doi.org/10.1016/j.cognition.2010.05.011

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35–54. https://doi.org/10.1037//0033-295X.109.1.35

Misyak, J. B., & Christiansen, M. H. (2010). When 'More' in Statistical Learning Means 'Less' in Language : Individual Differences in Predictive Processing of Adjacent Dependencies Paradigm.

Misyak, J. B., & Christiansen, M. H. (2012). Language Learning Statistical Learning and Language : An Individual Differences Study. *Language Learning*.

Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*, *2*(1), 138–153. https://doi.org/10.1111/j.1756-8765.2009.01072.x

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and Implicit Second Language Training Differentially Affect the Achievement of Native-like Brain Activation Patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933–947. https://doi.org/10.1162/jocn

Pearlmutter, N.J., Garnsey, S.M. & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427-456.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–8. https://doi.org/10.1016/j.tics.2006.03.006

Reber, A. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19489896

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning : Current pitfalls and possible solutions. *Behavior Research Methods*, *49*, 418–432. https://doi.org/10.3758/s13428-016-0719-z

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability : Theoretical perspectives and empirical evidence. *Journal of Memory and Langauge*, *81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Waters, G. S., & Caplan, D. (1996). Processing resource capacity and the comprehension of garden path sentences. *Memory & Cognition*, *24*(3), 342–355. https://doi.org/10.3758/BF03213298

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2017). The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental Science*, e12552, 1–13. https://doi.org/10.1111/desc.12552

CHAPTER THREE

Input complexity affects long-term retention of statistically learned regularities[2]

Abstract

Statistical learning involving sensitivity to distributional regularities in the environment has been suggested to be an important factor in many aspects of cognition, including language. However, the degree to which statistically-learned information is retained over time is not well understood. To establish whether or not learners are able to preserve such regularities over time, we examined performance on an artificial language learning task both immediately after training but also at a follow-up session two weeks later. Participants were exposed to an artificial language (Brocanto2), half of them received simplified training items in which only 20% of sequences contained complex structures, whereas the other half were exposed to a training set in which 80% of the items were composed of complex sequences. Overall, participants showed signs of learning at the first session and retention at the second, but the degree of learning was affected by the nature of the training they received. Participants exposed to the simplified input outperformed those in the more complex training condition. Additional analyses were conducted to determine the kind of information that participants employed in the two training conditions. The results indicate that participants in the complex training condition relied more on an item's

[2] Co-authored with Kate Brill-Schuetz, Kara Morgan-Short, and Morten H. Christiansen; currently in preparation for submission.

chunk strength than those in the simple training condition. Taken together, this set of findings show that statistically learned regularities are retained over the course of two weeks. The results also demonstrate that training on input featuring simple items leads to improved learning and retention of statistical regularities.

Introduction

Statistical learning (SL) has been identified as a domain-general cognitive ability that is integral to language processing, acquisition, and evolution (see Armstrong, Frost, & Christiansen, 2017, for an overview). For the purposes of this study, we can define SL as the process by which learners uncover the structure of the input from its distributional properties (Frost, Armstrong, Siegelman, & Christiansen, 2015). However, little is known about the extent to which statistically learned information is retained over time (see Gomez, 2017, for a review).

Initial studies of SL focused on the rapidity with which human infants could learn from predictable, structured sequences of input (e.g., Saffran, Aslin, & Newport, 1996). As a result, the literature has remained quite focused on measuring the ability of participants to learn these regularities within a single session, usually with a test-phase following some sort of training. There have been only a handful of studies examining the ability of adult participants to retain statistically learned information over long periods of time.

Two studies in particular stand out as examples of investigations into the long-term retention and consolidation of sequence learning abilities. Romano, Howard, and Howard (2010) demonstrated that participants seemed to retain sequence-specific learning and general skill effects a year after training on a serial reaction time task in which they had to press a key corresponding to a circle's location as they appeared on the computer screen. This retention seemed to be non-declarative and was found for each of their various training groups including younger and older adults, and experienced musicians or video game players. In effect, participants recalled frequent triplets more quickly than they did low-probability control trials, showing a learning effect over the course of the tasks at session one which persisted at session two, one year later.

Kobor, Janacsek, Takacs, and Nemeth (2017) attempted to extend these findings in a task designed to test consolidation along with retention, as they were interested in uncovering the core mechanism(s) that underlie long-term memory formation in such a task. This study investigated the role of retroactive interference in forgetting by training participants on a new set of items with an alternate statistical structure one day after the initial test session. The second test session in this study, which tested long-term retention of the initially trained patterns, took place a full year later, similar to the Romano et al. (2010) study. Again, the researchers found learning effects for highly frequent items relative to infrequent items, and also found no effect for the potentially interfering materials, and an additional test demonstrated that the knowledge gained in this task seemed to be implicit in nature. They took this to mean that long-term memory for statistically learned sequences does undergo a process of consolidation that appears to be robust and resistant to some kinds of interference. Moreover, learning scores seemed relatively stable between the first session, the interference training session the next day, and the final session a year later.

While these studies demonstrate the persistence of sequence learning abilities over a long stretch of time, they are still limited in some important ways. First, the tasks in both studies required only visuospatial to motor mappings, without any auditory or verbal component.

Second, and relatedly, the learned sequences did not contain any kind of meaning. While this is a common practice within the SL literature, it limits the study's ecological validity when it comes to addressing the mechanisms thought to underlie language learning. Third, the statistical structure underlying the training items was not very complex in either study, again somewhat undermining the claim that the kinds of relationships learned between items in a sequence are characteristic of those in natural language. Finally, neither of these studies demonstrated a quintessential feature of learning in SL and artificial grammar learning (AGL), the generalization of learned regularities to new items. The test sets in each task contained exclusively items on which participants had already been trained.

Within the SL literature, other studies have attempted to look at retention over relatively short gaps in time between training and testing. For example, Kim, Seitz, Feenstra, and Shams (2009) demonstrated that participants implicitly learned the statistical relationships governing a sequence of rapidly presented visual stimuli, and that this learning was retained over the course of 24-hours. Other work has shown that adults possess the ability to maintain information about the underlying relationship between visual stimuli in an SL task, with testing periods at 30-minute, 1-hour, 2-hour, 4-hour, and 24-hour delays (Arciuli & Simpson, 2012). Participants in each test group showed no difference in their ability to correctly identify grammatical sequences, suggesting that at least over the course of a day, the information gleaned within a SL task is relatively robustly retained. The authors of this study notably suggest that their findings do not indicate any sort of enhancement in retention for participants who slept between training and test.

A wide-range of additional research, however, has attempted to examine the role of sleep in the consolidation of associations learned within a SL paradigm. While the present study does not seek to examine the effects of sleep on SL, this is perhaps the most well-studied aspect of long-term retention within the literature. Napping appears to improve consolidation, as participants who slept during a 4-hour delay period between training and test outperformed those who did not when asked to choose between novel tone sequences, one of which followed the rules that were used to generate those encountered in training (Durrant, Taylor, Cairney, & Lewis, 2011). Interestingly, this enhancement was positively correlated with the amount of slow-wave sleep obtained by the participant. Similarly, researchers have shown that participants who slept were more likely to apply statistically learned constraints in a speech production task (Gaskell et al., 2014). This study also demonstrated a positive relationship between slow-wave sleep and learning effects. In general, it seems that over a relatively brief period of time, knowledge gained in a SL task can be retained, and that this retention may even be enhanced by sleep in some instances.

More recently, another sleep study by Frost and Monaghan (2017) demonstrated that participants who underwent a period of sleep between training and test within a non-adjacency SL paradigm outperformed those who stayed awake at both word learning and also in generalizing the rules of the grammar to new sequences that had not been see during training. In addition to showing that participants retain learned associations after several hours this study also provided evidence that these two processes may not be entirely separable, but rather rely on the same underlying mechanism, an idea which has been suggested by others as well (e.g. Frost & Monaghan, 2016; van den Bos, Christiansen, & Misyak, 2012). Given the frequently described links between SL and language, it would seem likely that the associations learned in commonly used SL paradigms should persist over longer periods of time than we currently have evidence for, an issue the current study seeks to address with the hypothesis that participants will retain learned information over the course of two weeks. If statistically learned information truly undergirds our language learning abilities, it must be retained beyond an immediate post-test in the lab.

The idea that while learners process the co-occurrence statistics of the input they are also acquiring the more abstract regularities of an underlying grammar is nothing new (Elman, 1990; Altmann, 2002). Meanwhile, other work has suggested that “less is more,” that is, beginning to learn without fully developed cognitive abilities could convey an advantage to children in terms of learning such grammatical regularities by forcing them to first learn the most basic information available from the input, and has been part of the literature for quite some time (Newport, 1990). This notion has been applied to the process of language learning and has been pointed out as a potential reason for the existence of sensitive periods in language acquisition (Johnson & Newport, 1989; Newport, 1990, 2016). The corresponding idea that “starting small” may be advantageous for learners shares similar longevity within the literature (Elman, 1993; Elman, et al., 1996), and emphasizes the possible benefit that reduced complexity within the learner’s input (e.g., in terms of length or syntactic complexity) has on learning.

While the evidence for these hypotheses has subsequently become somewhat less straightforward (for example, see Rohde & Plaut, 1999; Siegelman & Arnon, 2015), evidence is still emerging that, within the context of artificial language learning, participant performance may benefit from training that becomes progressively more challenging (Lai & Poletiek, 2011; Kersten & Earles, 2001). More recently, a study has shown that starting small leads to better learning of recursive structures, with the primary facilitation coming from a gradual increase in stimuli complexity rather than simply the effect of reduced length (Poletiek et al., in press). Other work has also shown that artificially biasing the kinds of chunks that adults form to be more child-like can lead to improved learning in a Hebb-repetition paradigm (Smalle, et al., 2016). Taken in conjunction with other recent ideas, chunking can be seen as an integral component of the statistical learning process as it applies to language (Isbilen & Christiansen, in press). Rapidly recoding and compressing information by chunking may allow learners to more efficiently process input, and to do so at higher levels of abstraction. In fact, stronger learners may show a decreased reliance on surface-level fragment information when tested due to the fact that they have already used that information to internalize the higher-order regularities, and no longer use them as a crutch.

The literature on second language acquisition also offers some insight into how learners acquire grammatical knowledge. Antoniou, Ettlinger, and Wong (2016) have shown that different kinds of training leads participants to learn different aspects of an artificial language's grammar. In their study, participants' individual differences in procedural memory positively corelated with the acquisition of simple rules, while their declarative memory abilities were positively correlated with the learning of complex rules. In addition, they also showed that training participants on simple rules prior to complex rules proved advantageous to learning.

This literature also has provided evidence of retention over long periods of time, albeit usually with rather extensive training. For example, researchers have previously described behavioral and neurophysiological evidence of retention over the course of three to six months for the same artificially learned language used in this study, Brocanto2, after participants were extensively trained to a high level of proficiency (Morgan-Short, Finger, Grey, Ullman, 2012). Using an artificial, as opposed to natural, language controls for prior experience with the second language and it is also more conducive to manipulating the qualities of input without changing any other linguistic properties. Moreover, the present study used an artificial language in order to maximize learning in a short time span while controlling for possible confounding variables, i.e., phonology and native-language transfer, as well as avoiding problems with experimental techniques while still eliciting learning processes typical of natural languages (Morgan-Short, 2007).

The delayed post-test is a hallmark of the second language acquisition literature as shown in a thorough meta-analysis (Norris & Orteaga, 2000), as nearly half of all studies in this field feature some kind of follow-up test phase more than a week after training. This research has shown that artificial languages can be learned and retained over time, although determining whether or not retention exists for more incidentally learned, less well-trained material would benefit the literature – over 80% of the reviewed studies had more than one hour of training. Much of the reviewed literature in their meta-analysis focused on determining optimal training techniques, rather than studying the ways that the input itself shapes the kinds of information to which learners have access.

To this end, the present study seeks to examine the different ways in which learners retain knowledge about an artificial grammar. Firstly, we predict that learners will retain knowledge from training over the span of two weeks. However, we also predict that both learning and retention will depend on the complexity of the items on which participants are trained. In order to mimic the constraints placed on young learners by the simplified input they tend to receive (Cameron-Faulkner, Lieven, & Tomasello, 2003), along with the more low-level features they are thought to be sensitive to due to processing and memory limitations, half of the participants in this study received a more simplified set of training items generated by the Brocanto2 grammar (B2; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012). Those in the simple training condition were eventually exposed to complex items, but the extensive experience they received with simple items before moving on to the more complex ones is expected to boost performance in the test phase of the experiment (Brill-Schuetz, 2016). Training sets with progressively increasing difficulty have been used in past AGL and SL studies for similar reasons (e.g., Christiansen, Conway, & Onnis, 2012; Conway, Ellefson, & Christiansen, 2003; Poletiek et al., in press).

The other half of participants received far less training with simplified items prior to exposure to the set of complex items yet obtained the same amount of total experience in terms of number of trials. These participants are thus predicted to have more trouble learning the underlying grammar, as they would have insufficient experience processing simple constructions before encountering the more difficult complex items. This may lead them to adopt poor learning strategies, disrupting their extraction of the relevant statistical structure embedded within the sequence. However, better working memory skills may ameliorate performance for participants in the complex training condition, whereas it would be less likely to do so for participants who underwent simple training.

We are also interested in finding out how the different training groups approach the task of endorsing items as grammatical, by looking into what features of the test items are most relevant to such judgments. Examining endorsement strategies is expected to provide insight into what each group of participants retained from the task across both sessions. The specific cues that participants rely on to make grammaticality judgments might vary between the training groups, and if participants in the complex training condition show the reduced sensitivity to the grammatical regularities of the language that we predict, they might be instead be found to rely more on fragment information. On the other hand, the simple training group will likely not be as distracted by surface-level similarities between training and test items and will rather demonstrate their superior knowledge of the grammatical regularities.

Method

Participants

Participants (*N* = 47; Male = 10) were young adult students at a large, Midwestern university, ranging in age from 18 to 24 (*M* = 19.43, *SD* = 1.98). Recruitment for the first session was conducted through a psychology department subject pool where participation earned class credit. For the second session, some participants received additional credit through the subject pool and others received monetary compensation ($5). Selection criteria limited participants to those who had no hearing, learning, or speaking impairments, and to native speakers of English.

The second session took place approximately two weeks after the original training session. Although every effort was made to schedule the delayed post-test exactly two weeks from the original session, the actual range was between 12 and 14 days from the training session. At this second session, some (n = 33) participants also completed an additional battery of cognitive tests.

Materials

Artificial Language

The artificial language learned by participants was Brocanto2 (B2; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Morgan-Short et al., 2012), which was adapted from the original version, Brocanto (Friederici, Steinhauer, & Pfeifer, 2002).

B2 follows basic patterns typical of many natural languages and is fully productive; it consists of 14 novel words: four nouns, two adjectives, two articles, four verbs, and two adverbs (see **Table 4** for a list of all words and their meanings). The grammatical structure of this language follows a syntactic pattern different from that of English; while English follows a subject-verb-object order, B2 follows a subject-object-verb order, which is found in languages such as Hindi and Japanese. For example, the B2 sentence "*Blom neimo lu neep troise li praz zayma"* corresponds to "Blom-piece square the neep-piece round the switch horizontally" and would be translated into English as "The square blom-piece switches with the round neep-piece horizontally." Participants learned this artificial language in order to play a computer-based game in which the tokens can move according to dictation in B2 (see **Figure 10**).

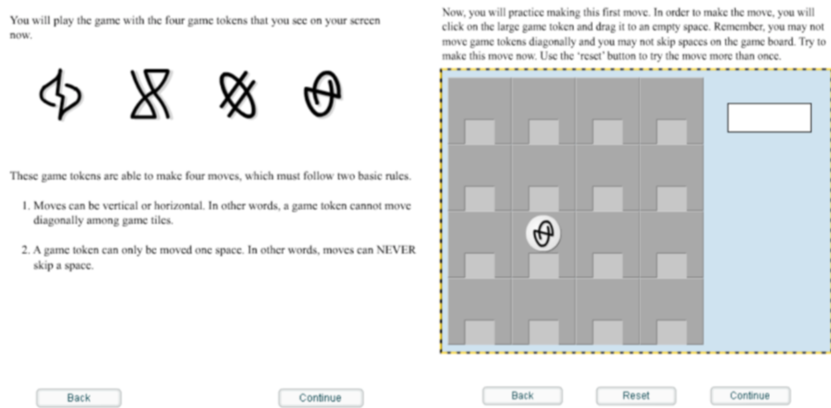


Figure 10. Screenshot of training on the B2 paradigm. Taken from Brill-Schuetz (2016).

