Adaptation as Statistical Learning: An Individual Differences Study

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Kelly Enochson
Master of Arts
George Mason University, 2007
Bachelor of Arts
University of Virginia, 2003

Director: Steven Weinberger, Associate Professor
Department of English, Program in Linguistics

Summer Semester 2015
George Mason University
Fairfax, VA
ACKNOWLEDGEMENTS

I would like to thank my husband Stephen for never-ending moral support and programming tips. Many, many thanks to Jenny Culbertson for all the time and effort she put into working with me. Thanks to Steven Weinberger for encouragement, advice, and making my graduate school experience run smoothly. I would like to thank Charlie Jones and Tuuli Morrill for helping me to flesh out my ideas and make sense of my arguments. I would also like to thank my fellow PhD students for constant support and assistance. Finally, thanks to Bill Badecker, Florian Jaeger, the audiences at CUNY 2014 and CUNY 2015, and anonymous reviewers for helpful comments on the experiments reported here. This dissertation was supported in part by NSF doctoral dissertation research improvement grant #1451652.
# TABLE OF CONTENTS

| LIST OF TABLES | vii |
| LIST OF FIGURES | viii |
| ABSTRACT | ix |

## CHAPTER 1: STATISTICAL LEARNING AND ADAPTATION

1. Introduction. ................................................................. 1
2. Statistical learning and language acquisition.............................. 2
3. Linguistic adaptation. .......................................................... 5
4. Individual differences in statistical learning as a predictor of language processing. ................................................................. 12
5. Integrating syntactic adaptation and statistical learning in an individual differences framework................................................................. 16

## CHAPTER 2: USING WEB-BASED METHODS FOR PSYCHOLINGUISTIC RESEARCH

1. Introduction. ................................................................. 17
2. Three classic effects in psycholinguistics.................................. 17
   2.1. Pronoun vs. DP processing...................................................... 17
   2.2. Filler gap dependency .......................................................... 18
   2.3. Agreement Attraction .......................................................... 19
3. Experiment 1: Replication using AMT ...................................... 19
   3.1. Method ................................................................................. 21
   3.2. Results ................................................................................. 24
4. Experiment 2: further investigations of agreement attraction ........... 31
   4.1. Method ................................................................................. 31
   4.2. Results and Discussion ......................................................... 32
5.4. Conclusion.................................................................................................................. 88
References......................................................................................................................... 90
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1: Test item schema (( t ) indicates the gap position)</td>
<td>20</td>
</tr>
<tr>
<td>Table 2: Example stimuli set, Experiment 1</td>
<td>23</td>
</tr>
<tr>
<td>Table 3: Regions of interest</td>
<td>25</td>
</tr>
<tr>
<td>Table 4: Estimates and (p-values) for agreement attraction models</td>
<td>29</td>
</tr>
<tr>
<td>Table 5: Example stimuli set, Experiment 2</td>
<td>32</td>
</tr>
<tr>
<td>Table 6: Example stimuli set, Experiment 3</td>
<td>36</td>
</tr>
<tr>
<td>Table 7: Sample stimulus set for the syntactic adaptation task</td>
<td>52</td>
</tr>
<tr>
<td>Table 8: T statistics and p-values for each factor in the simplified regression model</td>
<td>75</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1: A sample stimulus presentation for the item <em>rud balip pel.</em></td>
<td>15</td>
</tr>
<tr>
<td>Figure 2: Sample self-paced reading trial.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 3: Subject definiteness results. Error bars represent standard error of the mean.</td>
<td>27</td>
</tr>
<tr>
<td>Figure 4: Filler-gap results. Error bars represent error of the mean.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 5: Agreement attraction with PP modifiers. Error bars represent standard error of the mean.</td>
<td>34</td>
</tr>
<tr>
<td>Figure 6: Agreement attraction with RC modifiers. Error bars represent standard error of the mean.</td>
<td>38</td>
</tr>
<tr>
<td>Figure 7: Adaptation results, Block 2</td>
<td>42</td>
</tr>
<tr>
<td>Figure 8: Adaptation results, Block 3</td>
<td>43</td>
</tr>
<tr>
<td>Figure 9: A sample array for the trial <em>rud balip pel.</em></td>
<td>49</td>
</tr>
<tr>
<td>Figure 10: Results from participants who learned the dependency as measured by a score of 8 or above on the grammaticality judgment task.</td>
<td>51</td>
</tr>
<tr>
<td>Figure 11: Residual reading times for ambiguous and unambiguous sentences for the entire stimulus set.</td>
<td>54</td>
</tr>
<tr>
<td>Figure 12: Correlation between performance on the syntactic adaptation task and the statistical learning (grammaticality judgment) task.</td>
<td>60</td>
</tr>
<tr>
<td>Figure 13: Residual reading times for ambiguous and unambiguous items across the entire stimulus set.</td>
<td>67</td>
</tr>
<tr>
<td>Figure 14: Mean response times by block for the phonetic adaptation task. Error bars represent standard error of the mean.</td>
<td>70</td>
</tr>
<tr>
<td>Figure 15: Phonetic adaptation accuracy scores by block.</td>
<td>71</td>
</tr>
<tr>
<td>Figure 16: Correlation between performance on the phonetic adaptation task and the statistical learning task.</td>
<td>75</td>
</tr>
<tr>
<td>Figure 17: Correlation between performance on the syntactic adaptation task (at the final region) and the grammaticality judgment task.</td>
<td>76</td>
</tr>
</tbody>
</table>
ABSTRACT

ADAPTATION AS STATISTICAL LEARNING: AN INDIVIDUAL DIFFERENCES STUDY

Kelly Enochson, Ph.D.
George Mason University, 2015
Dissertation Director: Dr. Steven Weinberger

A rich body of research has shown that language learners can track and use distributional information in the input to acquire multiple levels of linguistic structure (see Krogh et al., 2013 for a review). There is reason to believe, however, that learning about the statistics of the input is not confined to language acquisition, but is part of ongoing language experience. In particular, language processing appears to be influenced by expectations—e.g., about probable sounds, words, structures—which are dynamic and can be rapidly updated based on the current linguistic environment (e.g., Fine et al., 2013). If a general mechanism like statistical learning underlies both acquisition and later processing and use, a clear prediction is made: performance on an independent measure of statistical learning should correlate with ability to adapt native language expectations based on novel information.

This dissertation reports three sets of experiments designed to address this prediction. The first set of experiments serves as proof of concept, demonstrating that...
response time data can be collected accurately and efficiently over the web using crowd-sourcing services like Amazon Mechanical Turk. The second set of experiments explores the relationship between statistical learning ability and syntactic adaptation, showing that the two processes co-vary within an individual. The third set of experiments tests whether statistical learning also underlies phonetic adaptation, addressing the novel prediction that syntactic adaptation and phonetic adaptation are driven by the same underlying cognitive mechanism.
CHAPTER 1: STATISTICAL LEARNING AND ADAPTATION

1.1. Introduction.

The task of language comprehension involves receiving noisy and variable input, and aligning perception of this input with the intended message from the speaker. This task is accomplished through the integration of a number of sources of information, including grammatical knowledge and distributional cues such as the frequencies of linguistic structures. From this information, a set of expectations is derived, which influences how a given message is processed and ultimately understood. Recent research suggests that these expectations are dynamic: language users can rapidly update them by tracking information about the current linguistic environment (Fine, Jaeger, Farmer, & Qian, 2013; Jaeger & Snider, 2013; Kleinschmidt & Jaeger, 2012). In other words, learning about the statistics of the input is not confined to language acquisition (Saffran, Newport, & Aslin, 1996), but is part of the ongoing language experience. Nevertheless there is evidence that statistical learning ability differs across individuals (Conway, Bauernschmidt, Huang, & Pisoni, 2010; Kidd, 2012; Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010). Thus if this ability underlies expectation adaptation, a clear prediction of this view of language use is that the two should be correlated. This dissertation will test this prediction by examining whether independent
measures of statistical learning ability relate to individuals’ ability to adapt to their linguistic environment.

This dissertation proceeds as follows: I begin with a brief literature review of statistical learning, adaptation, and individual differences in language processing, and demonstrate how I will merge these lines of research in my dissertation. In Chapter 2 I discuss the results of a set of studies functioning as proofs of concept of various aspects of my dissertation methodology. Chapter 3 explores the relationship between statistical learning and syntactic adaptation, with results showing that individual differences in statistical learning predict syntactic adaptation. In Chapter 4 I apply the same prediction to a different linguistic domain, testing whether statistical learning also underlies phonetic adaptation, with results again indicating that performance on one predicts performance on the other. Chapter 5 discusses the implications of these findings and possibilities for future research.

1.2. Statistical learning and language acquisition.

Language learners are able to track statistical information available in the input at multiple levels. Phonetic distributions of different sounds can be used to determine phoneme categories (Maye, Werker, & Gerken, 2002; McMurray, Aslin, & Toscano, 2009). Conditional probabilities between syllables in a continuous speech stream can be used to bootstrap into speech segmentation (Aslin, Saffran, & Newport, 1998; Pelucchi, Hay, & Saffran, 2009; Saffran, Newport, & Aslin, 1996; Saffran, 2003; but c.f. Yang, 2004). Low-level statistics such as word and collocation frequency (Biber et al., 1999; Bybee & Hopper, 2001; Roland, Dick, & Elman, 2007; Smith & Levy, 2008), can be
used to feed acquisition of higher-level phrase structure (Mintz, Newport, & Bever, 2002; Saffran, 2001; Thompson & Newport, 2007, but c.f. Lignos, 2013). Interestingly, not all statistical information is created equal; there is evidence for example that adjacent statistical dependencies are easier to track compared to non-adjacent dependencies (Newport & Aslin, 2004). Nevertheless, under certain conditions, transitional probabilities of non-adjacent dependencies can be learned. Peña, Bonatti, Nespor, and Mehler (2002) found that non-adjacent dependencies were learned when they were cued by phonological regularities. Gómez (2002) found successful learning when highly variable material intervened between adjacent elements.

Discussions of statistical learning are largely couched in terms of acquisition, suggesting that while infants make ready use of distributional information, once the target structure is acquired, learning does not necessarily continue. Recent research, however, probes the relationship between statistical information and language processing, proposing that statistical learning is not limited to acquisition, but actively used by adults to process language (see discussion below). This dissertation seeks to add to this body of work by investigating the extent to which individual differences in language processing in adults relate to differences in the ability to actively track statistical information.

While early research into statistical learning of language has shown that adults are able to track statistical information such as transitional probability (Saffran et al., 1996), only recently have researchers begun to explore the relationship between this ability and other language-related abilities such as syntactic processing and adaptation. Chang, Janciauskas, and Fitz (2012) trained a simple recurrent network using an error-driven
implicit learning mechanism to predict the next letter in a sequence based on previous experience. They demonstrated that this learning model parallels human syntactic priming behavior, suggesting that syntactic priming (or adaptation; see section 1.2 for an overview) is a result of an error-driven implicit learning mechanism.

If the mechanisms involved in statistical learning support error-driven implicit learning behavior, then performance in a statistical learning task, such as non-adjacent dependency learning, should relate to performance on a more directly language-based processing task. To address this relationship, Conway, Bauernschmidt, Huang, and Pisoni (2010) performed two experiments designed to explore whether people proficient at using sentence context to aid speech perception are also superior implicit learners. Using both visual and auditory serial response time tasks, as well as auditory and audiovisual speech perception tasks, Conway et al. (2010) demonstrated a correlation between implicit learning ability and speech perception. They suggest a direct causal relationship between the two skills, proposing that “superior implicit learning ability results in a more detailed and robust representation of word order probabilities in spoken language” (p.365).

In the domain of syntactic processing, Wells, Christiansen, Race, Acheson, and MacDonald (2009) advocate an explanation of the well-attested subject relative clause and object relative clause processing asymmetry based on experience (i.e., the interaction between frequency, regularity, and experience). To test this hypothesis, they expose one group of participants to both subject relatives (SRs) and object relatives (ORs)—known to be processed slower and comprehended less well than subject relatives (SRs)—while exposing the other group to the task procedure, but with no relative clause input. Through
exposure, they were able to almost entirely mitigate the additional processing cost for ORs as compared to SRs, suggesting that language processing may be directly related to statistical learning. To address this possibility more precisely, Misyak & Christiansen (2012) administer a variety of individual-differences tasks to a set of participants who also participate in a self-paced reading task measuring processing of SRs and ORs. They find that a statistical learning task measuring non-adjacent dependency learning correlates with better performance on OR processing, while none of the other measures, including lexical knowledge, reading experience, verbal working memory, short term memory, fluid intelligence, and cognitive motivation, correlated with relative clause processing.

Addressing a related question, Kidd (2012) proposes that statistical learning ability relates to persistence of syntactic priming, which he tests using a serial response time task as a measure of implicit learning and a production task designed to elicit syntactic priming effects. Results show that performance on the implicit learning task predicts persistence of syntactic priming effects, while performance on tasks measuring other cognitive functions such as IQ, explicit learning, and verbal ability did not. Kidd (2012) suggests that implicit statistical learning is a domain-general cognitive function that will correlate with any task that requires using experience to alter predictions of upcoming forms. I return to the relationship between statistical learning ability and linguistic behavior in section 1.3, but turn now to a discussion of linguistic adaptation, including syntactic priming.

1.3. Linguistic adaptation.
Encountering variability is an inherent part of perceiving and processing language, raising the question of how language processing systems address the challenges associated with variable linguistic input. There is evidence that processing is aided by expectations about upcoming linguistic structures—sounds, words, sentence structures, etc.—generated based on the statistics of previous linguistic experience (Fine et al., 2013). These expectations are optimal under a cost function that balances preparation cost and processing cost (Smith & Levy, 2008). Put another way, it is optimal to prepare for a frequent structure rather than for an infrequent one, as the likelihood of incurring a processing cost in addition to preparation cost is low. However, linguistic environments differ in their statistical distributions as individuals differ in their language use, necessitating that comprehenders be able to rapidly update their expectations with each novel environment.