Items used within the Brocanto2 artificial language

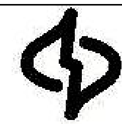
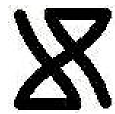
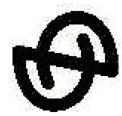
Word Category	Brocanto2 word	Symbol/meaning
Noun	pleck$_{m}$	
	neep$_{m}$	
	blom$_{f}$	
	vode$_{f}$	
Adjectives	trois(e$_{m}$/o$_{f}$)	circle
	neim(e$_{m}$/o$_{f}$)	square
Determiners	li$_{m}$/lu$_{f}$	the
Verbs	klin$_{intran}$	move
	praz$_{tran}$	switch
	nim$_{tran/intran}$	capture
	yab $_{tran/intran}$	release
Adverbs	noyka	vertically
	zayma	horizontally

Table 4. Complete list of words used within the artificial language learning task. Subscripts denote the gender of each noun and determiner along the corresponding marking for each adjective, and also the transitive nature of each verb. The adjectives described the shape of the area bordering the game piece, such as the circle that can be seen in **Figure 10**. Table adapted from Morgan-Short (2007).

These sentences could be either simple or complex in nature; simple stimuli were limited to words from three of the word categories (noun, article, verb) and could consist of three to five lexical items. Complex stimuli consisted of words from all five of the categories allowed in B2 (noun, adjective, article, verb, adverb) and a complex sentence could contain five to eight lexical items. For example, the sample sentence given above would be classified as a complex item due to the inclusion of the adjectives and the adverb. The presentation of each sentence was consistent in that all the noun phrases were simple or complex and all verb phrases were either simple or complex; for example, a sentence would not have a simple noun phrase followed by a complex verb phrase. See **Table 5** for examples of both complex and simple sentences. During training (but not test), some of the simple and complex stimuli included noun phrases presented without a corresponding verb or adverb. The simple phrases had only a noun and a determiner, while the complex phrases included noun, adjective, and determiner. **Figure 11** illustrates all possible word class combinations, and identifies the two kinds of phrases and four kinds of sentences that could be generated by the B2 grammar.

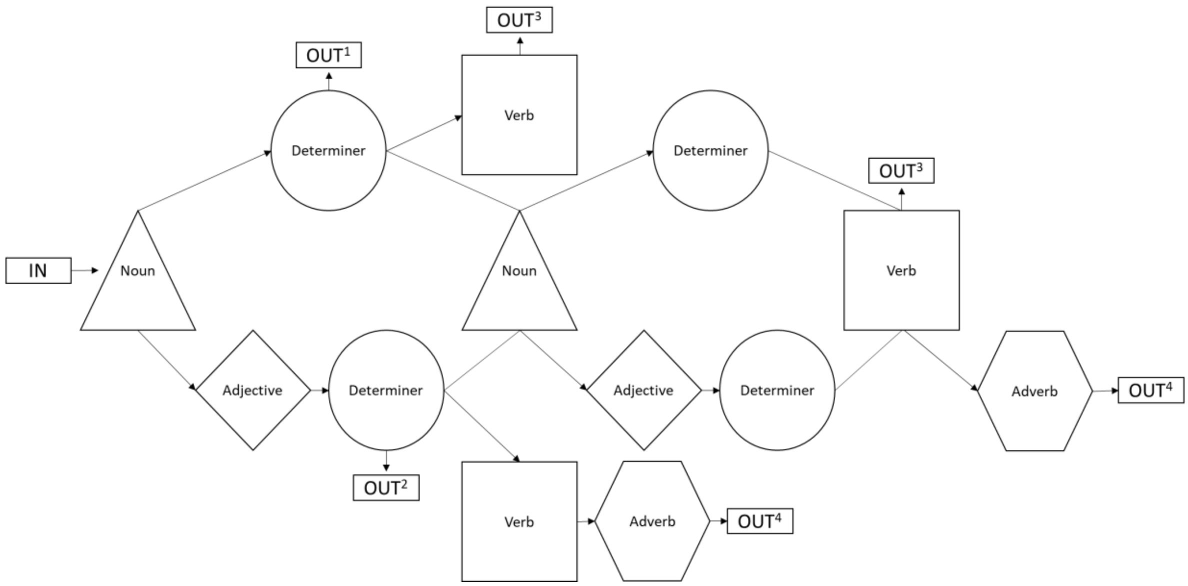


Figure 11. Chart depicting the possible word class combinations of items generated by the B2 grammar. The superscript at each output classifies the category of phrase or sentence that such a sequence produces; [1] denotes a simple phrase (noun + determiner), [2] denotes a complex phrase (noun + adjective + determiner), [3] denotes a simple sentence (noun + determiner + verb; noun + determiner + noun + determiner + verb), and [4] denotes a complex sentence (noun + adjective + determiner + verb + adverb; noun + adjective + determiner + noun + adjective + determiner + verb + adverb.

Examples of simple and complex input for klin *and* praz *in Brocanto2*

Simple input	Brocanto2 sentence	Word categories
Klin^	*Blom lu klin*	*N Det + V*
Praz+	*Blom lu neep li praz*	*N Det + N Det + V*
Complex input		
Klin^	*Blom neimo lu klin noyka*	*N Adj Det + V Adv*
Praz+	*Blom neimo lu neep troise li praz noyka*	*N Adj Det + N Adj Det + V Adv*

Table 5. Example sentences from both complexity conditions containing the two verbs that could not be both transitive and intransitive. Noun = N; determiner = Det; verb = V; adjective = Adj; adverb = Adv; ^ denotes intransitive verb and + denotes transitive verb. Table adapted from Brill-Schuetz (2016).

Procedure

Brocanto2 Artificial Language Learning Paradigm

The procedure for this study is that reported in Brill-Schuetz (2016); the data used in this study was originally collected for use in Brill-Schuetz (2016). Before training began, participants were taught the B2 vocabulary prior to starting any other aspects of the study. Participants were then presented with game training, which consisted of an introduction to the computerized board game they would be playing at a later point, thus providing a meaningful context for the artificial language on which they were subsequently trained. Participants read the rules of the game and viewed the four possible types of game moves (move, switch, capture, or release). They were then asked to practice making each move on the game board by selecting game tokens with a mouse and repeating the move that had just been visually presented, as can be seen in **Figure 12**. At no point were explicit translations of the symbols or movements provided. Once a perfect score was attained in terms of memorizing the vocabulary, participants continued on to language training. Note that all participants received the same vocabulary and game training — it was not part of the manipulation.

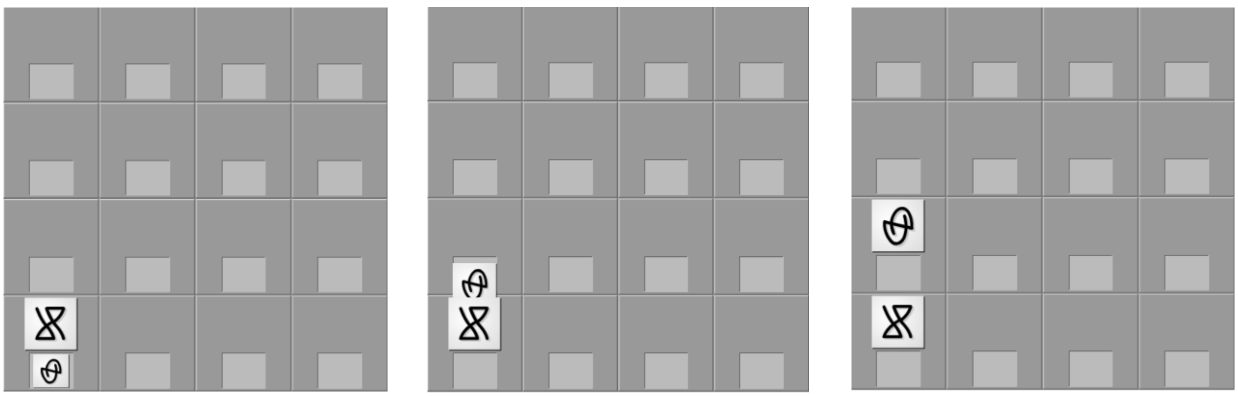


Figure 12. Example of progressive screen shots for an animated movement. The corresponding audio was *neep li vode lu yab* for simple or *neep neime li vode neimo lu yab noyka* for complex. Figure taken from Brill-Schuetz (2016).

Before training on the language the participants were instructed that they would receive training on an artificial language and would be presented with words, phrases, and sentences that would correspond to still and moving images on the game board. Participants were told they would complete a short quiz to test their memory and they would not be able to review this information again. They were also informed that they would then use the artificial language to play a board game at a later point. No other instructions were given; therefore, training was classified as implicit or uninstructed (not incidental) due to the lack of explicit information or explanation of the B2 language rules.

Participants were pseudorandomly assigned to either simple or complex input conditions, with every other learner assigned to the simple condition. All participants received training phrases and sentences featuring identical nouns and verbs, presented either in a simple or complex format (100 items). Thirty-six of the training items were phrases, while sixty-four were sentences.

As reported by Brill-Schuetz (2016), in the "simple" training condition, 80% of the sentences that participants received were simple while the other 20% was complex; in the "complex" training condition, 80% of the sentences were complex while 20% were simple. This particular ratio of stimuli was utilized so that participants would be exposed to every word category in B2 and its function in a sentence while still presenting a vast majority of one particular type of stimuli. Furthermore, a 1:4 ratio has also been used in previous cognitive linguistics literature that has examined the learning and generalization of grammar rules for novel verbs (e.g., Casenhiser & Goldberg, 2005). If not provided some examples of complex stimuli, participants in the simple training condition would not have been exposed to the function of an adjective or adverb in a sentence creating a potential confound that may have manifested in the language assessment (Brill-Schuetz, 2016). Participants received repeated aural examples of B2 and always received simple stimuli before complex stimuli regardless of the training condition. That is, all participants received the training items in the following order: simple phrases, complex phrases, simple sentences, complex sentences. This ordering of phrases being presented before full sentences follows the structure of previous B2 studies (e.g., Morgan-Short et al., 2014; Morgan-Short et al., 2010) and that of studies exploring the starting small hypothesis (e.g., Conway, Ellefson, & Christiansen, 2003; Kersten & Earles, 2001; Poletiek et al., in press). This means that, for example, a participant in the simple training condition would see twenty-nine simple phrases followed by seven complex phrases, and then fifty-one simple sentences followed by thirteen complex sentences. In the complex condition, the order would remain the same, but participants would instead be trained on seven simple phrases followed by twenty-nine complex phrases, and then thirteen simple sentences followed by fifty-one complex sentences.

Each training condition began with the visual presentation of the 36 individual symbols that corresponded to B2 noun phrases (simple and complex) and progressed to 64 full animated moves with corresponding sentences (simple and complex). Presentations for each noun phrase consisted of a single, static game piece while the audio was played. An animated movement involving one or more pieces on the game board accompanied the presentation of sentences and in this case, the audio was played before the animated movement occurred. At the conclusion of each noun phrase or animation, there was a one-second break before the next item appeared on screen. The game pieces and animations presented to participants were identical across the two conditions—the training only varied based on the audio.

The primary language assessment in this study consisted of a grammaticality judgment task (GJT). The GJT requires the participant to make a judgment regarding the grammaticality (yes or no) of a sentence and are commonly used across second-langugage learning literature (cf. Loewen, 2009). The GJT consisted of 72 novel sentences, half (36) of the stimuli were simple sentences and half were complex. Of the 36 simple sentences, half were correct and half contained a violation; this was also the case for the complex sentences.

Grammatical sentences for the GJT were novel, i.e., correct sentences that were not presented during training. In general, ungrammatical were generated by introducing violations in the novel, correct sentences. However, four ungrammatical simple sentences had to be created using violations of sentences that appeared in training due to the limited number of such sentences that could be generated by the grammar. Word order violations were created by replacing a word from one of the five word categories (e.g., noun) with a word from a different category (e.g., adjective, article, verb, adverb). Verb argument violations were created by replacing a transitive verb with an intransitive verb and vice versa, therefore these violations were constrained to the verbs *klin* and *praz*. Grammatical gender violations were created by replacing a feminine adjective or article with a masculine adjective or article, and vice versa. Violations never occurred on the first or final word, and violation position among words was distributed as evenly as possible. Word frequency within each grammatical category was also as equally distributed as possible across all sentences. Examples of each type of violation sentence can be found in **Table 6**.

Example correct Brocanto2 sentences and violations thereof

Sentence type	Brocanto2 sentence
Correct sentence	*Blom neimo lu neep neime li praz* Blom-piece square the neep-piece *square* the switch “The square blom-piece switches with the square neep-piece.”
Word Order (Syntactic) Violation sentence	*Blom *nim lu neep neime li praz* Blom-piece *capture the neep-piece square the switch “The *capture blom-piece switches with the square neep-piece.”
Verb Argument Violation sentence	*Blom neimo lu *praz* Blom-piece square the *switches “The square blom switches *(missing object)”
Morphosyntactic (Gender Agreement) Violation sentence	*Blom neimo lu neep *neimo li praz* Blom square the neep-piece $_{(fem)}$ *square$_{(masc)}$ the switches The square blom switches with the square$_{(m)}$ neep$_{(f)}$

Table 6. Adapted from Brill-Schuetz (2016). *Denotes the location of the violation.

The GJT was programmed in SuperLab 5 and the stimuli (the B2 sentences) were randomized. The GJT began by guiding participants through the instructions; all directions were presented in white font (size 30) on a black background. The initial screen informed participants that the task was to make a series of judgments regarding new sentences in the artificial language, and that they should make each judgment as quickly and accurately as possible. Participants also gave confidence ratings as part of the test. Although that data does not appear in the present study, it is worth mentioning as a potential task demand that could have influenced other results.

Working Memory Paradigms

Two working memory tasks were used to determine participants' working memory abilities. In the Symmetry Span Task (SymmSpan; Redick et al., 2012; Foster et al., 2014), participants were presented with a 4x4 grid, in which some of the cells contained a red box. After this array appeared and then left the screen, an intervening distractor task popped up prompting participants to judge whether or not a geometric figure was symmetrical along its vertical axis. After responding to the prompt, another red box would appear within the 4x4 grid, followed by another symmetrical judgment distractor item. At the end of this sequence of targets and distractors, participants were then presented with a new screen depicting an array of the previously seen 4x4 grid, except without any of the red boxes present. They were instructed to recall the locations of the previously encountered boxes in this 4x4 grid by clicking on the location(s) in which the red boxes appeared, in order of presentation. Trial lengths varied between two to five sets of symmetry-location pairs and one item of each length was presented within each of the three blocks. Scores were calculated by summing the number of correctly recalled box locations.

The letter-number ordering task (LNOT) is part of the WAIS-III Intelligence Scale (Wechsler, 1997). While it can be used to assess one's ability to process sequential information, the dual-task nature overlaps with WM tasks. This task was chosen specifically because of its linguistic component. The version used in this experiment was translated from van den Noort et al. (2006). In this WM test, the researcher read aloud a series of letters and numbers, starting at two letters and numbers per series and progressing to eight letters and numbers, meaning that participants would need to recall between four and sixteen items per series. The participant was then instructed to repeat the letters and numbers that were spoken but in numerical and alphabetical ordering, which conflicted with how they were presented. For example, if the researcher read *W, 1, K, 5,* the correct answer would have been *1, 5, K, W*. Participants received a point for each series correctly repeated and the task ends when the participant missed three series in a row. For the analyses reported below, a composite WM score was computed for each participant. This composite score was created by standardizing (using z-scores) the scores in each task and then averaging the two.

Results

General Participant Performance

Participants in the two training conditions showed different learning outcomes on the GJT across sessions. As found in Brill-Schuetz (2016), overall participant accuracy was above chance (i.e., 50%) when judging items as grammatical or ungrammatical at both sessions one ($t(46) = 4.774$, $p < .001$; mean: 56.9% correct) and two ($t(46) = 2.452$, $p = .018$; mean: 53.6% correct). This demonstrates that participants retained knowledge of the grammatical regularities over the course of two weeks.

Looking deeper to see what aspects of training affected accuracy and retention, a 2 (session) x 2 (training condition) mixed ANOVA analyzing accuracy showed significant main effects for both session ($F(1,45) = 9.058$, $p = .004$) and training group ($F(1, 45) = 6.872$, $p =$

.012), while the interaction effect did not reach significance ($F(1,45) = .796$, $p = .377$). As shown in **Table 7**, those in the simple training condition demonstrated above chance performance at both the first ($t(23) = 4.018$, $p = .001$) and second ($t(23) = 3.835$, $p = .001$) sessions. Those in the complex condition showed above chance accuracy at session one ($t(22) = 2.907$, $p = .008$), but not at session two ($t(22) = -0.172$, $p = .865$). While this set of results would seem to indicate that those in the complex training condition did not exhibit learning as well as those in the simple training condition, it is also possible that they were sensitive to other aspects of the items besides their grammaticality. That is, it is possible that they learned some features of the training set besides the grammar and used those as cues when accepting or rejecting items.

Participant performance by training condition

	Simple Training		Complex Training	
	Session 1	Session 2	Session 1	Session 2
Correct (SD)	59.9% (.12)	57.5% (.10)	54.2% (.07)	49.7% (.09)
Endorsed (SD)	55.1% (.11)	57.0% (.14)	51.8% (.12)	58.4% (.10)

Table 7. Percent correct and endorsement rates on the GJT for participants in each training condition at both sessions, along with standard deviations in parentheses.

An additional analysis of performance focused on participants for whom the working memory task had been administered ($n = 33$; 15 within the simple training group, 18 for complex training group). Overall, only participants within the complex training condition showed a correlation between accuracy on the GJT and working memory ability. They showed an overall effect across both sessions ($r = .611$, $p = .007$), and also a significant relationship at both sessions one ($r = .536$, $p = .022$) and two ($r = .575$, $p = .012$) separately.

Modeling Predictors of Item Endorsement

In order to get a clearer picture of the type(s) of information to which participants in either group showed sensitivity, endorsement rates were calculated by looking at the proportion of 'yes' responses when participants were asked if they thought the test sequence was grammatical. Endorsement rates for each group at both sessions can be found in **Table 7**.

To determine the strongest predictors of item endorsement, we used a series of generalized linear mixed effect models (GLMMs) to examine the effects of training condition, chunk strength, and time (session) on item endorsement using the LME4 package in R (Bates et al., 2014). The model included as fixed effects: training group, chunk strength of GJT item, and time. We included as a random effect the intercepts for GJT endorsement by subject. This controlled for individual differences in GJT endorsement, making it easier to detect fixed effects of our variables of interest.

The chunk strength of each item was calculated in order to determine the extent to which the participants used this kind of fragment information when endorsing items. The chunk strength referenced here was measured as the sum of the frequency of occurrence in the training items of each of the fragments in a test item, weighted by the number of fragments in that item (Knowlton & Squire, 1994). For example, the associative chunk strength of the item ZVX would be calculated as the sum of the frequencies of the fragments ZV, VX and ZVX divided by 3. A higher number indicates that a test item is well supported by chunk information in the training items. With the sets of training and test items used in this study, chunk strength actually was significantly greater for grammatical versus ungrammatical test items, meaning that it was a potentially useful cue for performing accurately on this task for both the simple ($t(70) = 2.267$, $p = 0.026$) and complex ($t(70) = 2.396$, $p = 0.019$) training groups. Note that these comparisons were computed separately given that the two groups had different training sets, even though the test sets were exactly the same.

The initial model (Model 1) with separate fixed effects is reported in **Table 8**. However, due to the nature of the manipulation and the variables of interest, another model with three two-way interaction terms was built. This model (Model 2) was built based on the hypothesis that the training condition (Group) would interact with both the session (Time) and item chunk strength (Chunk). Additionally, we hypothesized that the effect of chunk strength on item endorsement may potentially degrade with time due to the nature of memory, thus we included an interaction term between these variables. The results for Model 2, which include these interaction terms, are also reported in **Table 8**. To test if the inclusion of interaction terms improved upon Model 1, a deviance test was conducted (Singer and Willett, 2003). The interaction terms improved model fit, $\chi^2(3) = 76.681$, $p < .0001$.

A further desire to also include a potential three-way interaction between training condition, session, and chunk strength led to the creation of Model 3. This model outperformed Model 2 ($\chi^2(1) = 13.716$, $p = .0002$), supporting the hypothesis that the effect of training on retention would differ between groups. **Figure 13** depicts this interaction nicely, showing that the effect of chunk strength on item endorsement decreases over time and illustrating the greater impact of chunk strength on endorsement for participants in the complex training condition.

Overview of GLMM statistics

Fixed Effects	Model 1	Model 2	Model 3
Intercept	-0.847***	-1.961***	-2.527***
	(.134)	(.217)	(.268)
Group	0.095	1.093***	2.100***
	(.137)	(.229)	(.357)
Chunk	0.126***	0.281***	0.374***
	(.007)	(.025)	(.036)
Time	0.180***	0.682***	1.05***
	(.052)	(.118)	(.155)
Group*Chunk		-0.110***	-0.279***
		(.015)	(.048)
Group*Time		-0.227*	-0.887***
		(.105)	(.207)
Chunk*Time		-0.064***	-0.125***
		(.015)	(.022)
Group*Chunk*Time			0.111***
			(.030)
Random effects			
Subject (Intercept)	0.189	0.192	0.193
	(.435)	(.438)	(.439)
Goodness of Fit			
Log likelihood	-4231.2	-4192.9	-4186.0
AIC	8472.5	8401.8	8390.1
BIC	8506.4	8456.0	8451.1

Table 8. Summaries of the two generalized linear mixed effects models. Estimated coefficients are listed while standard errors are reported in parentheses. * indicates $p < .05$, *** $p < .001$.

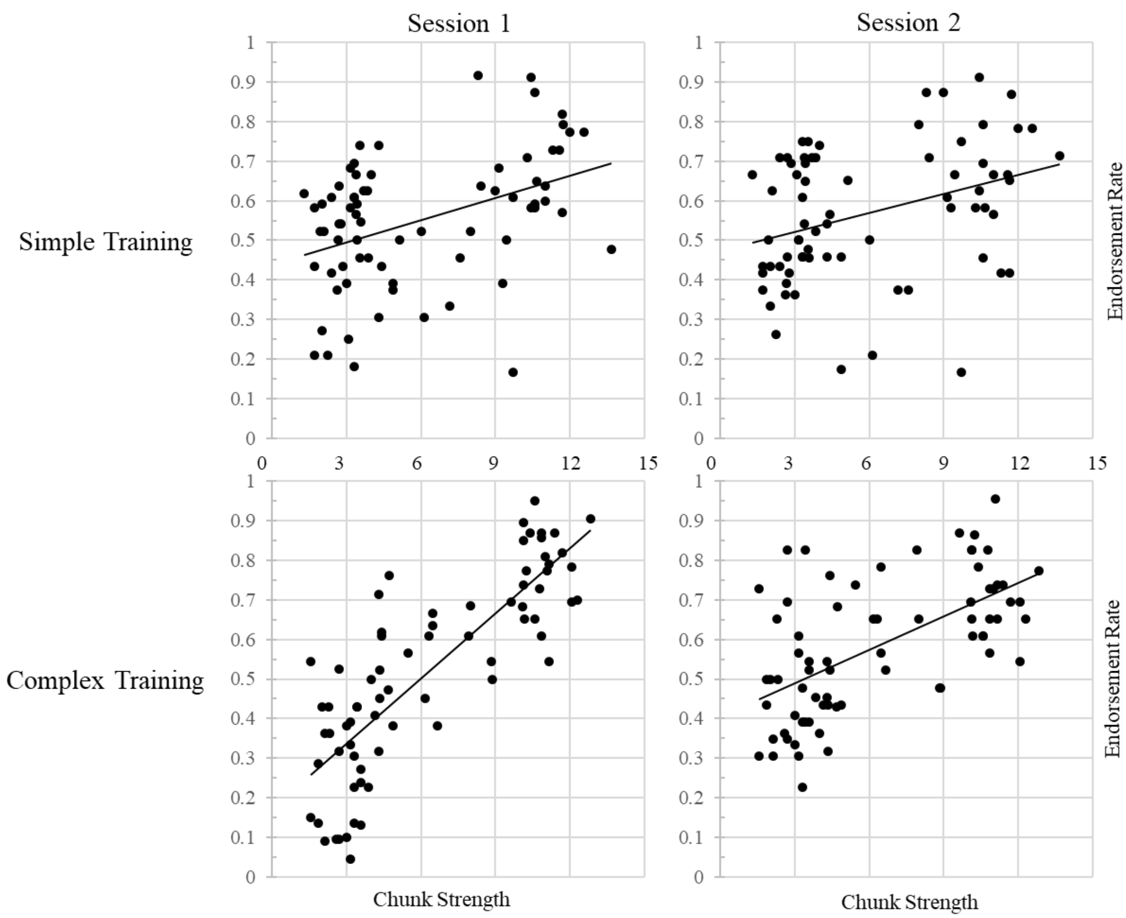


Figure 13. Endorsement rates correlated with chunk strength across sessions for each training group, illustrating the three-way interaction effect in Model 3. Trend lines represents linear lines of best fit.

Item Level Analyses

While we have already reported differences in accuracy between participants based on their training group, we can also look at their profiles of performance based on the features of the test items. This will allow us to determine what aspects of the test items participants used to determine whether or not those test items followed the rules of training. As in Brill-Schuetz (2016), first, we can see that participants in the simple training group performed significantly above chance on accuracy when judging simple test items at sessions one ($t(23) = 2.794$, $p = .010$) and two ($t(23) = 3.799$, $p = .001$). They also performed above chance on complex items at both session one ($t(23) = 4.847$, $p < .001$) and two ($t(23) = 2.600$, $p = .016$). However, participants in the complex training condition only showed above chance accuracy on complex items, and did not retain any knowledge of the grammar between sessions. They performed at chance on simple items at session one ($t(22) = -0.079$, $p = 0.937$) and session two ($t(22) = -0.438$, $p = .666$). This can be contrasted with their above chance performance on complex items at session one ($t(22) = 5.460$, $p < .001$), but not at session two ($t(22) = 0.196$, $p= 0.8468$). See **Table 9** for means and standard deviations.

Participant performance on different item subtypes

	Simple Training		Complex Training	
	Session 1	Session 2	Session 1	Session 2
Simple Items (SD)	61.1% (.17)	59.6% (.19)	49.5% (.24)	48.9% (.14)
Complex Items (SD)	58.5% (.12)	55.3% (.14)	58.2% (.24)	50.5% (.23)

Table 9. Percent correct on the GJT for participants in each training condition at both sessions for both simple and complex test items, along with standard deviations in parentheses.

With the aim of extending the GLMM's findings, we also chose to examine the ways in which accuracy and endorsement varied depending on the surface-level features of each test item at both sessions within either training group. In order to do so, we conducted subsequent analyses on by-items data rather than collapsing across participants. As described in the methods, this meant that the twenty-three participants in the complex training condition and twenty-four in the simple training condition constituted the number of observations across the seventy-two test items, and due to the differing fragment statistics for each training condition, all subsequent analyses treated these groups separately.

To further explore the results of the GLMMs, traditional, frequentist analyses were conducted. Both training groups exhibited a correlation between an item's chunk strength and their endorsement rate. Notably, while the simple training group showed moderate correlations at both sessions one ($r = .409$, $p < .001$) and two ($r = .342$, $p = .003$), the complex training group showed an extremely strong correlation at session one ($r = .819$, $p < .001$), as well as at session two ($r = .598$, $p < .001$). A comparison of these correlation coefficients shows that two groups' correlations are significantly different from one another at both sessions one ($z = -4.23$, $p < .001$) and two ($z = -1.96$, $p = .05$). Also note that both groups showed a reduced reliance on chunk strength at session two when comparing their correlation coefficients, which corroborates the three-way interaction found in Model 3, although this difference was only significant for the complex training group ($z = 2.72$, $p = .006$). To verify the validity of these contrasts, we examined the variation between sub-sets of test items and found that they did not significantly differ between training groups. The mean chunk strength of grammatical items was not significantly different between the simple and complex training conditions ($t(70) = -0.456$, $p = .649$), a pattern that also held true for ungrammatical items (t(70 = -0.429, $p = .670$). This shows that chunk strength was not a stronger cue for either group of participants, suggesting that the complex group's reliance on it was not merely because it was more useful for them in terms of differentiating grammatical and ungrammatical items at test.

A key difference between training groups also emerged when looking at how the chunk strength of each item correlated with participants' accuracy when judging the grammaticality of that item. Only participants in the complex training condition showed a statistically significant relationship between accuracy and chunk strength, and they did so at session one ($r = .300$, $p = .010$), as well as at session two ($r = .248$, $p = .035$), while those in the simple training condition did not at either session one ($r = .187$, $p = .116$) or session two ($r = .139$, $p = .244$). This underscores the complex training group's reliance on the surface level properties of the test stimuli when engaged in the GJT.

Discussion

The set of results described above demonstrates that first, learners seem to be able to retain the regularities of an artificial grammar over the span of two weeks. This is a much longer time interval than what is typically found in the literature on SL. Extensive research on other kinds of learning and memory has found that participants can recall learned items at rather long intervals, and much like the present study, such memory may even be implicit in nature (Mitchell, 2006; Roediger, 1990; Schacter, 1987; Tulving, Schacter, & Stark, 1982). However, the test items in this study were not present in the training set and were only seen once previously during a test session, where half of the trials were foils. The ability of participants to retain their knowledge of statistically learned dependencies over time is crucial to understanding the way in which experience with linguistic constructions affects later processing (Reali & Christiansen, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). In order for SL to impact language processing in the way it has long been hypothesized (Saffran, 2001), the learned statistical patterns must be retained in memory. This research demonstrates that such retention is possible, and adds support for such theories. Determining the limits of retention for statistically learned regularities should be a priority for future research, as the SL literature has long rested on the assumption that such associations form the foundation for language learning.

In addition, the fact that those trained extensively on simple items exhibited above-chance accuracy performance at both sessions while those trained primarily on the complex stimuli showed fewer signs of learning at either session provides evidence that “starting small” with extensively scaffolded, staged training leads to better learning and retention of grammatical regularities. While participants in the complex training condition did perform above-chance on grammaticality judgments in session one for complex test items, they did not learn the overarching regularities of the grammar that would have allowed them to correctly judge simple test items. Moreover, they did not retain this knowledge between sessions for either subset of test items.

While both training conditions within the present study started small, participants in the simple training condition were given significantly more time to learn from the simpler items. Intentionally reducing the problem space for learners during the early phases of acquisition seemed to improve learning outcomes in this study (see also Conway et al., 2003). Poletiek and colleagues (in press) have recently demonstrated that participants are able to use their memory of previously encoded, simple structures to facilitate their learning of newer, more complex ones. They also point out the importance of incrementally exposing learners to increasingly complex items, rather than simply longer ones. The present research also shows a similar trend to other studies that demonstrate how overrepresenting simplified input in training can lead to improved learning (Pine, 1994; Perfors, Tenenbaum, & Regier, 2011). In the current study, the simple training group benefitted more from these factors than did those in the complex training condition. Such scaffolding reflects the way in which young learners typically acquire language, which suggests that forcing adults to adopt more immature strategies when learning a novel language may confer benefits. Future research into the relationship between second-language learning in adults and intentionally constrained input could be important to shaping pedagogical strategies and our understanding of language acquisition more generally.

Looking at the item-level performance also gives a window into the kinds of information learners are sensitive to and use when making grammaticality judgments. The endorsement rates for each item depict a subtly different strategy for participants in either training condition. While the complex training group did not consistently show signs of learning in that they failed to accurately choose grammatical items and reject ungrammatical ones, they did show sensitivity to the low-level surface-structure of the training items at both sessions, as evidenced by the strong correlation found between item chunk strength and endorsement rate. Notably, this relationship was much stronger for the complex training group. This means that while those in the complex training group did not learn the grammatical regularities as expected, they did demonstrate sensitivity to and retention of the fragment statistics that they acquired during training. Interestingly, they also exhibited a correlation between accuracy on the GJT and item chunk strength while the simple training group showed no such relationship for accuracy; this pattern even persisted at session two, showing that while the complex training group failed to retain knowledge of the general grammatical regularities, they did encode the surface-level information from training and relied on it when processing items at test.

This set of findings fits in well with recent proposals about how the constraints placed on learning by our cognitive abilities shape the way in which we process, and thereby learn, language (Christiansen & Chater, 2008, 2016). The proposed "Now-or-Never bottleneck" refers to the process by which language users must continuously recode and compress linguistic input in order to keep up with comprehension. In this framework, language processing *is* language learning; during comprehension, we must effectively process the input as quickly and accurately as possible before it is overwritten or interfered with by new incoming information. Learners take the information that makes it through the bottleneck as far as they can – in the simple training condition of the present study, the more exposure to simple items may have allowed them to process subregularities more efficiently and thereby better deal with similar patterns in the more complex items, while those in the complex condition were only able to rely on the more surface-level information contained within the chunks that they learned and retained.

Further evidence for this comes from the finding that having better working memory skills ameliorated the negative effects of complex training on grammar learning. Participants with this advantage may have been more successful in initially processing the input in a way that allowed them to learn the higher-order relationships between individual items. The fact that experience with certain types of linguistic constructions has been shown to have major effects on learning and subsequent processing (Reali & Christiansen, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2009) also fits together with these results. It seems likely that learners in the complex condition did not have the experience with more simple sequences required to appropriately process items during training but were able to at least begin learning some lower level associations (i.e. chunk strength), which they heavily relied upon when making grammaticality judgments.