Adaptation is referred to by various labels including syntactic or structural priming (in the syntactic domain), accommodation (in phonetic production), entrainment, and alignment. In this dissertation, I will refer to all these processes as adaptation, although I recognize that there are some differences between them. Adaptation refers to the ability to more efficiently produce or process a linguistic form based on its similarity to a recently-experienced form (Bock, 1986). In the domain of phonology, in production this typically presents as two talkers’ acoustic characteristics converging, so that the individual speakers come to be more similar-sounding (Babel, 2009). Laboratory studies indicate that phonetic accommodation can occur when speakers are producing only single words (Namy, Nygaard, & Sauerteig, 2002) as well as under cooperative and social
conditions (Pardo, 2006). Speakers accommodate various aspects of their phonology including vowel quality (Babel, 2009, 2012), artificially-lengthened VOT (Shockley, Sabadini, & Fowler, 2004), formant frequency (Babel & Bulatov, 2012) and vowel category (Maye, Aslin, & Tanenhaus, 2008).

Phonetic accommodation in comprehension is typically termed perceptual learning. Studies in perceptual learning have demonstrated that participants rapidly adjust their phonetic categories in response to a new distribution of VOT (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kraljic & Samuel, 2006), vowel category (Escudero & Williams, 2014; Maye et al., 2008), or to an entirely novel accent (Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004).

Clayards et al. (2008) found that given input in which VOT is not a reliable cue to voicing, participants learn to ignore unreliable phonetic cues, and consequently accept, for example, highly p-like tokens as instances of b. At least under some condition, such adjusted perceptual category information is generalized across different speakers and test continua (e.g., Kraljic & Samuel, 2006). To explore the role of distributional information in phonetic adaptation, Kleinschmidt & Jaeger (2012) exposed participants to prototypical instances of /b/, and showed that they subsequently identified fewer tokens along the /d/ to /b/ continuum as being instances of /b/. On the other hand, after exposure to ambiguous instances of /b/, participants identified more tokens along the /d/ to /b/ continuum as being instances of /b/. They argue that these results derive from adaptation to statistics of the current input, suggesting both syntactic adaptation (priming) and phonetic adaptation (accommodation) may result from dynamic expectation updating.
Based on the premise that foreign-accented speech contains phonological regularities that a listener can exploit, Clarke & Garrett (2004) suggest that tracking these regularities could lead to more efficient decoding (and subsequent processing) of accented speech. They demonstrate that over a very short exposure period—less than one minute—processing speed of foreign-accented speech improves to the point that it is comparable to that of “unaccented” speech. Participants in the test condition listen to 16 English sentences, in which the final word cannot be predicted from context, spoken by a native speaker of Spanish. After each sentence a word appears on the screen, and participants must decide whether or not the word on the screen is the final word of the previous sentence. Reaction times in the final block of 4 sentences were comparable to those of the control group, who heard a native English speaker reading the sentences, suggesting that listeners can very rapidly adapt the perceptual system to novel acoustic-phonetic patterns.

Turning to the domain of syntax, in one of the first studies to directly probe syntactic priming, Weiner and Labov (1983) found that exposure to a sentence in the passive voice strongly predicts subsequent use of the passive voice in interviews. Shortly after this, in a seminal work, Bock (1986) demonstrated that after reading prime sentences, participants show priming effects for both active and passive voice as well as for direct objects and prepositional objects with verbs participating in the dative alternation. Priming occurs regardless of whether specific lexical items (Bock, 1989; Fox Tree & Meijer, 1999) or morphology (Ferreira, 2003) are repeated, and irrespective of register (Bock, 1986). Additionally, priming is not limited to production and occurs when
participants simply read, instead of speaking, the prime (Bock, Dell, Chang, & Onishi, 2007; Potter & Lombardi, 1998) as well as in exclusively comprehension-based studies (Thothathiri & Snedeker, 2008).

Two accounts of syntactic priming that have gained widespread support are activation-boost accounts and error-driven learning accounts. Activation-boost accounts (e.g., Pickering & Branigan, 1998) attribute syntactic priming solely to a short-term boost in activation of recently processed material, and do not predict that frequency of words or structures impacts level of priming. On the other hand, error-driven learning accounts suggest that priming occurs as the result of updating expectations about the distribution of syntactic structures to better match the distribution of the input, with a natural consequence of this being greater priming effects for less frequent structures than more frequent ones. Jaeger and Snider (2013) test this prediction by reanalyzing production data from Thothathiri & Snedeker (2008) and comparing results to norming studies of their own. Jaeger & Snider (2013) gathered sentence completion preference information for a variety of verbs which participate in the dative alternation using gated sentence completion, and compared this distributional information to the strength of priming found in Thothathiri and Snedeker’s data. They found that less frequent structures result in stronger priming effects, suggesting that priming is in fact a form of implicit error-driven learning and a result of expectation adaptation rather than short-term activation of recently processed structures.

Given that the magnitude of adaptation effects depends on the predictability of the linguistic forms involved, structures that are temporarily ambiguous provide a rich
environment for testing hypotheses related to adaptation. Fine et al. (2013) explore how people use statistical information that differs significantly from typical distributions to update their expectations about the likelihood of a particular syntactic structure in a novel linguistic environment. For instance, sentences very commonly follow the form subject – main verb – object, but are certainly not limited to this form. It is possible for the subject to be followed by some kind of modifier, such as a relative clause, which can then be reduced, resulting in a verb which is not the main verb immediately following the subject. In this type of construction, comprehenders encounter a temporary ambiguity at the relative clause, which is mitigated at the main verb.

For example, consider the sentences in (1).

1. a) The experienced soldiers spoke about the dangers before the midnight raid.
   b) The experienced soldiers who were told about the dangers conducted the midnight raid.
   c) The experienced soldiers [warned about the dangers] before the midnight raid.
   d) The experienced soldiers [warned about the dangers] conducted the midnight raid.

Given the preamble *The experienced soldiers*..., there are several possible sentence continuations. The most likely is that *soldiers* will be followed by a main verb, as in (1a) and (1c). The verb *spoke* in (1a) is unambiguously a main verb, since it cannot function in another capacity here. Another possibility is that the preamble will be followed by a relative clause, as in (1b) and (1d). In (1b), the continuation is unambiguously a relative clause, since it includes the overt relative pronoun *who*. However, the continuations
beginning with the word *warned* in (1c) and (1d) are potentially ambiguous; *warned* could be a main verb, as in (1c), or it could be introducing a reduced relative clause, as in (1d). Again, (1c) is more likely, given the relative frequencies of the structures. Conditions (1c) and (1d) are ambiguous during the bracketed region, and are not disambiguated until the underlined word.

Fine et al. (2013) manipulated the distribution of syntactic structures in the current linguistic environment by using stimuli such as those above in (1) instantiated in a self-paced reading task to probe the time course of adaptation effects. Participants were assigned to one of two conditions: relative clause (RC) first or filler first. The RC first group read 16 relative clauses, 8 of which were ambiguous, in block 1, while the filler first group read 16 filler items in block 1 and, crucially, no relative clauses. In blocks 2 and 3 both groups were exposed to the same stimuli: in block 2, both groups read 10 relative clauses, 5 of which were ambiguous, and some filler items; in block 3, both groups read 10 main verb constructions, 5 of which were ambiguous, and some filler items. In typical linguistic environments, main verb continuations are much more likely than relative clause continuations for sentences such as those presented in (1). However, this task specifically manipulated the distribution that participants encounter, exposing the RC first group to drastically different distributional information in block 1 than what is typical. If participants are able to use the distributional information available in this novel linguistic environment to adapt their expectations, then this predicts that in block 2, the RC first group should be faster at reading ambiguous relative clause items than the filler first group based on their immediately previous experience with such structures.
This also predicts that in block 3, the filler first group should be faster at reading ambiguous main verb items than the RC first group, because at this point the RC first group should experience a greater degree of surprisal when encountering a main verb than the filler first group because of their experience with this particular task, despite the fact that main verbs are more frequent in general.

The results from Fine et al. (2013) support these two predictions. The RC first group was faster at reading ambiguous items in block 2, but slower at reading ambiguous items in block 3. This suggests that language comprehenders are able to use new distributional information to adapt their expectations about upcoming linguistic structures, and that they are able to do this rapidly, over the course of 26 relative clauses in a 15-minute experiment.

1.4. Individual differences in statistical learning as a predictor of language processing.

Individual differences in language development and language processing are substantial and pervasive. The debate surrounding the underlying cause of these individual differences typically centers around two dominant accounts: capacity-based accounts and experience-based accounts. Capacity-based accounts attribute individual differences to variability in cognitive or psychological resources, such as differences in working memory capacity (Just & Carpenter, 1992; Waters & Caplan, 1996). Experience-based accounts instead appeal to differences in experience with particular linguistic structures to account for language-based intra-individual variation (MacDonald & Christiansen, 2002; Wells et al., 2009). As suggested by error-driven implicit learning
models, statistical learning is a means of learning *from experience*, while language processing may be directly related to experience as well (Wells et al., 2009); therefore, individual differences in statistical learning should relate to variation in language processing.

To further address this prediction, Misyak, Christiansen, and Tomblin (2010) explore whether statistical learning and syntactic processing involve the same cognitive mechanisms. They approach this question via two tasks: a serial response time task, which measures an individual’s statistical learning ability, and a self-paced reading task, which measures syntactic processing. Statistical learning experiments typically rely on a training block in which participants are exposed to a miniature artificial language, and a testing block in which participants must discriminate items that are grammatical in the artificial language from items that are not. Misyak, et al. present the artificial grammar learning experiment in a serial reaction time (SRT) task as a novel paradigm for measuring an individual’s online statistical learning ability. If syntactic processing is related to statistical learning ability, then performance on the SRT task should correlate with performance on the syntactic processing task. Since the methodology used in this dissertation is based on work done in this study, I will elaborate here about their methods and results.

The statistical learning measure came from an artificial grammar learning task of non-adjacent dependencies. It was instantiated in an SRT task using the stimuli from Gómez (2002). Non-adjacent dependencies exist when two related linguistic elements co-occur, with intervening material separating the dependent elements. For instance, the
dependency between $a$ and $b$ in the sample $a\times b$ represents a non-adjacent dependency. The English formation of the progressive aspect exemplifies this structure, requiring the auxiliary verb “be” and the suffix “-ing” together to denote the progressive. While intervening between these two elements is a main verb unrelated to the progressive aspect dependency. In order to acquire the dependency a learner must disregard the highly variable intervening material, and instead track the relationship between the flanking elements. Non-adjacent dependency learning has been successful in studies where the intervening material is highly variable, as in this case where it consists of 24 elements, but performance is impeded when the intervening material is less variable (Gômez, 2002).

Misyak et al. (2010) measured performance in an online SRT task, which presents auditory stimuli accompanied by a corresponding visual representation. A computer screen was divided into six boxes: three columns and two rows, with each column corresponding to a portion of the stimulus and each row presenting two alternative possible stimuli. The participant clicked on each box that corresponds with part of the stimulus, for a total of three clicks per trial, one for each element in $a\times b$ as shown in Figure 1.
Figure 1: A sample stimulus presentation for the item *rud balip pel*.

Initially, click times for the final column should be as slow as or slower than click times for the first column; as the experiment progresses, if the participant acquires the dependency, click times on the third column will speed up as the participant anticipates the third element based on its dependency with the first. Misyak et al. (2010) used a SRT task as an independent measure of statistical learning ability, similar to how researchers often use a reading span task as an independent measure of verbal working memory.

In addition to the statistical learning task, Misyak et al. (2010) used the same participants for a syntactic processing task, which measures online comprehension of subject relative (SR) and object relative (OR) clauses, instantiated in a self-paced reading task. As mentioned above, a well-established asymmetry exists between these two structures, with SRs being more common and easier to process than ORs. Results indicated that, for all participants, SRs were processed faster than ORs, which is consistent with a robust body of psycholinguistic research. However, the difference in processing cost of the two structures varied by individual; for some participants the difference was much smaller than for others. Similarly, on the artificial grammar learning
SRT task, some participants were able to learn the non-adjacent dependency, while some were not. Misyak et al. (2010) divided the participants based on whether or not they learned the dependency in the artificial grammar; they found that participants who learned the non-adjacent dependency had a significantly smaller difference in SR/OR processing than participants who did not learn the dependency, suggesting that statistical learning and syntactic processing do seem to involve the same cognitive mechanisms.

1.5. Integrating syntactic adaptation and statistical learning in an individual differences framework.

In this dissertation I aim to merge these two lines of research related to individual differences in statistical learning and rapid expectation adaptation. I will explore whether individuals who are more successful at statistical learning (as opposed to learning by, for example, by parameter resetting, which is not statistical, or effects of a memory boost, which is not learning) are also better able to adapt their expectations—either more quickly or more accurately—based on distributional information in a novel linguistic environment. Adaptation occurs most strongly in instances of novel or infrequent structure, suggesting that it is error-driven, and consequently the outcome of an implicit learning mechanism. Statistical learning is such a mechanism, as it allows error-driven learning to proceed accurately given each novel linguistic context. Individuals with superior statistical learning ability—those who more accurately generate expectations about upcoming structure based on distributional information—should be able to more rapidly adapt to a new linguistic environment.
CHAPTER 2: USING WEB-BASED METHODS FOR PSYCHOLINGUISTIC RESEARCH

2.1. Introduction.

Typical psycholinguistic research in the lab requires substantial resources, such as access to a sizeable population of undergraduate students, infrastructure for recruiting, and means of compensation, making completing experiments expensive and time consuming, and prompting researchers in psychology and cognitive science to turn to online services as an inexpensive and fast alternative (Crump, McDonnell, & Gureckis, 2013). MTurk is one such service: an online crowd-sourcing marketplace that allows researchers to post experiments, and automatically recruit and compensate participants from disparate demographic groups. In order to demonstrate proof of concept for further psycholinguistic research methods related to this project, this section shows that MTurk provides a promising framework for research in psycholinguistics, including when accurate response time measurements are crucial.