Conclusion

In sum, it appears that increased early exposure to simplified grammatical structures confers benefits to learners. Importantly, the learning of this artificial language is retained in long-term memory in a way that has not been shown previously. This falls in line with theories about childhood language acquisition, and also with new ideas concerning the role of processing constraints on language learning. Overly challenging and complex input seems to derail learners and affects the kind of information they are sensitive to, leading them to rely more on low-level fragment statistics than higher-order associations in comprehension. This pattern of results contrasts with learners who were provided scaffolded input, as they demonstrated better acquisition of the higher-order regularities and relied less on low-level cues when choosing to endorse items as grammatical or ungrammatical.

References

Altmann, G. T. M. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15250–15251.

Antoniou, M., Ettlinger, M., & Wong, P. C. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PloS one*, *11*(7), e0158812.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is lasting and consistent over time. *Neuroscience Letters*, *517*(2), 133–5. https://doi.org/10.1016/j.neulet.2012.04.045

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: Past, present and future. *Philosophical Transactions of the Royal Society B, 372*(1711), 20160047. https://doi.org/10.1098/rstb.2016.0047

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Brill-Schuetz, K. A. (2016). *The Complexity of Input and How It Contributes to Adult Second Language Acquisition*. Unpublished dissertation, University of Illinois at Chicago, Chicago, IL.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873. https://doi.org/10.1016/j.cogsci.2003.06.001

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The Behavioral and Brain Sciences*, *31*(5), 489-508; discussion 509–58. https://doi.org/10.1017/S0140525X08004998

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *The Behavioral and Brain Sciences*, *39*, e62.. https://doi.org/10.1017/S0140525X1500031X

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *00*(August), 1–26. https://doi.org/10.1080/01690965.2011.606666

Conway, C. M., Ellefson, M. R., & Christiansen, M. H. (2003). When less is less and when less is more: starting small with staged input. *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, 270–275.

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322–1331. https://doi.org/10.1016/j.neuropsychologia.2011.02.015

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*, 179–211.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, *48*(1), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2014). Shortened complex span tasks can reliably measure working memory capacity. *Memory and Cognition*, *43*(2), 226–236. https://doi.org/10.3758/s13421-014-0461-7

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. https://doi.org/10.1016/j.tics.2014.12.010

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147,* 70–74. https://doi.org/10.1016/j.cognition.2015.11.010

Frost, R. L. A., & Monaghan, P. (2017). Sleep-driven computations in speech processing. *PLoS ONE, 12*(1), 1–14. https://doi.org/10.1371/journal.pone.0169538

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, 99*(1), 529-534.

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., & Stackhouse, A. (2014). Sleep Underpins the Plasticity of Language Production. *Psychological Science*, *25*(7), 1457–1465. https://doi.org/10.1177/0956797614535937

Gomez, R. L. (2016). Do infants retain the statistics of a statistical learning experience ? Insights from a developmental cognitive neuroscience perspective.

Isbilen, E. S., & Christiansen, M. H. (in press). Chunk-based memory constraints on the cultural evolution of language. *Topics in Cognitive Science*. DOI: 10.1111/tops.12376

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn Psychol*, *21*, 60–99. https://doi.org/0010-0285(89)90003-0 [pii]

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–9. https://doi.org/10.1016/j.neulet.2009.06.030

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 79–91. https://doi.org/10.1037/0278-7393.20.1.79

Kobor, A., Janacsek, K., Takacs, A., & Nemeth, D. (2017). Statistical learning leads to persistent memory: Evidence for one-year consolidation. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-00807-3

Lai, J., & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition*, *118*(2), 265–273. https://doi.org/10.1016/j.cognition.2010.11.011

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. *Implicit and explicit knowledge in second language learning, testing and teaching*, 94-112.

Mitchell, D. B. (2006). Nonconscious Priming After 17 Years. *Psychological Science*, *17*(11), 925–929. https://doi.org/10.1111/j.1467-9280.2006.01805.x

Morgan-Short, K. (2007). A neurolinguistic investigation of late-learned second language knowledge: The effects of explicit and implicit conditions. Unpublished dissertation, Georgetown University, Washington, DC.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language learning*, *60*(1), 154-193.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and Implicit Second Language Training Differentially Affect the Achievement of Native-like Brain Activation Patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933–947.

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *PLoS One*, *7*(3), e32974.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, *17*(1), 56-72.

Newport, E. L. (1990). Maturational constraints on language learning.

Newport, E. L. (2016). Statistical language learning: Computational, maturational, and linguistic constraints. *Language and Cognition*, *8*, 1–15. https://doi.org/10.1017/langcog.2016.20

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning*, *50*(3), 417-528.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306–338. https://doi.org/10.1016/j.cognition.2010.11.001

Pine, J.M. (1994). The language of primary caregivers. In C. Gallaway & B.J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 109-149). Cambridge, UK: Cambridge University Press.

Poletiek, F., Conway, C.M., Ellefson, M.R., Lai, J., Bocanegra, B.R. & Christiansen, M.H. (in press). Under what conditions can recursion be learned? Effects of starting small in artificial grammar learning of recursive structure. DOI: 10.1111/cogs.12685

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1–23. https://doi.org/10.1016/j.jml.2006.08.014

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164-171.

Roediger, H. L. (1990). Implicit memory. Retention without remembering. *The American Psychologist*, *45*(9), 1043–1056. https://doi.org/10.1037/0003-066X.45.9.1043

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*(1), 67–109. https://doi.org/10.1016/S0010-0277(99)00031-1

Romano, J. C., Howard, J. H., & Howard, D. V. (2010). One-year retention of general and sequence-specific skills in a probabilistic, serial reaction time task. *Memory*, *18*(4), 427–441. https://doi.org/10.1080/09658211003742680

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*(4), 493-515.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19489896

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 501–518. https://doi.org/10.1037/0278-7393.13.3.501

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Siegelman, N., & Arnon, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language, 85*, 60-75.

Smalle, E. H. M., Bogaerts, L., Simonis, M., Duyck, W., Page, M. P. A., Edwards, M. G., & Szmalec, A. (2016). Can chunk size differences explain developmental changes in lexical learning? *Frontiers in Psychology*, 6.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming Effects in Word-Fragment Completion are Independent of Recognition Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 336–342.

van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language, 67*(4), 507–520. https://doi.org/10.1016/j.jml.2012.07.008

Van den Noort, M. W., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, *11*(4), 289-296.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–71. https://doi.org/10.1016/j.cogpsych.2008.08.002

CHAPTER FOUR

Right-hemisphere dominance indicates improved language outcomes in patients with left-frontal brain tumors[3]

Abstract

The brain possesses a limited capability to compensate for injury. Such plasticity of brain function has been demonstrated in a number of clinical scenarios including translocation of left language function to the right-sided Broca's and Wernicke's area homologues in brain tumour patients. The present study sought to identify whether language-related right-frontal functional Magnetic Resonance Imaging (fMRI) activation exists as a group-level trend in patients with left-frontal tumours. We also sought to examine the possibility that translocation of language function would lead to a better functional outcome. We therefore conducted a retrospective analysis of 197 brain tumour patients who had undergone pre-surgical fMRI language mapping. Patients with left-frontal brain tumours were found to be more likely to show right- or co-dominant fMRI activation during language mapping tasks compared to patients who had tumours elsewhere in the brain. Further, patients with left-frontal tumours who were identified as right- or co-dominant for language were found to possess more intact language function as measured by the Boston Naming Test. In order to highlight the major findings, three illustrative case studies

[3] Co-authored with Nicole P. Brennan, Kyung K. Peck, Viviane Tabar, Cameron Brennan, Morten H. Christiansen, and Andrei I. Holodny; currently in preparation for submission.

are presented in further detail, which also depict the individual differences we find within this patient cohort.

Introduction

Those studying the cognitive neuroscience of language have described the brain's left hemisphere as specialized for language ever since the seminal work of Paul Broca and Carl Wernicke. These studies, along with the century and a half of research they inspired, have focused primarily on uncovering the roles of the left frontal and temporal lobes in the learning and processing of language (Dronkers, Plaisant, Iba-Zizen, & Cabanis, 2007). This has led to an emerging characterization of the role that various regions, sub-regions, and pathways within the left-hemisphere's language network play in language processing (Hagoort, 2014; Friederici, & Gierhan, 2013).

However, there is a more limited understanding of the role of the right hemisphere in language, and the debate on the degree of specialization in each hemisphere remains open to further exploration. Even with the focus of research on left-hemisphere language function, some theories posit a complementary role for right-hemisphere processing (Jung-Beeman, 2005). In addition, there is much evidence that the right hemisphere, specifically Broca's homologue, can take over function from the left hemisphere following extensive early brain damage (Thal et al., 1991; Vicari et al., 2000; Tivarus, Starling, Newport, & Langfitt, 2012).

Yet the literature investigating whether adults with brain lesions retain this capacity is currently inconclusive. Some smaller-scale work has suggested that increased right-hemispheric activation may be correlated with better language outcomes in patients with left-frontal lesions, and in healthy patients who undergo targeted rTMS (Krieg et al., 2013). Others have found that in stroke patients with aphasia, increased activation in the right superior temporal sulcus was correlated with improved outcomes, while activation of the left dorsal pars opercularis was associated with poorer outcomes (Skipper-Kallal, Lacey, Xing, & Turkeltaub (2017). Furthermore, recent work suggests certain patterns of lesion can lead to oscillatory dynamics that shift language function to the right hemisphere as a compensatory mechanism (Piai, Meyer, Dronkers, & Knight, 2017). It seems likely that while the right hemisphere can play a role in recovery for patients with aphasia, the dynamics of this potential reorganization may differ from patient to patient, and etiology to etiology (Turkeltaub, 2015).

Determining the degree to which the adult brain can reorganize language function, and under what conditions this occurs, promises to yield important insights that not only may inform clinical prognoses, but also lead to a better understanding of the brain’s equipotentiality. Examining brain tumor patients may provide a fruitful source of data for uncovering the potentially more subtle plasticity that exists in adult populations than acute stroke and brain damage patients, as they may exhibit different compensatory mechanisms and competencies (Fisicaro et al., 2016).

Case studies with brain tumor patients have indicated that damage to traditional language cortex in both the frontal (Holodny, Schulder, Ybasco, & Liu, 2002) and temporal (Petrovich, Holodny, Brennan, & Gutin, 2004) lobes can result in contralateral reorganization of the brain’s language network, although how this affects patients’ cognitive outcomes is not currently known. In addition, the fact that ipsilateral reorganization also occurs in this patient population (Brennan, 2008), creates an ideal setting in which to test for both the relative frequency of contralateral reorganization and its outcomes with respect to language. Does the right hemisphere demonstrate compensatory activation when there is damage to the left-frontal lobe more often than the occasional case study would suggest? Also, do patient outcomes differ depending on whether the patient exhibits ipsilateral versus contralateral compensation? This set of questions led to the present study, in which a large database of patients with brain tumors of varying grade, size, location, and etiology was queried to answer these critical questions. We hypothesized that the increased right-frontal activation seen in such case studies exists as a group-level trend among patients suffering from left-frontal tumors, and also that this activation is compensatory in nature.

Results

Plasticity in the Brain's Language Network

In order to determine the possible extent of both ipsilateral and contralateral reorganization following tumor infiltration of the left-frontal lobe, we divided all right-handed patients in the original sample (n = 159, 72 female; mean age = 50.64, range: 10-83) into groups based on tumor location. Patients with tumors impacting the left-frontal lobe (n = 81, 37 female; mean age = 49.65 years, range: 11-75) can be considered the experimental group, while patients with tumors elsewhere in the brain – including other regions within the left hemisphere – (n = 78, 35 female; mean age = 52.35, range: 10-83) served as the control. There were no significant differences between patient groups in terms of age ($t(156) = -1.145$, $p = .254$), sex ($\chi2 = .01$, $p = .920$), or tumor grade ($\chi2 = 2.843$, $p = .092$).

As can be seen in **Table 10**, patients with tumors localized in the left-frontal lobe were much more likely to be reported as right- or co-dominant for language function than would be expected given the control group ($\chi 2 = 9.51$, $p = .002$, $\varphi = .245$). This finding suggests that the frontal component of the typical left-hemisphere language network can shift to the right hemisphere in patients with tumor impacting putative Broca's area in the left-frontal lobe. It also demonstrates that this contralateral reorganization happens with some frequency, and is not isolated to a small number of cases. However, it remains unclear whether the plasticity encountered in this patient population has any effect on language function. It is possible that the increased right-hemisphere activation is not compensatory in nature, and may even be deleterious.

χ^2-table corresponding to reported laterality

	Patients reported as left-dominant	Patients reported as right-or co-dominant	(Total)
Patients with left-frontal tumors	67 (72.85)	14 (8.15)	81
Patients with tumors elsewhere in brain	76 (70.15)	2 (7.85)	78
(Total)	143	16	159

Table 10. χ^2-table exhibiting the difference in report laterality between patients with left-frontal tumors and those with tumor elsewhere in the brain. Expected values reported in parentheses.

Patterns of Reorganization Affect Language Outcomes

To elucidate the role that potential right-hemisphere reorganization plays in actual language outcomes, patients with left-frontal tumors who completed a Boston Naming Task

(BNT) during their pre-surgical assessment ($n = 28$) were evaluated more closely. To determine the effect of laterality on language performance, the BNT scores of patients with left-dominant reports ($n = 20$; $M = 48.30$; $SD = 14.71$) were compared to those with right- or co-dominant reports ($n = 8$; $M = 56.38$; $SD = 2.72$). Again, no significant differences were found between groups for age ($t(156) = -1.145$, $p = .254$), or tumor grade ($\chi 2 = 2.05$, $p = .152$), although sex and hemisphere of language dominance were not found to be independent ($\chi 2 = 4.01$ $p = .045$, $\varphi = .223$), as males were more likely to be right- or co-dominant than would be expected if the two variables were randomly distributed.

As depicted in **Figure 14**, this analysis found that patients identified as being right- or co-dominant in terms of language function had significantly more intact language abilities than did patients with left-dominant maps exhibiting ipsilateral, and often perilesional activation to language tasks ($t(22) = -2.36$, $p = .028$). This independent-samples t-test did not pass Levene's test for equality of variances ($F = 4.82$, $p = .037$), so corrected degrees of freedom were used. This set of analyses demonstrates that the right hemisphere can effectively take over language function in adults who were likely left-dominant prior to tumor development. It also opens the possibility that therapeutic techniques which focus on transferring language function to the right hemisphere in the face of left-hemisphere damage may result in better patient outcomes.

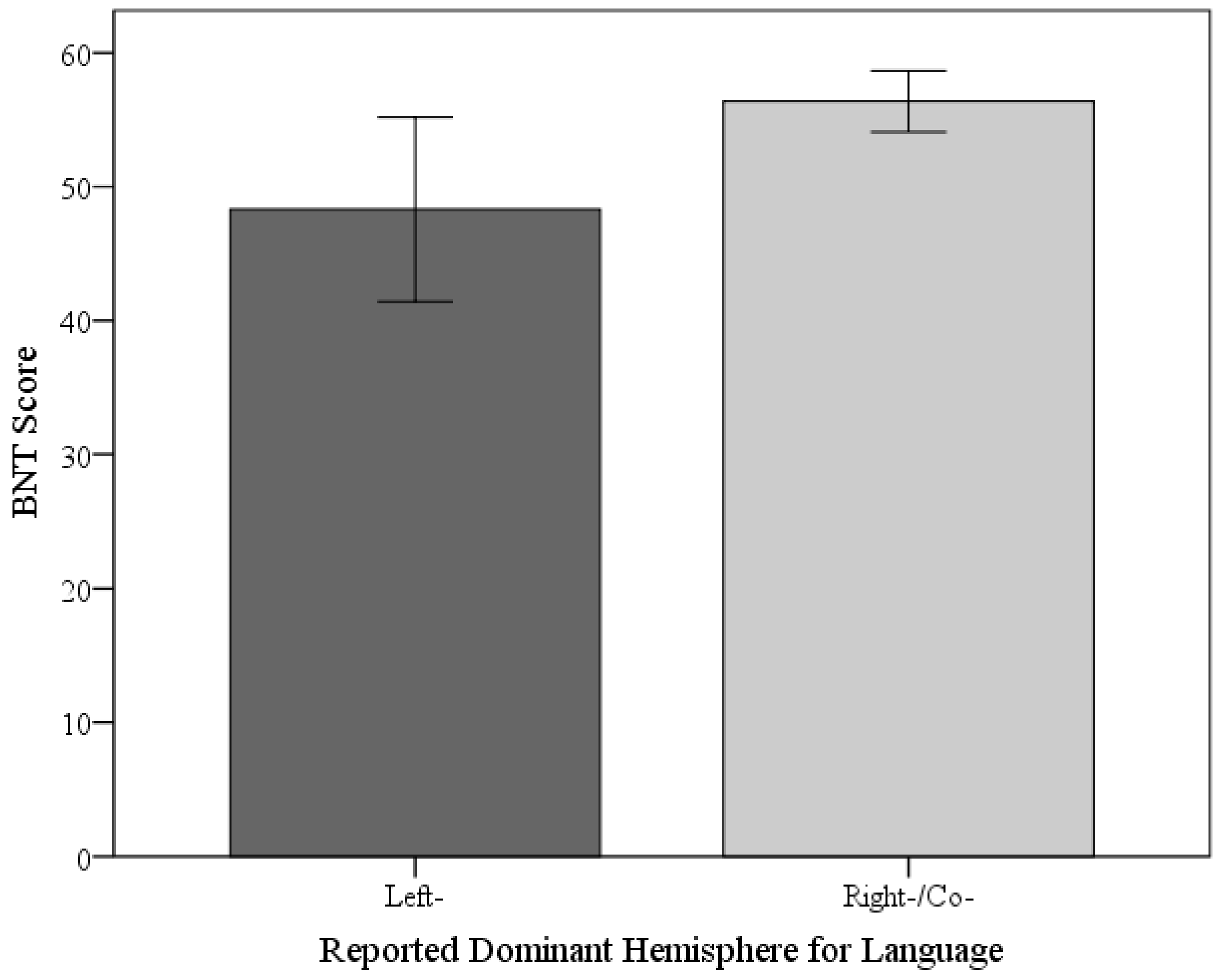


Figure 14. Bar graph displaying the difference in BNT performance between patients with left-dominant reports and those with right- or co- dominant reports. Error bars indicate 95% CIs.