2.2. Three classic effects in psycholinguistics

2.2.1. Pronoun vs. DP processing

A number of psycholinguistic studies have shown that referring expressions which differ in definiteness are processed or judged differently. This has been claimed to be the result of relative ease of accessing a referring expression in memory
through factors like distance from last mention, number of competing referents, and notions of topicality or givenness (Ariel, 1990; Gundel, Hedberg, & Zacharski, 1993). Pronouns are relatively more accessible compared to full DPs (Gordon, Grosz, & Gilliom, 1993), and indeed self-paced reading experiments have revealed faster reading of pronouns as in (2b) compared to definite DPs as in (2a) (Warren & Gibson, 2002).

2. a) George never thinks about how others will feel.

b) He never thinks about how others will feel.

While we target this particular distinction, similar differences can be seen in processing definite compared to indefinite DPs (Hofmeister & Sag, 2010).

2.2.2. Filler gap dependency

Another well-known effect in psycholinguistics concerns so-called filler-gap constructions, which feature a grammatical dependency between a fronted element (the filler) and its original syntactic position (the gap). The thematic role of the filler is precisely the one which would have been assigned to the element in the position of the gap. For example, in (3), there is a gap after the main verb find that would be filled by a DP. Which cars is the filler associated with the gap.

3. Which cars did the salesman find _______ easiest to sell?

Filler-gap dependencies are claimed to increase processing load because the filler must be held in working memory until the gap is identified, while all the other information encountered must be processed simultaneously (Hawkins, 1999). In a landmark study, Wanner & Marastos (Wanner & Marastos, 1978) used a combined comprehension and memory task to test the effects of distance between filler and gap,
finding that as distance increases, comprehension decreases. Subsequent studies using self-paced reading tasks have found that in a number of different syntactic constructions, reading times slow precisely where a gap would occur (Crain & Fodor, 1985; Stowe, 1986).

2.2.3. Agreement Attraction

The phenomenon of agreement attraction describes an instance in which, instead of agreeing in number with its grammatical subject, a verb spuriously agrees with some nearby constituent, as in (4a–b) (Kimball & Aissen, 1971). A robust finding in the literature is that agreement attraction is much more likely to occur with a nearby attractor that is plural (Bock & Miller, 1991). In other words, sentences like (4a) are much more common than (4c).

   4. a. The key to the cabinets are missing.

   b. The people who Clark think are in the garden…

   c. The keys to the cabinet is missing.

Agreement attraction has been attested in a number of distinct construction types, including with prepositional phrase modifiers (Bock & Miller, 1991; Pearlmutter, Garnsey, & Bock, 1999), relative clauses (Bock & Miller, 1991; Wagers, Lau, & Phillips, 2009), auxiliary inversion (Vigliocco & Nicol, 1998), and wh-fronted constructions (Badecker, in prep).

2.3. Experiment 1: Replication using AMT

In order to investigate whether pronoun/DP processing effects, filler-gap effects, and agreement attraction can be replicated over the web using AMT, we used stimuli
comparable in structure to those in Badecker (in prep) which make it possible to test all three effects in a single self-paced reading experiment. This experiment was presented to participants using ScriptingRT, as described in more detail below. Self-paced reading tasks typically reveal an increase in reading time at a structure that is difficult to process (King & Just, 1991), for example one which is ungrammatical. Following Badecker (in prep) we used grammatical stimuli, with all critical items involving a fronted object *wh*-phrase with an associated gap in post-verbal position, and a DP or pronominal subject. Because attraction effects are expected to occur more often with plural DPs, the number of the subject and *wh*-attractor were also manipulated. These stimuli are schematized in Table 1.

Table 1: Test item schema (*t* indicates the gap position)

<table>
<thead>
<tr>
<th>Subject, Verb</th>
<th><em>wh</em>-Attractor</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG.</td>
<td>Which X has he/DP verbed <em>t</em>&lt;sub&gt;wh&lt;/sub&gt;?</td>
</tr>
<tr>
<td>PL.</td>
<td>Which X have they/DP verbed <em>t</em>&lt;sub&gt;wh&lt;/sub&gt;?</td>
</tr>
</tbody>
</table>

We make the following general predictions. Reading-times will be slower for DPs compared to pronouns, and a slow-down in reading times will occur at or just following a gap. As for attraction effects, these occur in ungrammatical sentences when participants fail to detect erroneous agreement between a verb and an intervening non-subject DP
(attractor). However, in grammatical sentences they occur when participants falsely identify as ungrammatical a sentence in which the verb does not agree with an attractor. Therefore here agreement attraction effects are predicted to manifest as a slow-down in reading after encountering an attractor DP followed by a verb with agreement that does not match it (Pearlmutter et al., 1999).

2.3.1. Method

Participants

Participants were workers recruited through Amazon Mechanical Turk (AMT). In total, 35 external survey human intelligence tasks (HITs) were posted. Because of a software problem, one participant received an incomplete test list, therefore data analyzed here come from 34 AMT workers. This sample size is comparable to that used in other studies investigating these same effects (Badecker, in prep; Hofmeister & Sag, 2010; Warren & Gibson, 2002). Two features of the HIT were designed to recruit only participants who were native speakers of English: first we included a locale qualification specifying that workers be located in the United States. Second, the informed consent document specified that workers must be native English speakers, have no known language disorders, and be above 18 years of age. Recruiting participants in this way gives up some control (relative to lab-based studies), as confirmation of all these factors is based on self-reporting. However, willfully misrepresenting oneself to complete an AMT HIT is a violation of the worker’s terms of service. Workers were compensated $1.00 for participation in the study, which took approximately 20 minutes. For the sake of comparison with lab-based studies, we were able to recruit and gather data from 35
participants in 7 hours, for a total of $35.00 in payment to participants and $10.50 to AMT ($0.30 per participant).

**Apparatus, Stimuli & Design**

The experiment was presented to workers as a Flash movie embedded in an HTML page, as shown in Figure 2. A link to the wrapper HTML page was posted on AMT as an external survey HIT. Flash is currently a very popular solution to providing dynamic content over the web, and Flash plug-ins are either built into or available for most popular browser applications (Chrome, Firefox, Safari, Internet Explorer). By using Flash to capture response times on the client side, we hope to capture psycholinguistic processing effects with accuracy comparable to that of laboratory software such as Linger (Rohde, 2003).

Figure 2: Sample self-paced reading trial.
The critical stimuli used in the task consisted of 48 sentence sets arranged in a 2x2x2 design as in Table X above, with *wh*-number (singular, plural), auxiliary/subject number (singular, plural), and subject type (pronoun, DP) manipulated. Stimuli were designed to replicate the structure of the test items used in Badecker (in prep). The first two words of each critical item were always the *wh*-phrase, followed by the auxiliary *was* or *were*, then a DP or pronominal subject, then a main (uninflected) verb. Following Badecker (in prep) an adjective or adverb was added to the sentences with pronominal subjects in order to match the length of the corresponding DP subject stimuli. Example critical stimuli are shown in Table 2.

Table 2: Example stimuli set, Experiment 1.

<table>
<thead>
<tr>
<th>Stimulus sentence</th>
<th>Subj. Number</th>
<th><em>Wh</em> Number</th>
<th>Subj. Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Which antique was the maid polishing in the study?</td>
<td>Singular</td>
<td>Singular</td>
<td>DP</td>
</tr>
<tr>
<td>(b) Which antique was she polishing in the upstairs study?</td>
<td>Singular</td>
<td>Singular</td>
<td>Pronoun</td>
</tr>
<tr>
<td>(c) Which antique were the maids polishing in the study?</td>
<td>Plural</td>
<td>Singular</td>
<td>DP</td>
</tr>
<tr>
<td>(d) Which antique were they polishing in the upstairs study?</td>
<td>Plural</td>
<td>Singular</td>
<td>Pronoun</td>
</tr>
<tr>
<td>(e) Which antiques was the maid polishing in the study?</td>
<td>Singular</td>
<td>Plural</td>
<td>DP</td>
</tr>
<tr>
<td>(f) Which antiques was she polishing in the upstairs study?</td>
<td>Singular</td>
<td>Plural</td>
<td>Pronoun</td>
</tr>
<tr>
<td>(g) Which antiques were the maids polishing in the study?</td>
<td>Plural</td>
<td>Plural</td>
<td>DP</td>
</tr>
<tr>
<td>(h) Which antiques were they polishing in the upstairs study?</td>
<td>Plural</td>
<td>Plural</td>
<td>Pronoun</td>
</tr>
</tbody>
</table>

The 48 critical items were combined with 72 filler items, for a total of 120 trials. Fillers were all questions, but did not feature a fronted object *wh*-phrase (e.g., *Was the advertisement for the club colorful?* or *Who paid for the snacks at last month’s*
meeting?). This resulted in 60% of the total sentences being filler items (comparable to similar studies, e.g. (Wagers et al., 2009). The items were distributed among 8 counterbalanced test lists, and each participant was randomly assigned to a list upon accepting the HIT.

Procedure

The experiment was a self-paced reading task (King & Just, 1991) implemented using ScriptingRT, which compiles into a Flash movie. Workers browsing AMT could see instructions and a link to the consent document. Once a participant accepted the task, a JavaScript function was called to randomly load one of the 8 test lists. The experiment began with on-screen instructions that described the task; participants were told that they would be reading sentences one word at a time, and that pressing the space bar would reveal each subsequent word and hide the previous word. Unlike in a moving-window self-paced reading task, other words in the sentence did not remain on the screen masked. Participants were instructed to read at a natural pace, but slowly enough to comprehend what they read.

Items were randomized using the branching function available through the ScriptingRT library with some additional code written by the author (available at https://code.google.com/p/enochson-amt/). Reading times were captured in milliseconds for each word. A yes/no comprehension question followed each item (e.g., Was the antique in the bedroom?). Participants pressed the “y” key for yes, and the “n” key for no to indicate their response, with feedback given for incorrect responses.

2.3.2. Results
Sentences for which the participant answered the comprehension question incorrectly were removed from the analyses. It is typical in self-paced reading experiments to discard data from participants who answered more than 20% of the comprehension questions incorrectly or whose reading times are more than 2.5 standard deviations from the mean (e.g., Wagers et al., 2009). No participants met either criterion; therefore, analyses were run on all 34 participants. Response times below 100 ms or above 2500 ms were removed; this resulted in loss of between 2% and 2.5% of the data in each of the experiments reported here.

Each sentence can be thought of as comprised of several regions of interest: the *wh* region, the auxiliary region, the subject region, the main verb region, what we will call the V+1 region one word after the verb, and the V+2 region two words after the verb. These regions are illustrated in Table 3 for an example sentence, with regions of most interest highlighted.

<table>
<thead>
<tr>
<th>Table 3: Regions of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Which antiques</em> was <em>the maid</em> polishing <em>in</em> the <em>study?</em></td>
</tr>
<tr>
<td><em>wh</em> auxiliary subject verb V+1 V+2</td>
</tr>
</tbody>
</table>

Because the critical regions differed in the number of letters, and in the case of the subject region, in the number of words, all analyses were performed on residual reading times (Ferreira & Clifton, 1986). Log transformed RTs (Baayen & Milin, 2010) were input into a linear mixed-effects model, using subject number as a random effect, to
calculate fitted and residual reading times. Subsequent linear mixed-effects models of residual RTs, using both participant and item as random effects, were used to analyze the effect of each factor of interest and any interactions between them. All models included the maximal random effects structure justified by the data (Barr, Levy, Scheepers, & Tily, 2013). All statistical modeling and hypothesis testing was performed in R (R Core Team, 2014), and all mixed-effects models were run using the lme4 package (Bates, Maechler, & Bolker, 2012).

**Pronoun vs. DP processing effect**

If subject type (pronoun vs. DP) impacts processing, we expect to see a slow-down in reading time at the subject region for DPs as compared to pronouns. To test whether we can capture this effect using AMT, a linear mixed-effects model was fit using residual reading time as the dependent variable and subject type as a fixed effect. Results indicate that at the subject region, pronouns are read significantly faster than DPs ($\beta = -0.508 \pm 0.02$, $p < 0.0001$). This is illustrated in Figure 3.
Filler-gap effect

The filler-gap effect is characterized by increased processing difficulty at the position of a gap. In our case, this predicts a slow-down in reading time following the verb, where the $wh$- phrase gap is filled (e.g., Crain & Fodor, 1985; Stowe, 1986). To test whether we have captured this effect, a linear mixed-effects model was fit using residual RT as the dependent variable, region as the fixed effect, and participant and item as random effects. Tukey-adjusted pairwise comparisons indicate a significant slowdown...
from the verb region to the V+1 region ($\beta = 0.170 \pm 0.02$, $p < 0.001$), indicative of a filler-gap effect. This effect is illustrated in Figure 4.

![Figure 4: Filler-gap results. Error bars represent error of the mean.](Image)

**Agreement attraction effects**

Recall that for the type of stimuli we are using here—namely grammatical sentences—we expect agreement attraction to present as a slow-down in reading when the subject and verb differ in number from the fronted *wh*-phrase. This would indicate that participants are (erroneously) expecting the verb to agree with the attractor. Here we expect the fronted *wh*-phrase to act as a potential attractor because of its subject-like pre-verbal position in the sentence (Badecker, in prep; MacWhinney, Bates, & Kliegl, 1984).
As mentioned above, agreement attraction is typically found with plural attractors, and thus we also predict a difference in reading-time slow down depending on the number of the *wh*-phrase.

Agreement attraction effects typically spill over from the verb to later regions, (e.g., Pearlmutter et al., 1999; Wagers et al., 2009). Therefore, to assess potential agreement attraction effects, linear mixed-effects models were fit for the verb region, the V+1 region, and the V+2 region, using residual reading time as the dependent variable, and a binary factor coding number mismatch between subject and *wh* number as the fixed effect. Number mismatch was not significant in any of the regions. Since agreement attraction is *more* likely when the subject is singular and the attractor is plural, as discussed above, it could be that agreement attraction is found only in the latter case. A significant interaction between *wh* number and mismatch was indeed found in the verb region, however it did not persist into any of the spillover regions. To assess whether pronoun subjects might block agreement attraction effects (by providing a particularly strong cue to subjecthood), a model was also fit using subject type and number mismatch as fixed effects. This interaction was not significant in any region. Estimates and *p*-values for all models are shown in Table 4.