Laterality Index Verification of Report Groups

To confirm the radiologist report data, a laterality index assessment also was conducted on the patients with BNT data. The patients in both the left-dominant and right-/co-dominant groups were also evaluated using laterality indices (LI), with interior frontal gyrus (IFG), middle frontal gyrus (MFG), frontal lobe, and hemispheric ROIs (see **Figures 15-18**). Note that the frontal lobe ROI was computed as a combination of the MFG and IFG ROIs.

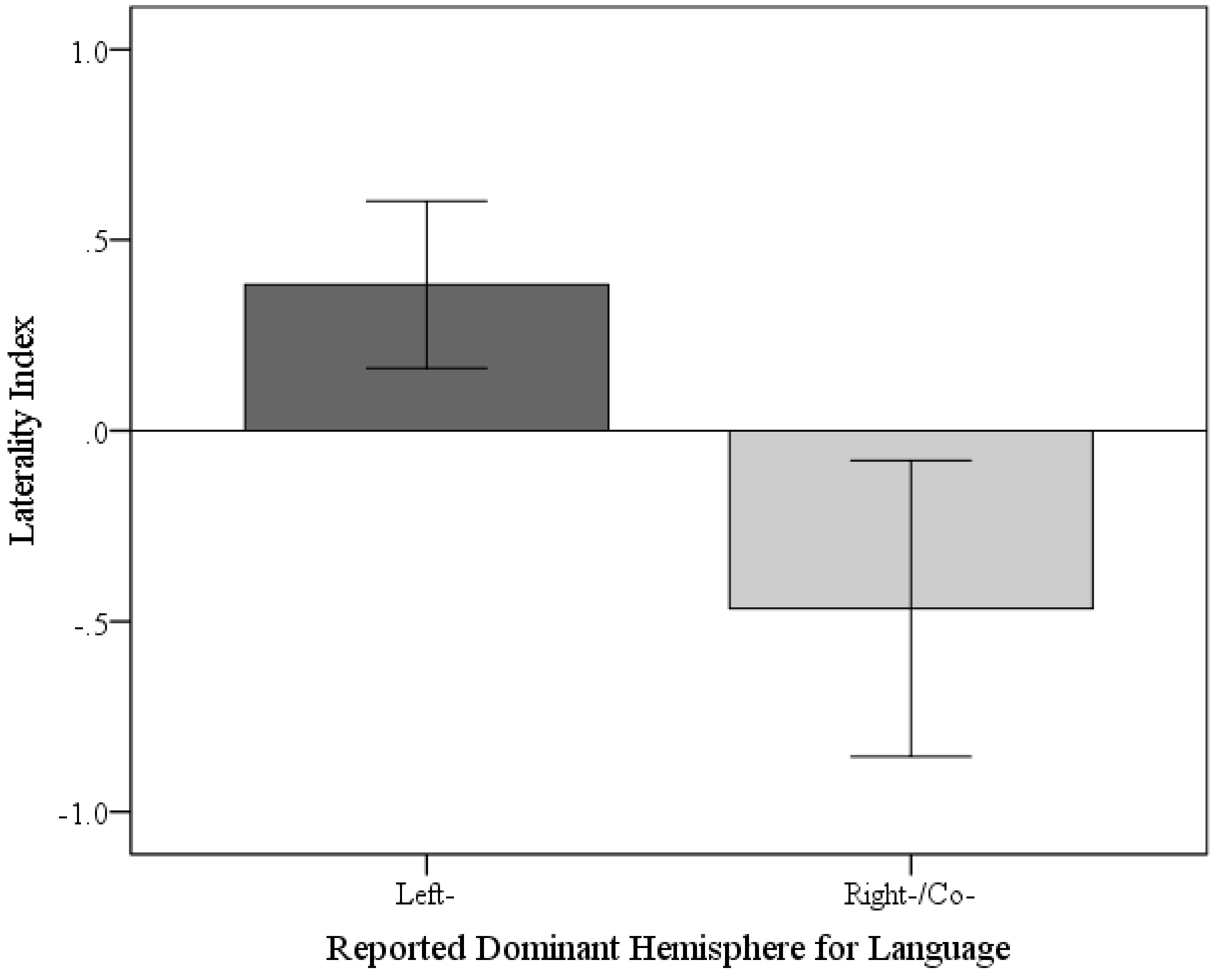


Figure 15. Difference in LI between patients with left-dominant reports and those with right- or co- dominant reports in the IFG ROI. Error bars indicate 95% CIs.

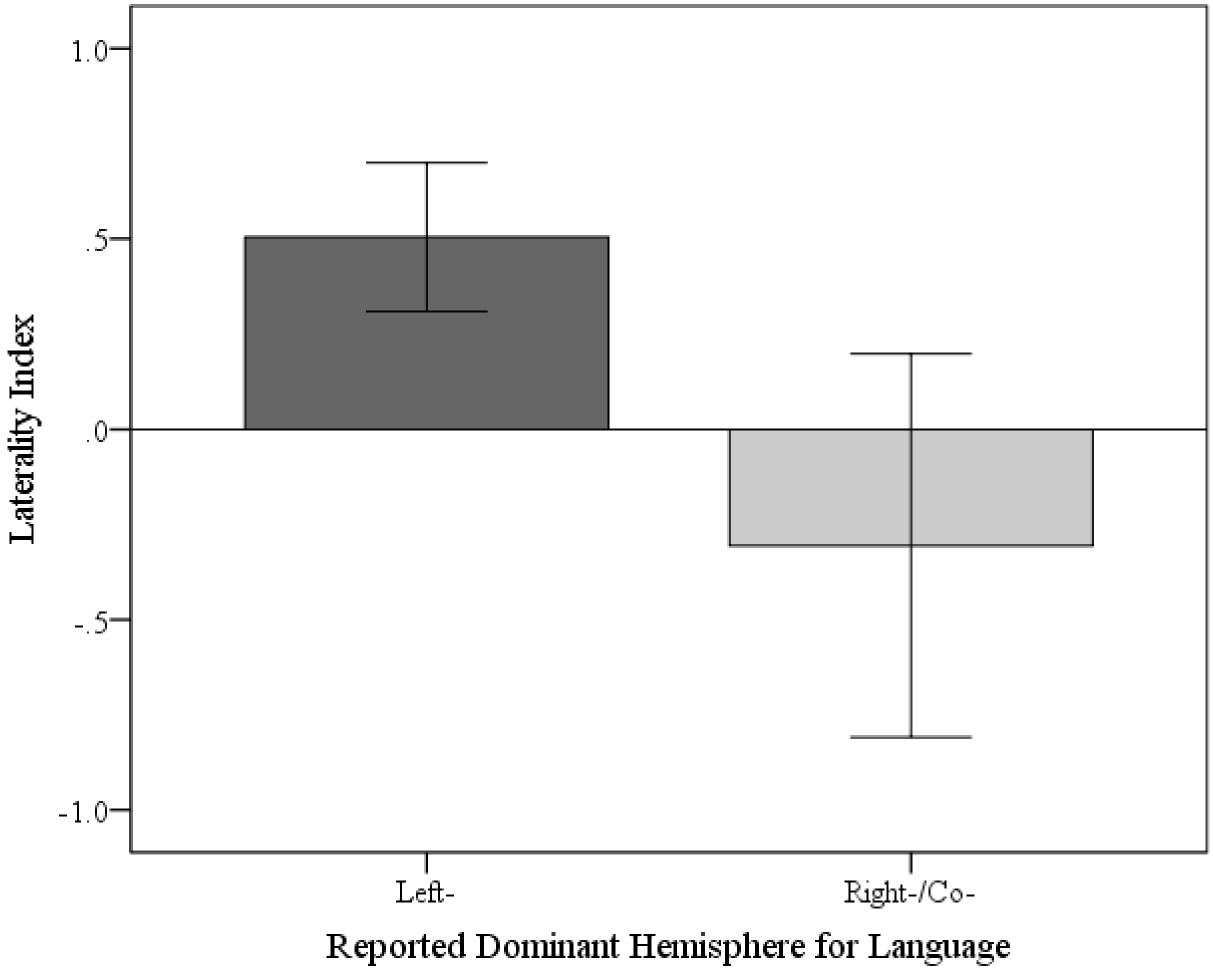


Figure 16. Difference in LI between patients with left-dominant reports and those with right- or co- dominant reports in the MFG ROI. Error bars indicate 95% CIs.

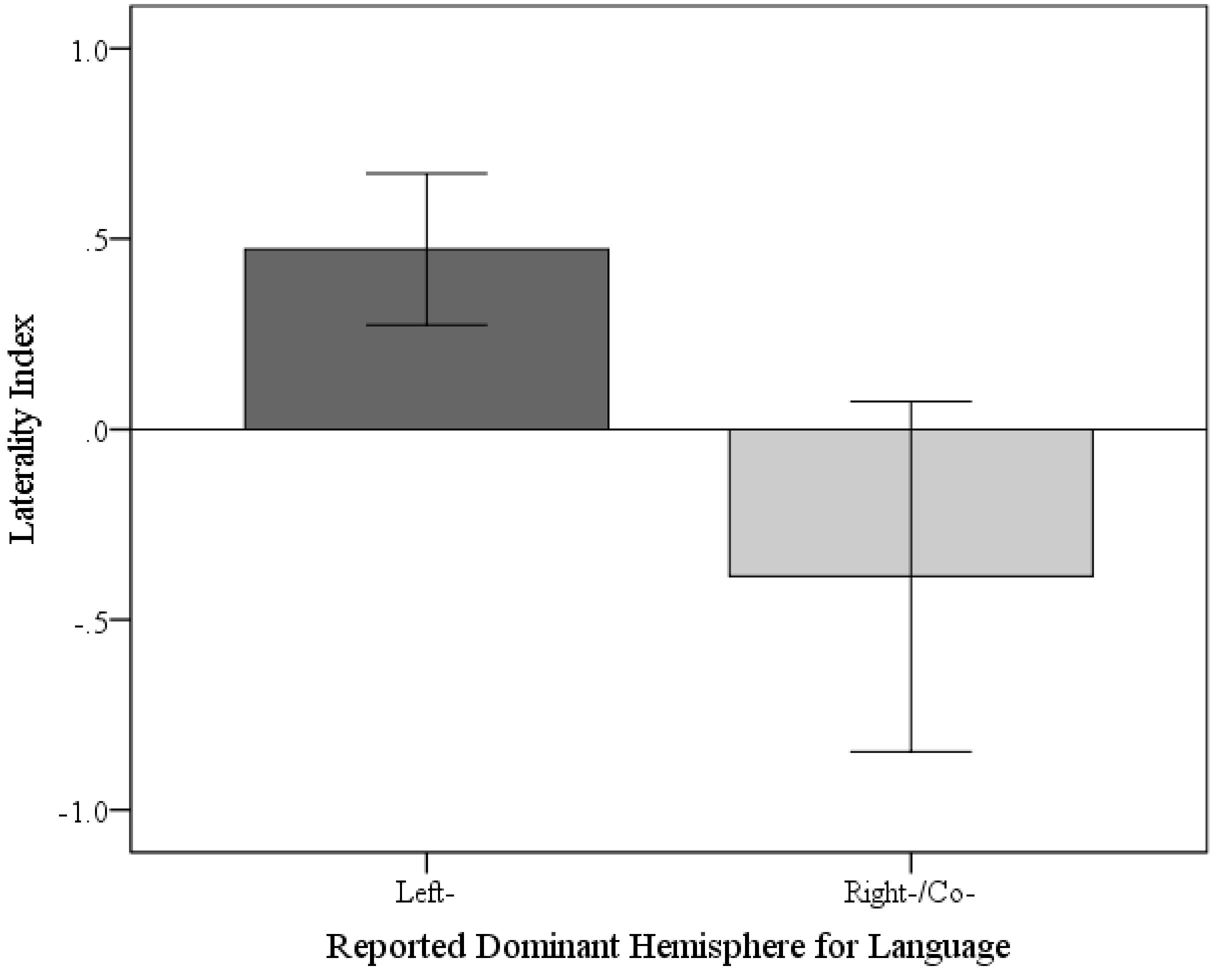


Figure 17. Difference in LI between patients with left-dominant reports and those with right- or co- dominant reports in the frontal lobe ROI. Error bars indicate 95% CIs.

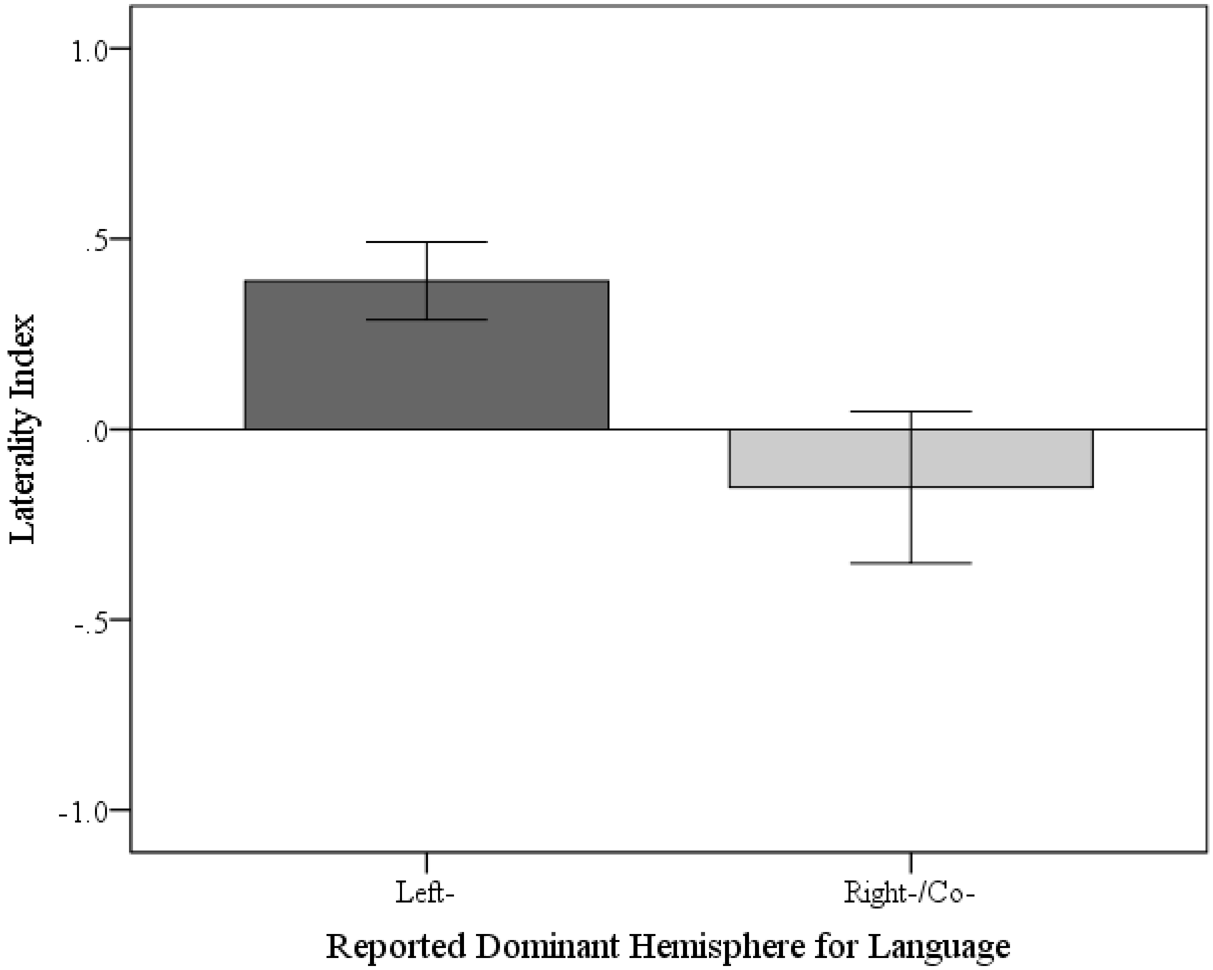


Figure 18. Difference in LI between patients with left-dominant reports and those with right- or co- dominant reports in the hemispheric ROI. Error bars indicate 95% CIs.