Table 4: Estimates and (*p*-values) for agreement attraction models.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Verb region</th>
<th>V+1 region</th>
<th>V+2 region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1. mismatch</td>
<td>−0.002 (0.91)</td>
<td>−0.004 (0.87)</td>
<td>−0.013 (0.47)</td>
</tr>
<tr>
<td>Model 2. <em>wh</em>-number * mismatch</td>
<td><strong>0.075 (0.04)</strong></td>
<td>0.039 (0.25)</td>
<td>0.020 (0.58)</td>
</tr>
</tbody>
</table>
The significant slowdown in reading time found when the subject was singular and the attractor plural suggests that we may have uncovered evidence of agreement attraction. However, the fact that the effect was found only in the verb region and not in any later spillover regions contrasts it with previous findings of agreement attraction in other construction types (e.g., Wagers et al., 2009). We address this issue in detail below.

Interim summary
The experiment reported here used a single set of stimuli to test whether three important psycholinguistic effects could be replicated with response time data gathered over the web via AMT. The first two effects—the pronoun vs. DP processing difference and the filler-gap effect—were successfully replicated, using a similar number of participants and items as a more traditional lab-based study. The third effect—agreement attraction—was not convincingly replicated. This could potentially be related to the particular construction we used, or to the fact that we used only grammatical stimuli. While other researchers (e.g., Pearlmutter et al., 1999) have found attraction effects in grammatical sentences in production and self-paced reading studies, Pearlmutter et al. (Pearlmutter et al., 1999) suggests that the magnitude of attraction effects in grammatical sentences is generally smaller than in ungrammatical sentences. Badecker (in prep) did report attraction effects using grammatical wh-questions, however our methodology was not the same; Badecker (in prep) used a production task rather than a comprehension task. Since no other studies demonstrate agreement attraction using grammatical wh-questions in a self-paced reading task, it may be that attraction in this context is not as
robust or reliable as some other contexts. Additionally, although Experiment 1 used a comparable number of participants to other studies of filler-gap dependency and subject definiteness (e.g., Hofmeister & Sag, 2010; Warren & Gibson, 2002), the number is low compared to other agreement attraction studies (e.g., Pearlmutter et al., 1999; Wagers et al., 2009). In Experiments 2 and 3, we seek to replicate agreement attraction using structures that have consistently demonstrated robust effects, specifically prepositional phrase modifiers and relative clause modifiers, using the same number of participants as the corresponding laboratory studies.

2.4. Experiment 2: further investigations of agreement attraction

Experiment 2 attempts to replicate the agreement attraction effects reported in Pearlmutter et al. (1999). Here we focus on agreement attraction in sentences with prepositional phrase modifiers, e.g., *The slogan on the poster(s) was/were designed to get attention*. In such sentences, the DP contained in the prepositional phrase intervenes between the subject and the agreeing verb, potentially triggering agreement attraction (Bock & Miller, 1991; Pearlmutter et al., 1999; Solomon & Pearlmutter, 2004; Vigliocco & Nicol, 1998; Wagers et al., 2009).

2.4.1. Method

Participants

In order to match as closely as possible the task reported in Pearlmutter et al. (1999), we recruited an identical number of participants, namely 82. These participants were recruited and compensated via AMT in the same manner as in Experiment 1 over the course of one week.
Apparatus, Stimuli & Design

As in Experiment 1, this experiment was presented to workers as a ScriptingRT Flash movie embedded in an HTML page. Stimuli for Experiment 2 were taken from Pearlmutter et al. Experiment 1 (1999). All critical stimuli had singular subjects (since attraction is typically most likely with plural attractors), and we manipulated attractor number and sentence grammaticality. A sample stimuli set is shown in Table 5, where the most likely context for agreement attraction should be sentences of type (d). As in Pearlmutter et al. (1999), there were 16 critical test items and 94 filler items for a total of 110 trials. Stimuli were distributed among 4 counterbalanced test lists to which participants were randomly assigned.

Table 5: Example stimuli set, Experiment 2.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Attractor Num.</th>
<th>Grammaticality</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) The slogan on the poster was designed to get attention.</td>
<td>singular</td>
<td>grammatical</td>
</tr>
<tr>
<td>(b) The slogan on the posters was designed to get attention.</td>
<td>plural</td>
<td>grammatical</td>
</tr>
<tr>
<td>(c) The slogan on the poster were designed to get attention.</td>
<td>singular</td>
<td>ungrammatical</td>
</tr>
<tr>
<td>(d) The slogan on the posters were designed to get attention.</td>
<td>plural</td>
<td>ungrammatical</td>
</tr>
</tbody>
</table>

Procedure

HITs were posted to AMT in the same manner as Experiment 1, and the self-paced reading task procedure was identical. This differs slightly from Pearlmutter et al., (1999), who use a moving window design.

2.4.2. Results and Discussion
Sentences for which the participant answered the comprehension question incorrectly were removed from the analyses. As in Experiment 1, no participants answered more than 20% of the comprehension questions incorrectly, and no participants had reading times more than 2.5 standard deviations from the mean. Therefore, analyses were run on all 82 participants. Note that Pearlmutter et al., (1999) ultimately excluded 2 from the analyses due to low comprehension question performance, therefore our replication includes data from two additional participants. Data were processed in the same manner as in Experiment 1.

Recall that our stimuli included both grammatical and ungrammatical sentences. In general, grammatical sentences should be read faster than ungrammatical sentences. However if the intervening DP in the prepositional phrase serves as an agreement attractor, then we expect a decrease in reading time associated with ungrammaticality in sentences that include a number mismatch between the subject and the verb. Put another way, since all subjects are singular, ungrammatical sentences with a plural attractor should be read faster than ungrammatical sentences with a singular attractor. To investigate this, a linear mixed-effects model was fit using residual RT as the dependent variable, grammaticality and number mismatch as fixed effects, and participant and item number as random effects. Pearlmutter et al. (1999) report a spill-over agreement attraction effect two words after the verb. Given the example sentence The slogan on the poster(s) was/were designed to get attention, we would then expect agreement attraction to present as a slow down at the word “to”. A simple effect of ungrammaticality should occur at the verb, with effects potentially spilling over onto the next two regions. Our
data reveal a significant interaction between ungrammaticality and number mismatch at the region corresponding with the word “to” ($\beta = -0.09 \pm 0.03$, $p = 0.005$), such that when an item is ungrammatical and the subject number and attractor number do not match, the item is read faster. This indicates a successful replication of the study in Pearlmutter et al. (Pearlmutter et al., 1999) and is illustrated in Figure 5.

![Figure 5: Agreement attraction with PP modifiers. Error bars represent standard error of the mean.](image)

Experiment 2 demonstrates that agreement attraction effects can be captured using AMT, indicating that our failure to capture such effects in Experiment 1 is likely a function of the stimuli rather than the method. Experiment 3 attempts to extend this
result, replicating attraction effects reported in Wagers et al. (2009) using relative clause modifiers.