Independent-samples t-tests confirmed that there was a significant difference between the report groups in laterality for each of the ROIs that were assessed, the IFG ($t(26) = 4.33$, $p < .001$), MFG ($t(26) = 4.08$, $p < .001$), frontal lobe ($t(26) = 4.44$, $p < .001$), and hemispheric ($t(26) = 5.78$, $p < .001$). See **Table 11** for a report of means for each group’s LI within each ROI. This set of analyses corroborates the qualitative findings of the report data and suggests that most patients falling into the left- and right-/co-dominant groups were categorized in the same way with both methods.

Overview of LIs within each ROI

	IFG	MFG	Frontal	Hemispheric
Left-dominant	0.383	0.504	0.472	0.389
Right-/co-dominant	-0.466	-0.306	-0.387	-0.153

Table 11. Group LI means for each ROI, separated by groups identified within the radiology team's reports.

Further Examination Using Quantitative Methods

By utilizing radiologist reports for evaluations of laterality, the previously described analyses are partially based on qualitative measures. Because several patient scans suffered from issues related to drop-out artifacts caused by prior surgery along with other common issues faced when scanning this patient population, traditional quantitative analyses may not be optimal for determining laterality in this patient population.

However, to make the results more transparent and interpretable, we explored this relationship between language outcomes and functional organization using traditional quantitative measures as well. Specifically, we examined the correlation between LIs and BNT scores directly. As depicted in **Figure 19**, the relationship one would expect based on the qualitative analyses between higher scores on the BNT and right-hemisphere dominance stands, although it does not reach significance (for the IFG ROI, r = -.239, p = .22).

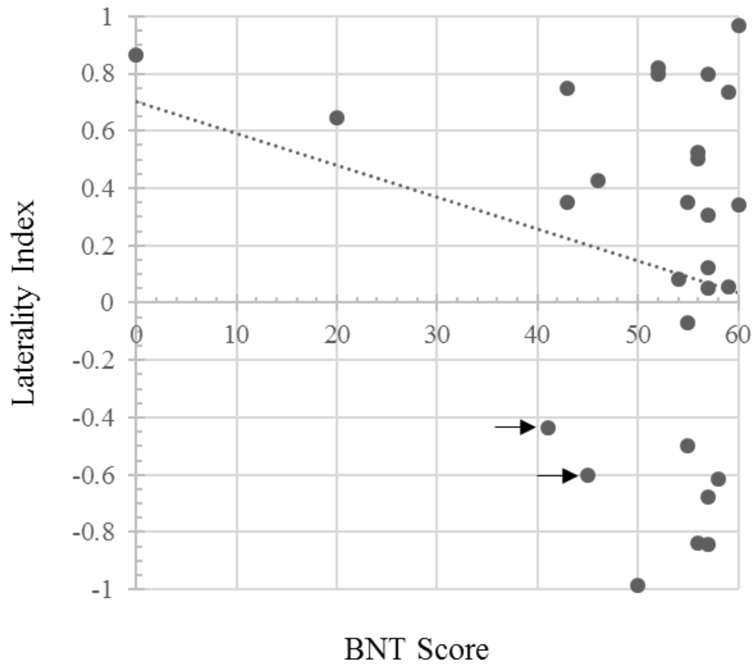


Figure 19. Negative trend in the correlation between BNT scores and LI indicates some agreement between quantitative and qualitative assessments of the relationship between right-hemispheric activation and preserved language abilities. Note the data points highlighted with arrows, as these were patients who had negative LIs due to low voxel counts, and were reported as left-dominant by the radiology team based on additional scans and patient history.

This is due in large part to two patients who exhibited an incongruence between language dominance as determined by LI and the radiology report. Both patients (highlighted with arrows in **Figure 19)** were determined to be left-dominant for language by the radiology team, but due to drop-out artifacts (caused by susceptibility from hemorrhages or prior surgery) had diminished activation in the affected hemisphere, leading to a negative, right-dominant, LI as determined by a purely quantitative ROI.

This relationship depicts the limitations of using purely quantitative analyses in clinical settings (Peck et al., 2009). While at the group-level the trends match up well, when looking at individual patients there are strong reasons to qualify and interpret what the raw numbers are saying. Even when working with a relatively large sample for this type of patient cohort, a few patients with less than ideal scans can have a dramatic influence on the data. It is also important to note that while patients with left-dominant maps were more likely to experience language deficits, most of them scored rather well on the BNT. The heterogeneity of patient populations does not lend itself to a one-size-fits-all approach to using LIs for diagnosis and eventual treatment. Rather, the patient report data seems to offer more clarity due to the fact that it can more readily adapt to the specifics and potential idiosyncrasies involved in working with patient data that was collected for clinical purposes.

Case Reports Highlighting Individual Differences in Language Outcomes

To further illustrate the complex relationship between tumor location, functional organization, and language abilities, we discuss a set of patients with similar etiologies but with a range of outcomes. These three patients, all with large, left-frontal tumors, were first examined after experiencing language problems. All underwent pre-surgical language mapping and were also assessed using the BNT (see **Table 12**). This section aims to connect traditional case studies in the field with the present studies' attempts to extend such findings.

Case-report patient details

				Calculated LI			
Patient	Report	BNT	Grade	Hemispheric	IFG	MFG	Frontal
A	Left-dominant	55	Low	0.43	-0.50	0.60	0.58
B	Left-dominant	41	High	0.59	0.73	0.88	0.84
C	Right-dominant	57	High	-0.55	-0.84	-0.99	-0.92

Table 12. Overview of language and scanning data for three illustrative cases.

Patient A (**Figure 20**), a 50-year-old female, presented with a single episode of speech arrest, and was diagnosed with a low-grade glioma. Her tumor impacted a large portion of the left-frontal lobe, including the inferior frontal gyrus. The results of her pre-surgical fMRI mapping procedure indicated left-dominance for language as reported by the radiology staff, and was corroborated by a later LI calculation (see **Table 12**). Note that she had very little activation (8 total voxels) within the IFG ROIs, leading to a skewed LI. Patient A is an example of a patient with a left-frontal tumor who is left-dominant for language and maintained relatively intact language function (BNT = 55). Her case is also an illustration of why using purely quantitative measures may occasionally not tell the whole story when dealing with certain patient populations.

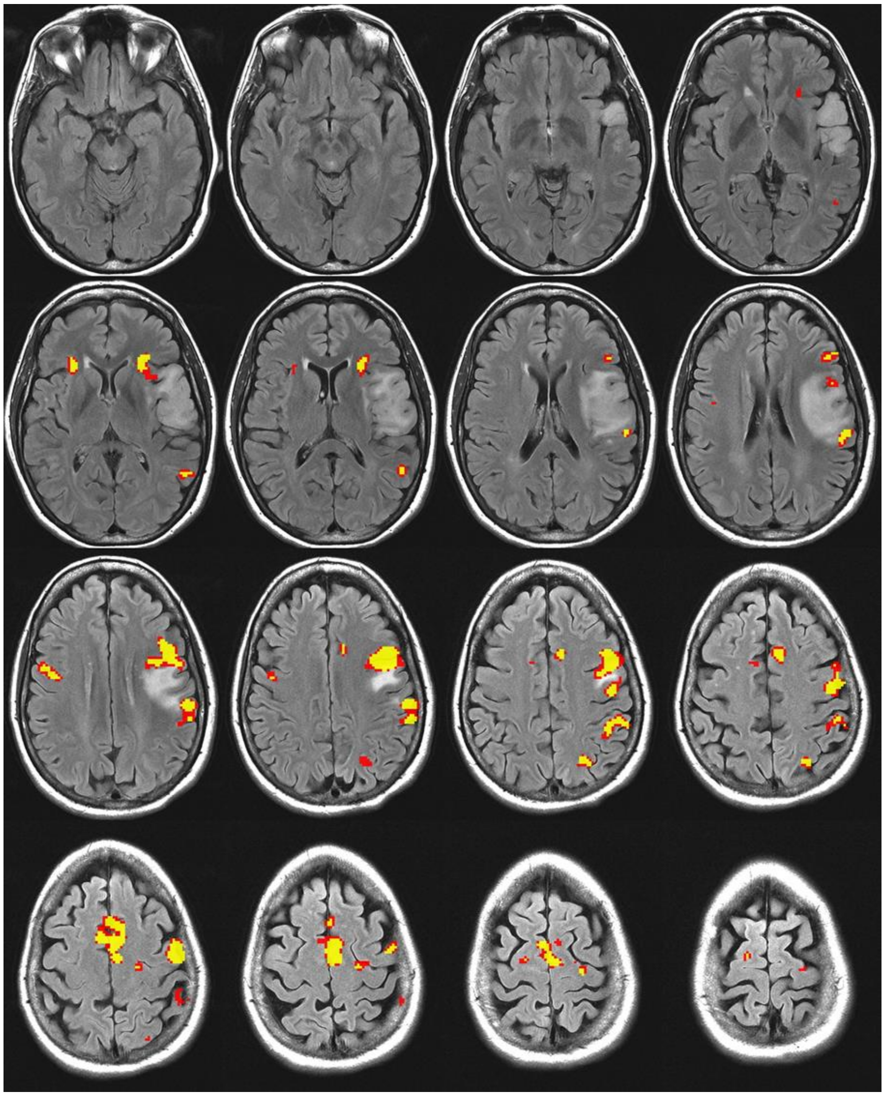


Figure 20. fMRI of activation during a Phonemic Fluency task by Patient A, indicating left-dominance for language.

However, as described above, patients with left-dominant language maps were actually more likely than those with right- or co-dominant maps to exhibit poor language outcomes. Patient B (**Figure 21**), a 69-year-old female, presented with paraphasia and word-finding difficulties, and was diagnosed with a large, high-grade glioblastoma impacting the frontal lobe, including the posterior portion of the inferior frontal gyrus. The results of her pre-surgical fMRI mapping procedure indicated left-dominance for language as reported by the radiology staff, and was corroborated by a later LI calculation (Hemispheric LI: 0.59; see **Table 12**). Patient B is an example of a patient with a left-frontal tumor who is left-dominant for language yet performed poorly on a language assessment (BNT = 41).

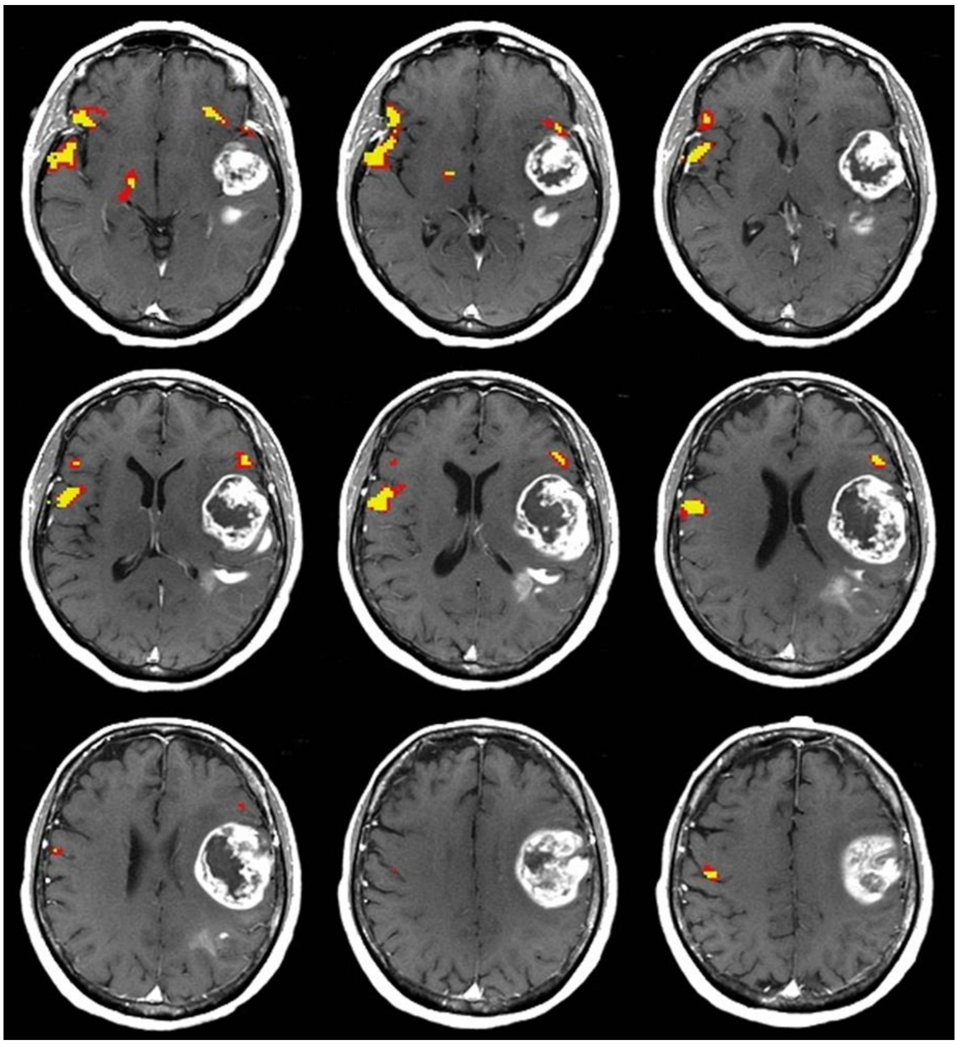


Figure 21. fMRI of activation during a Phonemic Fluency task by Patient B, indicating left-dominance for language.

Finally, Patient C (**Figure 23**), a 43-year-old female, presented with some dysarthria and was subsequently diagnosed with a high-grade astrocytoma impacting the left-inferior frontal gyrus. The results of her pre-surgical fMRI mapping procedure indicated robust right-dominance for language as reported by the radiology staff, which was corroborated by a later LI calculation (Hemispheric LI: -0.55; see **Table 12**). Patient C is an example of a typical right- or co-dominant patient from this cohort, as she did not experience clinically meaningful language dysfunction, scoring high on her assessment (BNT = 57).

Taken together, these patients represent the three sets of outcomes seen within the patient cohort examined in this study. Patients A and B illustrate the divided outcomes that patients with ipsilateral/perilesional compensation are prone to. While most patients with left-dominant language maps and left-frontal tumors did have relatively preserved language function, they were also much more likely to have clinically significant language deficits – seven out of these twenty patients had BNT scores under the clinical threshold of 50, while none of the eight right- or co-dominant patients performed below that level.

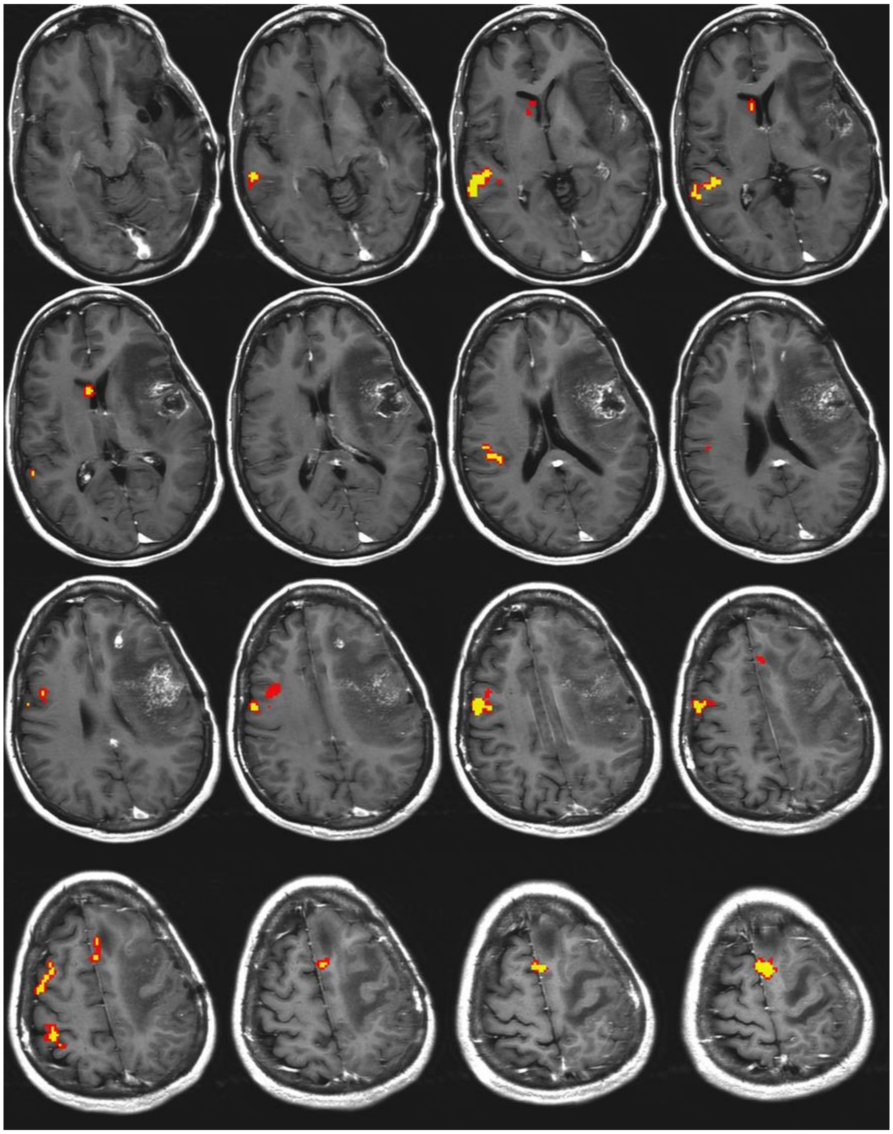


Figure 22. fMRI of activation during a Phonemic Fluency task by Patient C, indicating right-dominance for language.

Discussion

This is the first large scale study investigating the complex relationship between reorganization in the brain's language network and behavioral outcomes in patients with brain tumors. While past research with stroke patients has incorporated large datasets (Turkeltaub, Messing, Norise, & Hamilton, 2011), case studies have made up much of the past research on atypical language lateralization in adult brain tumor patient populations.

The findings of the present study point toward a degree of plasticity in the adult brain's language network that is somewhat unexpected. The frontal lobe, in the right hemisphere, likely Broca's homologue, seems to be able to take over language function when the left hemisphere is compromised, and may even be a better candidate for reorganization than the ipsilateral cortex. It is also of note that the patients with left-hemisphere damage exhibiting right- or co-dominant activation maps fell within the normal-range of scores on the BNT, demonstrating the right hemisphere's ability to maintain healthy levels of language performance. Patients with left-dominant maps indicating ipsilateral, and often peri-lesional compensation suffering from left-frontal lesions exhibited a much wider range of outcomes.

Such a pattern of results suggests that while typical development results in an overwhelming majority of adults having left-dominant language networks (Knecht et al., 2000), the right hemisphere may also possess the requisite capacities for language processing. This falls in line with the work of Bates and colleagues, which suggests that early language development can occur in children with severe damage to the left hemisphere and in children who are lacking their left hemisphere entirely (Thal et al., 1991; Vicari et al., 2000; Bates & Roe, 2001). It is likely that the left hemisphere is not uniquely suited or specifically adapted for language learning and processing. Rather, as suggested in these examples of prior research, relatively ubiquitous and subtle biases in early development likely lead to the typically left-hemisphere language network. Future work determining the nature of these biases would be a fruitful addition to our understanding of the developmental intricacies of the brain's language network.

Other clinical studies have also indicated that the right hemisphere may be able to serve what is traditionally the role of the left hemisphere in language in some circumstances (Fisicaro et al., 2016). An investigation of a large number of stroke patients suffering from aphasia demonstrates that patients suffering from left-frontal damage recruit the right hemisphere more often than healthy controls, although the findings of this study did not determine whether the compensatory activation is deleterious or beneficial (Turkeltaub et al., 2011).

Past case-study research has demonstrated an unclear and even conflicting relationship between the role of right-hemispheric activation and language outcomes in patients with aphasia at different points in recovery (Turkeltaub et al., 2012). Unlike this past research, however, the present study demonstrates a clear relationship between right-hemisphere activation during language tasks and positive patient outcomes in a large cohort. Determining whether this is true for stroke patients, or patients suffering from brain damage with other etiologies constitutes an intriguing question for future research.

Studies utilizing therapeutic techniques attempting to transfer language function to healthy right-hemisphere cortex after damage to the left-frontal lobe in two stroke patients have demonstrated that the right hemisphere may be able to take over such function when the basal ganglia remains intact, and that this compensation leads to positive language outcomes for the patient (Crosson et al., 2005). Other recent research also seems to suggest that a rightward shift in language related function is related to the amount of tumor infiltration suffered by the basal ganglia (Shaw et al., 2016).

This has led to the hypothesis that the basal ganglia activation serves as the causal mechanism by which language function can transfer between hemispheres following left-hemisphere damage (Shaw et al., 2016). Further research examining the trajectories and limitations of this transfer will hopefully not only elucidate mechanisms for contralateral reorganization in clinical populations, but also increase the understanding of how the hemispheres communicate during normal language processing. Doctors and patients can also perhaps change expectations for surgery and treatment as a result of this research. These findings, along with future work demonstrating the feasibility of potential language transfer therapies, could lead to more aggressive tumor resections, and improved outcomes for patients post-injury and post-surgery (Duffau, 2006).

The present study is limited in some regards. By utilizing radiologist reports for evaluations of laterality, the main analyses are partially based on qualitative measures. Several patient scans suffered from issues related to drop-out artifacts due to prior surgery and other common issues faced when scanning this patient population, rendering traditional quantitative analyses less useful. However, the limited laterality index findings reported here suggest that the radiologist report data overlaps substantially with more quantitative measures, at least at the group level, mitigating such concerns. In addition, the radiology team's reports were able to account for such issues in their analyses. For example, several patients whose scans exhibited multiple artifacts were difficult to categorize using traditional laterality indices due to low voxel counts in the affected hemisphere. In fact, recent research suggests that individualized fMRI language maps, like those used to generate the radiology team's reports in the present study, actually outperform fixed threshold maps for this patient population (Benjamin et al., 2017). Ideally future research will be able to control for artifacts and scanning issues in a more well-controlled prospective study. Future prospective research would also benefit from utilizing a broader range of neuropsychological assessments than were available in the present study.

The detailed explanations included for three illustrative cases should ameliorate some of these concerns as well, as they depict the range of outcomes seen within this sample. While the desire to move beyond the case study tradition strongly motivated this work, we hope that including these exemplars serves to more accurately and clearly depict the kinds of issues faced by doctors and researchers working with similar patient cohorts.

The present study is the first to indicate a group-level trend in putative right-hemisphere reorganization following damage to traditional left-frontal language cortex. It is also the first to demonstrate positive outcomes for patients because of such reorganization in a large sample. These findings show that the left hemisphere is not likely to be uniquely adapted to subserve language function. Rather, the right hemisphere seems to be able to take over for the left hemisphere when the latter is damaged, and can maintain a normal-level of language abilities. This study also highlights the potential for therapies attempting to facilitate contralateral shifts in language function in patients suffering from language problems. As the right hemisphere seems to be just as adept at subserving language function as the left hemisphere, therapy targeted at shifting language function to healthy right hemisphere cortex seems like an increasingly attractive option.

Method

Participants

A database was compiled involving all patients who underwent a pre-surgical fMRI language procedure at Memorial Sloan Kettering Cancer Center over a five-year time period, excluding patients with sub-optimal scans or incomplete patient information (n = 197, 95 female;

mean age = 50.52 years, range: 10-83). Patients initially presented with a range of symptoms and tumor etiologies. **Table 13** indicates the proportion of patients falling into different relevant categories. The IRB of Memorial Sloan Kettering Cancer Center exempted this retrospective analysis as it was considered a study of existing data.

Patient information

	Female	High-grade tumor	Left-frontal tumor	Right-handed	Left-dominant language
Proportion of Patients	0.482	0.538	0.482	0.807	0.873

Table 13. General information of patients included in the present study.

Data Acquisition

Data were acquired with a 1.5-T or 3.0 T scanner (General Electric, Milwaukee, WI) using an 8-channel head coil. Based on localizer images, a set of 26 T1-weighted (repetition time [TR], 600ms; echo time [TE], 8ms; thickness, 4.5-mm) and T2-weighted (TR, 4000ms; TE, 102ms; thickness, 4.5-mm) spin-echo axial slices, covering the whole brain, was obtained for coregistration with the functional data. Functional images were acquired with a gradient-echo echo-planar imaging sequence (TR, 4000ms; TE, 30/40ms (for 3T/1.5T); matrix, 128x128; field of view, 240 mm; thickness, 4.5 mm; flip angle, 90°). Head motion was minimized using straps and foam padding.

Task Administration

Patients performed one or more covert block designed fMRI language tasks as part of the pre-surgical language task panel. Tasks consisted of a phonemic fluency task where they were required to generate words that began with a high frequency letter, semantic fluency where they were required to generate words that fit a category, verb generation where they were required to generate verbs to given nouns, or auditory responsive naming where they were required to answer simple questions. All tasks were delivered aurally (Ruff et al., 2008). The radiology team's reports were based on all scans collected for each patient, while reported laterality indices were taken from the task with the highest quality scan, with a preference for the phonemic fluency task.

There were 90 images in total for each patient scan, consisting of 5 activation images (20 sec) followed by 10 rest images (40 sec) repeated 6 times (6 min total). Subjects were monitored continuously while performing the task. Subject participation was confirmed using real-time imaging software, which provided real-time acquisition, processing, and display of functional results (Brainwave RT, GE Healthcare, Milwaukee, WI).

All language fMRI tasks were visually inspected for lateralization patterns and discrepancies in language lateralization by a board certified neuroradiologist. Clinical reports indicated language lateralization, language localizations in the peri-tumoral region, handedness, and if the BNT was performed, the patient's score.

References

Bates, E., & Roe, K. (2001). Language Development in Children with Unilateral Brain Injury. In C.A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience.* Cambridge, MA: MIT Press.

Benjamin, C. F. et al. (2017). Presurgical language fMRI: Mapping of six critical regions. *Human Brain Mapping, 38*, 4239–4255.

Brennan N.P. (2008). Preparing the patient for the fMRI Study and optimization of paradigm selection and delivery. In A.I. Holodny (Ed.), *Functional neuroimaging: A clinical approach*. New York: Informa Healthcare USA, Inc.

Crosson, B. et al. (2005). Role of the right and left hemispheres in recovery of function during treatment of intention in aphasia. *Journal of Cognitive Neuroscience, 17*(3), 392-406.

Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., & Cabanis, E. A. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain, 130*(5), 1432-1441.

Duffau, H. (2006). Brain plasticity: from pathophysiological mechanisms to therapeutic applications. *Journal of Clinical Neuroscience, 13*(9), 885-897.

Fisicaro, R., Jost, E., Shaw, K., Brennan, N.P., Peck, K.K., & Holodny, A.I. (2016). Cortical plasticity in the setting of brain tumors. *Top Magn Reson Imaging.* (1):25-30.

Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology, 23*(2), 250-254.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current opinion in neurobiology, 28*, 136-141.

Holodny, A. I., Schulder, M., Ybasco, A., & Liu, W. C. (2002). Translocation of Broca's area to the contralateral hemisphere as the result of the growth of a left inferior frontal glioma. *Journal of computer assisted tomography, 26*(6), 941-943.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in cognitive sciences*, *9*(11), 512-518.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E.B., & Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain, 123*(12), 2512-2518.

Krieg, S. M., Sollmann, N., Hauck, T., Ille, S., Foerschler, A., Meyer, B., & Ringel, F. (2013). Functional language shift to the right hemisphere in patients with language-eloquent brain tumors. *PLoS One*, *8*(9), e75403.

Peck, K. K., et al. (2009). Presurgical evaluation of language using functional magnetic resonance imaging in brain tumor patients with previous surgery. *Neurosurgery, 64*(4), 644-653.

Petrovich N.M., Holodny A.I., Brennan C.W., & Gutin P.H. (2004). Isolated translocation of Wernicke's area to the right hemisphere in a 62-year-man with a temporo-parietal glioma. *American Journal of Neuroradiology, 25*, 130–133.

Piai, V., Meyer, L., Dronkers, N. F., & Knight, R. T. (2017). Neuroplasticity of language in left-hemisphere stroke: Evidence linking subsecond electrophysiology and structural connections. *Human Brain Mapping*, *38*(6), 3151-3162.

Ruff, I. M. et al. (2008). Assessment of the language laterality index in patients with brain tumor using functional MR imaging: effects of thresholding, task selection, and prior surgery. *American journal of neuroradiology*, *29*(3), 528-535.

Shaw, K. et al. (2016). Infiltration of the basal ganglia by brain tumors is associated with the development of co-dominant language function on fMRI. *Brain and Language, 155*, 44-48.

Skipper-Kallal, L. M., Lacey, E. H., Xing, S., & Turkeltaub, P. E. (2017). Functional activation independently contributes to naming ability and relates to lesion site in post-stroke aphasia. *Human Brain Mapping*, *38*(4), 2051-2066.

Thal, D. J. et al. (1991). Early lexical development in children with focal brain injury. *Brain and language*, *40*(4), 491-527.

Tivarus, M. E., Starling, S. J., Newport, E. L., & Langfitt, J. T. (2012). Homotopic language reorganization in the right hemisphere after early left hemisphere injury. *Brain and language*, *123*(1), 1-10.

Turkeltaub, P. E. et al. (2012). The right hemisphere is not unitary in its role in aphasia recovery. *Cortex*, *48*(9), 1179-1186.

Turkeltaub, P. E. (2015). Brain stimulation and the role of the right hemisphere in aphasia recovery. *Current Neurology and Neuroscience Reports*, *15*(11), 72.

Turkeltaub, P. E., Messing, S., Norise, C., & Hamilton, R. H. (2011). Are networks for residual language function and recovery consistent across aphasic patients? *Neurology*, *76*(20), 1726-1734.

Vicari, S. et al. (2000). Plasticity and reorganization during language development in children with early brain injury. *Cortex*, *36*(1), 31-46.

CHAPTER 5

Distributional language learning: Entrenchment and plasticity

While language learning is complex, the SL literature suggests that learners are able to acquire language at least in part thanks to the regularities that exist within the distributional properties of the input (Christiansen, Allen, & Seidenberg, 1998; Monaghan & Christiansen, 2008). The three studies reported in the previous chapters elucidate several important characteristics of the human language system. I will discuss the two chapters on SL (Chapters 2 and 3) in terms of how they illustrate that the basic cognitive process of SL underlies language learning and processing. I will also detail the ways in which Chapter 4 interfaces with ideas about how the adult brain's language network may exhibit more plasticity than previously thought, and how this demonstrates support in favor of theories which claim that this network is not highly modular. Together, the results presented in this dissertation help shape our understanding of language acquisition - suggesting that language learning and processing is underpinned by cognitive and neural systems characterized by individual differences, entrenchment, and plasticity.

CHAPTER 5: CONCLUSION

Individual Differences in Statistical Learning and Language Processing

As anticipated, findings from this research provide further evidence for the claim that learners can draw on the distributional properties of the input during language acquisition (Saffran, 2001). Importantly though, this work also sheds light on the nature of this learning across learners, in terms of individual differences, and the reliability of the learning assessments we use at test. In addition to acting as a psychometric evaluation of commonly used AGL paradigms, Chapter 2 highlights the interaction between general cognitive processes and how they bias language learning and processing. This work comes in close contact with a rich tradition of investigations into how individual differences in cognitive processes, like SL, may be reflected in the linguistic domain, providing evidence for the link that exists between them.

Reber (1993) initially stated that due to the fundamentally ancient nature of implicit learning, it was unlikely that there would be profound individual variation in related abilities. While he has since reconsidered this claim (Reber & Allen, 2000), his initial hypothesis has had a great deal of influence on the field of SL research. However, recent evidence has pointed towards individual variation in SL abilities, while others have attempted to elucidate how these individual differences contribute to differences in language abilities (see Frost et al., 2015, for a discussion).

Shafto, Conway, Field, and Houston (2012) have provided developmental evidence for direct links between individual differences in SL and language abilities. Pre-linguistic infants aged eight-and-a-half months had their learning abilities evaluated on a visuo-spatial SL task, and then five months later were assessed for their early language skills. Early SL abilities were found to predict language development, as infants who were able to track the statistical relationships in the visual learning paradigm showed better language outcomes than those who did not. More longitudinal studies investigating the relationship between SL and language would greatly benefit our understanding of their relationship (Arciuli & Torkildsen, 2012).

Other individual differences studies with adult participants have demonstrated co-variation between SL and language abilities. One study found that individuals' performance on a visual SL task was correlated with performance on a task designed to test linguistic knowledge by querying whether or not they were able to decipher a predictable word in degraded auditory conditions (Conway, Bauernschmidt, Huang, & Pisoni, 2010). Individuals' SL scores have also been found to be a better predictor of language comprehension than performance on a verbal working memory task (Misyak & Christiansen, 2012). Another study in which implicit learning was identified as a distinct cognitive ability found it to be associated with verbal analogical reasoning (Kaufman et al., 2010).

A study by Misyak, Christiansen, and Tomblin (2010) found an association between SL ability and reading-time at the main verb in a sentence containing an object-relative clause (e.g., *the reporter that the senator attacked <u>admitted</u> the error*). Individuals who were better at learning the non-adjacent dependencies in the SL task also processed the long-distance dependency between the head noun and main verb more efficiently in a self-paced reading paradigm. Importantly, the better learners did not show significantly faster reading times when reading the main verb in subject-relative clauses (e.g., *the reporter that attacked the senator <u>admitted</u> the error*).

In a study that served as a direct inspiration for the research described in Chapter 2, a similar reading-time effect was found to exist for individuals who are more sensitive to a grammar relying on the learners' ability to track adjacent dependencies (Misyak & Christiansen, 2010). The better an individual was at learning the adjacent dependencies in a SL task the more interference they experienced when processing subject-verb number agreement with conflicting local information (e.g., *the key to the cabinets was rusty*). This suggests that such learners are hyper-sensitive to adjacent relations even when it was misleading, as all sentences of this type were grammatical. Of note is the fact that individual differences in adjacent and non-adjacent SL ability are not correlated with one another (Misyak & Christiansen, 2010). The finding reported in Chapter 2 that the ability of participants to chunk local information within the bi-/tri-gram familiarity judgment task in the Standard AGL paradigm was correlated with reading time differences when processing sentences with this same kind of potentially confusing adjacency information reflects a replication of this effect.

In combination with the findings reported in Chapter 2, the individual differences literature on SL clarifies the relationship between SL and language. Findings which demonstrate that SL abilities are related to language skill validate the idea that SL itself is a contributing factor to language learning and processing. The nuanced literature surrounding the relationship between SL and language contributes to the idea that this domain-general process plays an important role in language.

Following from this, individuals with greater experience tracking the types of relationships involved in processing sentences with non-adjacent dependencies should not only show higher performance on language tasks involving such dependencies, they should also show similar performance on tasks that rely on the same types of structure in other domains. This is consistent with other evidence pointing towards the effect that frequency has on processing (e.g., Reali & Christiansen, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). As individuals repeatedly track the same types of relationships in language, we would expect them to learn the underlying associations between elements that reduce uncertainty if they possess a mechanism for extracting such patterns. Wells et al. (2009) have shown that experience with the reading of relative clause sentences facilitates object-relative clause reading times in adults, demonstrating the importance of experience for language processing, and also providing compelling evidence for the plasticity of entrenchment throughout development. Learners track relationships between linguistic elements over the course of experience, and use the information in these relationships to continuously update their expectations and representations – SL abilities can be thought of as mediating the effect of linguistic experience.

Importantly, the finding that statistically learned information is retained over time, as described in Chapter 3, provides much-needed empirical support for the assumption that long-term effects of experience on processing can be accounted for by the process of SL. That learned associations can remain intact over time lends credence to the idea that experience can drive learning and processing biases within language users. Thus, even adults can become better at processing complex linguistic structures once those structures have become entrenched through experience-dependent learning mechanisms, indicating that it is a continuous, lifelong process of learning in language use (see Christiansen & Chater, 2016, for discussion).

In combination with the individual differences literature, the findings reported in Chapter 2 show that there is variation across individuals' capacity for detecting statistical regularities given their linguistic experience. These differences highlight the importance of SL in the entrenchment of linguistic structures, and linguistic relationships more generally; increased experience with certain structures leads to more automatic processing thereof. They also tie into the idea that language learning and processing are two sides of the same coin (Christiansen & Chater, 2016); processing input affects subsequent processing

CHAPTER 5: CONCLUSION

Chunking and Entrenchment Facilitate Language Learning and Processing

As shown in Chapter 3, constraining the input that participants are exposed to can change the way that they use and retain statistically learned information. Participants trained on increased amounts of simplified input showed improved signs of learning the artificial language to which they were exposed, and also demonstrated superior retention of the grammatical regularities embedded within that language. On the way to developing entrenched representations of learned associations, increased exposure to simplified input can provide a sort of scaffolding that facilitates learning in both the short-term and in the long-term. Giving learners more time and opportunity to "start small" (Elman, 1993) by providing extensive training for learners on simplified input allowed them to better acquire the grammatical regularities of the artificial language to which they were exposed (Lai & Poletiek, 2011; Poletiek, et al., in press). This reduced-complexity training set allowed participant to more efficiently and effectively learn the associations between items within the artificial language. Yet it was not that the training group exposed primarily to more complex input failed to learn anything about the input. Rather, they were able to use the surface-level similarities between items when choosing to endorse them, suggesting that they had begun to chunk the input in some way.

This links well with usage-based approaches to understanding language (e.g., Tomasello, 2003; Goldberg, 2003), which argue that grammatical knowledge is learned via the chunking/entrenchment of multi-word utterances, rather than relying on innate language-specific knowledge (e.g., Pinker, 1999). Language users have since been shown to rely on such chunks when processing language (see Arnon & Christiansen, 2017, for a review). In Chapter 3, we report strong evidence corroborating this claim, as participants relied heavily on the chunk strength of an item when judging its grammaticality. Other studies have reported how this reliance affects processing as well, for example, young children are able to repeat words in highly frequent non-idiomatic chunks more rapidly and accurately than when the same words form lower frequency chunks (Bannard & Matthews, 2008). Adults have also been found to have a processing advantage for high-frequency multiword chunks (Arnon & Snider, 2010; Janssen & Barber, 2012), an effect that is modulated by the meaningfulness of the utterance (Jolsvai, McCauley, & Christiansen, 2013). This set of findings indicates the importance of entrenchment to language processing and also highlights the importance of conventionalized form-meaning mappings, supporting construction grammar approaches to language (e.g., Goldberg, 2003). Language users seem to chunk multiple words together in ways that improve processing; these constructions are best understood as entrenched linguistic elements.

Taken together, this leads to the question of how the process of SL might aid in the construction of such chunks. Sensitivity to statistical relationships, like the backward transitional probabilities that infants as young as eight-months are capable of tracking (Pelucchi, Hay, & Saffran, 2009), has been built into certain models attempting to understand how children might form their early lexicon through the construction of these entrenched chunks. The peaks and dips in forward transitional probability have also been identified as potential cues for placing phrasal boundaries when computed over word classes (Thompson & Newport, 2007).

McCauley and Christiansen (2011; in press) have created a model which is capable of tracking the statistical relationships between single words and, based on these relationships, forming chunks. The model is trained on corpora of child-directed speech from the CHILDES database (MacWhinney, 2000), giving it a naturalistic input from which to learn. The model is able to accurately place boundaries between phrases, and also out-performs competing models when attempting to re-produce the utterances of the children in the corpora. In addition, the model parallels child performance in an AGL paradigm (Saffran, 2002) when the learning takes place over individual items, rather than classes of items, mirroring its relative performance in the analyses of language production and comprehension, contradicting the findings of Thompson and Newport (2007). This model demonstrates that entrenched units can be formed on the basis of distributional information alone, identifying SL as a mechanism of entrenchment in the contexts of both natural and artificial language.

In the context of the findings reported in Chapter 3, this makes it seem likely that extensive training with simplified input may have given those participants greater abilities to accurately process the relevant multiword sequences within test items. While participants in the complex condition were distracted by the chunk strength of items when performing grammaticality judgments, those in the simple training condition showed better command over how they incorporated such information into their judgments – accuracy was not correlated with item chunk strength, even while their endorsement rate was. Overall, learners' ability to acquire grammatical regularities and retain that information over time suggests that entrenchment and chunking could be part of the process of SL. A language system characterized by entrenchment, as argued here, requires some way to transfer learned associations into long-term memory and subsequently influence processing, a series of steps evidenced by participants in the aforementioned study.

Plasticity in the Brain's Language Network

The language system as described and illustrated throughout this dissertation is underpinned by a range of cognitive processes, including SL. As discussed in Chapter 4, this may be advantageous when the system is perturbed – if we did indeed possess a highly modular language organ (Chomsky, 1965, 1978; Fodor, 1983) that is responsible for all aspects of language learning and processing, then an injury to it would be necessarily catastrophic. Instead, it seems likely that our brain's language network is characterized by a level of interconnectedness with other networks that are involved in different kinds of processing (Aboitiz & Garcia, 1997). This idea in itself is not particularly new, dating at least back to Sapir (1921), who suggested that language relies on a network which evolved to subserve other functions.

While the findings reported in Chapter 4 do not specify the cognitive functions besides language that are embedded within this network, they do undercut the hypothesis that we possess a language organ, as initially proposed (Chomsky, 1965, 1978; Fodor, 1983). The results demonstrate that damage to the left inferior frontal gyrus, often pointed towards as one of the seats of language processing in the brain (Geschwind, 1970), does not necessarily lead to a loss of language function. Rather, the right-hemisphere homologue appears capable of taking over function when the left-frontal lobe is impacted by a tumor.

Past research had begun to demonstrate such plasticity in the developing brain (e.g., Thal et al., 1991; Vicari et al., 2000), yet there still exists a paucity of such evidence in adults. Showing that adults, too, can exhibit neuroplasticity within the brain's language network is an important step in characterizing the neural system that underpins language learning and processing. It shows that language does not rely upon a unique, highly specified neural circuit, as suggested by several important contributors to the field (Pinker & Bloom, 1990; Chomsky, 1965, 1978). Instead, we can see that parts of the brain not typically associated with language function able to maintain one's linguistic abilities.

CHAPTER 5: CONCLUSION

The findings presented in Chapter 4 portray a language network that is able to change and adapt over time, even through adulthood. Extrapolating beyond these findings, this could be thought of as providing indirect support for accounts of experience-dependent plasticity within said network. In order for our language system to change and adapt based on the input it receives (e.g., Reali & Christiansen, 2007; Wells et al., 2009), it would have to exhibit some degree of plasticity. While the results reported in Chapter 4 may not directly demonstrate this kind of plasticity, they do hint at the possibility that the language network is able to change over time, and that this ability reaches into adulthood.

Conclusion

In sum, the ability to track and learn probabilistic dependencies between elements seems to underlie human learning in multiple domains, including language. Whether the elements are tones (Saffran, Johnson, Aslin, & Newport, 1999), syllables (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996), word-like units (Gómez, 2002), visual sequences (Fiser & Aslin, 2002), or complex audiovisual stimuli (Mitchel, Christiansen, & Weiss, 2014; van den Bos, Christiansen, & Misyak, 2012), humans are able to learn about the statistical structure underlying their co-occurrence. This evidence points towards SL as a robust, domain-general process (Saffran & Thiessen, 2007) that is linked to language processing abilities, as shown in Chapter 2, and is likely implemented in separate modality-specific neural networks relying on similar computational principles (Frost et al., 2015).

The manner in which SL operates, by tracking relational and distributional information for items across space and time leads to the entrenchment of learned relationships and, crucially, Chapter 3 demonstrates that statistically learned information can indeed be retained over time. Additionally, the degree of entrenchment can vary between items as a function of frequency (Reali & Christiansen, 2007), meaningfulness (Jolsvai et al., 2013), and predictability (Aslin et al., 1998). Processing biases putatively influenced by entrenchment are also fundamentally plastic throughout the lifespan (Wells et al., 2009), a set of ideas that meshes well with the findings reported in Chapter 4.

When combined, this general understanding of how SL leads to the construction of units that contain meaning fits well into emergent, experience-based theories about language (i.e., Goldberg, 2003; Bybee, 2006; Elman et al., 1996; Christiansen & Chater, 2015), and identifies it as integral to theories postulating that language learning and processing rely on sensitivity to multiple cues in the input (Christiansen et al., 1998). Highly entrenched items can be stored as chunks, which can become the building blocks of language in development (McCauley & Christiansen, 2011), and which can also affect language processing (Bannard & Matthews, 2008). These entrenched representations are built up over the course of development as a result of SL, allowing higher level linguistic features to be learned.

References

Aboitiz, F., & Garcıa, R. (1997). The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Research Reviews*, *25*(3), 381-396.

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing Our Understanding of the Link between Statistical Learning and Language Acquisition: The Need for Longitudinal Data. *Frontiers in Psychology*, *3*(August), 324.

Arnon, I. & Christiansen, M.H. (2017). Multiword units as building blocks for language. *Topics in Cognitive Science*, *9*, 621-636.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological science*, *9*(4), 321-324.

Bannard, C., & Matthews, D. (2008). Stored Word Sequences in Language Learning The Effect of Familiarity on Children's Repetition of Four-Word Combinations. *Psychological Science*, *19*(3), 241–248.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711–733.

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

Chomsky, N. (1978). *Rules and Representations*. New York: Columbia University Press.

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.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The Behavioral and Brain Sciences*, *31*(5), 489–558.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *The Behavioral and Brain Sciences, 39,* e62.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356–371.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, *48*(1), 71–99. https://doi.org/10.1016/0010-0277(93)90058-4

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

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

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Geschwind, N. (1970). The organization of language and the brain. *Science, 170*(3961), 940-944.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224.

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

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PloS One*, *7*(3), e33202.

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2013). Meaning overrides frequency in idiomatic and compositional multiword chunks. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kaufman, S. B., Deyoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–40.

Lai, J., & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition*, *118*(2), 265–273. https://doi.org/10.1016/j.cognition.2010.11.011

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk, Vol. II: The database. Mahwah, NJ: LEA.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 1619-1624)*. Austin, TX: Cognitive Science Society.

McCauley, S.M. & Christiansen, M.H. (in press). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*. DOI: 10.1037/rev0000126

Misyak, J. B., & Christiansen, M. H. (2010). When ‘more’ in statistical learning means ‘less’ in language: individual differences in predictive processing of adjacent dependencies. In

Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 2686–2691).

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Language Learning*, *62*(1), 302–331.

Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*, *2*(1), 138–153.

Mitchel, A. D., Christiansen, M. H. & Weiss, D. J. (2014). Multimodal integration in statistical learning: Evidence from the McGurk illusion. *Frontiers in Psychology, 5,* 407. doi: 10.3389/fpsyg.2014.00407.

Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. *Trends in corpus research: Finding structure in data*, 139-63.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247.

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: HarperCollins.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707–784.

Poletiek, F., Conway, C.M., Ellefson, M.R. & Christiansen, M.H. (in press). Under what conditions can recursion be learned? Effects of starting small in artificial grammar learning of recursive structure. *Cognitive Science.* DOI: 10.1111/cogs.12685

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1–23.

Reber, A. S. (1993). Implicit learning and tacit knowledge: An essay on the cognitive unconscious. Oxford University Press: New York.

Reber, A. S. & Allen, R. (2000). Individual differences in implicit learning. In R. G. Kunzendorf & B. Wallace (Eds.), *Individual differences in conscious experience (pp. 227-248).* Philadelphia: John Benjamins.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*(4), 493-515.

Saffran, J. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, *47*(1), 172–196.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–8.

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. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.

Saffran, J. R. & Thiessen E. D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Handbook of language development* (pp. 68–86). Cambridge: Blackwell**.**

Sapir, E. (1921). *Language: An introduction to the study of speech.* New York: Harcourt, Brace and Company.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*(3), 247–271.

Thal, D. J., Marchman, V., Stiles, J., Aram, D., Trauner, D., Nass, R., & Bates, E. (1991). Early Lexical Development in Children with Focal Brain Injury, *527*, 491–527.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*(1), 1–42.

Tomasello, M. (2003). Constructing a language: A usage- based theory of language acquisition. Cambridge: Harvard University Press.

van den Bos, E., Christiansen, M. H., & Misyak, J. B. (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, *67*(4), 507–520.

Vicari, S., Albertoni, A., Chilosi, A. M., Cipriani, P., Cioni, G., & Bates, E. (2000). Plasticity and reorganization during language development in children with early brain injury. *Cortex*, *36*, 31–46.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271.