2.5. Experiment 3

In Experiment 3 we attempt to replicate agreement attraction effects in relative clause modifiers reported in Wagers et al. (2009). For example, in a sentence like *The runner(s) who the driver see(s) during the commute every morning always wave(s) to say hi*, the main clause subject *runner(s)* functions as a potential attractor for the agreeing relative clause verb *see(s)*. Because the original task as described in Wagers et al. (2009) is substantially longer than Experiments 1 and 2 (a total of 192 items compared to 120 and 110 respectively), here we reduce the number of items and increase the number of subjects.

2.5.1. Method

Participants

Wagers et al. (2009) use data from 30 participants; we doubled this to 60 participants, recruited and compensated as in Experiments 1 and 2. Participants were recruited over the course of 4 days.

Apparatus, Stimuli & Design

As in Experiments 1 and 2, this experiment was presented to workers as a ScriptingRT Flash movie embedded in an HTML page. Stimuli for Experiment 3 come from Wagers et al. Experiment 2 (2009). As mentioned above, in order to keep Experiment 3 consistent with Experiments 1 and 2 in terms of time and compensation, we used a subset of the items; in particular we used the first half of the critical items from
Wagers et al. stimuli, and half the number of filler items. Thus we had 24 critical items and 72 fillers, resulting in 96 total items. All items used singular subjects, and attractor number and grammaticality were manipulated. A sample stimuli set is shown in Table 6, where agreement attraction is expected to be most likely in sentences of type (d). The stimuli were distributed among 4 counterbalanced test lists to which participants were randomly assigned.

Table 6: Example stimuli set, Experiment 3.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Attractor Number</th>
<th>Grammaticality</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) The runner who the driver sees during the commute...</td>
<td>singular</td>
<td>grammatical</td>
</tr>
<tr>
<td>(b) The runners who the driver sees during the commute...</td>
<td>plural</td>
<td>grammatical</td>
</tr>
<tr>
<td>(c) The runner who the driver see during the commute...</td>
<td>singular</td>
<td>ungrammatical</td>
</tr>
<tr>
<td>(d) The runners who the driver see during the commute...</td>
<td>plural</td>
<td>ungrammatical</td>
</tr>
</tbody>
</table>

**Procedure**

HITs were posted to AMT in the same manner as Experiments 1 and 2, and the self-paced reading task procedure was identical. This differs slightly from Wagers et al. (2009), who use a moving window design.

2.5.2. Results and Discussion

Sentences for which the participant answered the comprehension question incorrectly were removed from the analyses. One participant was excluded for having mean reading times more than 2.5 standard deviations from the mean. No participants answered more than 20% of comprehension questions incorrectly. Therefore, analyses
were run on 59 participants. Data were processed in the same manner as in Experiments 1 and 2.

As in Experiment 2, agreement attraction should lead to a decrease in reading time associated with ungrammaticality in sentences that include a number mismatch between the subject and the verb. To assess this, a linear mixed-effects model was fitted using residual RT as the dependent variable, grammaticality and number mismatch as fixed effects, and participant and item number as random effects. The region of interest in these stimuli is two words after the relative clause ends, so in the example sentence *The runner(s) who the driver see(s) during the commute*..., agreement attraction would likely present at the word “the” in “the commute”. Ungrammaticality occurs at the verb “see(s)”, and again effects should spill over onto the next two regions. Our data reveal a significant interaction between ungrammaticality and number mismatch at the region corresponding with the word “the” (β = −0.02 ± 0.008, p = 0.001), indicating agreement attraction effects. This is illustrated in Figure 6.
Figure 6: Agreement attraction with RC modifiers. Error bars represent standard error of the mean.


The first three experiments reported here demonstrate the feasibility of using Amazon Mechanical Turk to collect accurate response time data. The second pilot study, Experiment 4, reported here is a replication of Fine et al., (2013), described in section 1.2 above. This study examines adaptation in sentences that are temporarily ambiguous—the first verb encountered could be a main verb or could introduce a reduced relative clause, as in (1) above, reproduced here as (5a) and (5b), which are ambiguous up until the end of the bracketed region.

5. a) The experienced soldiers [warned about the dangers] before the midnight raid.
b) The experienced soldiers [warned about the dangers] conducted the midnight raid.

c) The experienced soldiers who were told about the dangers conducted the midnight raid.

d) The experienced soldiers spoke about the dangers before the midnight raid.

Importantly, the main verb reading is much more frequent, and thus, all things equal, English speakers will expect it and encountering the second verb will cause a processing slow down. Fine et al., (2013) investigate whether participants can rapidly adapt to changes in this expected distribution.

Half of the participants first read a set of exposure sentences containing either 16 relative clause continuations (8 of which are ambiguous, the rest use *that* or *who* as in 1c), while the other half first read 16 filler items (no relative clauses). In two subsequent blocks, both groups are exposed to the same stimuli: in block 2, 10 relative clauses (5 ambiguous, the rest with main clause only verbs as in 1d), along with filler items; in block 3, 10 main verb constructions (5 ambiguous), along with filler items. Because relative clause continuations are much less frequent than main verb constructions, the relative clause first group has received distributional information that is very different from their prior expectations. If participants are able to use this new information to adapt their expectations, then in block 2, this group should be faster at reading ambiguous relative clause items than the filler first group. A further prediction is that in block 3, the filler first group should be faster at reading ambiguous main verb items than the relative clause first group. This is because the relative clause first group should now have
reversed their expectations about how these ambiguous structures will be resolved in the current linguistic environment. The results from Fine et al., (2013) support these two predictions. The relative clause first group was faster at reading ambiguous items in block 2, but slower at reading ambiguous items in block 3. This suggests that language comprehenders are able to use new distributional information to adapt their expectations about upcoming linguistic structures. In this case, they do so rapidly, over the course of 26 relative clauses in a 15-minute experiment.

2.6.1. Method

Participants

In our replication, 80 participants were recruited, following Fine et al., (2013).

Apparatus, Stimuli & Design

This experiment differs from the previous three reported here in that it uses JavaScript to instantiate the self-paced reading task. It is the same in all other respects. Stimuli were the same as those used in Fine, et al., (2013), presented in the same order. Participants were randomly assigned to either the relative clause first group or the filler first group.

Procedure

HITs were posted to AMT as external HITs. The self-paced reading procedure is identical to the previous experiments.

2.6.2. Results and Discussion

Two participants were excluded for answering more than 20% of comprehension questions incorrectly, and 1 was excluded for reading times greater than 2.5 standard
deviations from the mean; therefore data from a total of 77 participants were analyzed. Participants were randomly assigned to the relative clause (RC) first group or filler first group, and stimuli were the same sentences and the same presentation order used in Fine et al., (2013), presented over the web using original JavaScript code.

The prediction is that the RC first group will be faster than the filler first group at reading relative clauses in block 2, based on exposure in block 1. To test this, a linear mixed-effects model was fit using residual reading time as the dependent variable, condition (filler first vs. RC first) and item type (ambiguous vs. unambiguous) as fixed effects, and participant and item as random effects. At the final word region, there was a main effect of item type such that ambiguous items were read slower than unambiguous items ($\beta = 55.82 \pm 19.3, p = 0.004$) and an interaction between item type and condition, indicating that the ambiguity effect was smaller for the RC first group than for the filler first group ($\beta = -63.14 \pm 27.51, p = 0.02$), demonstrating syntactic adaptation.

To determine whether the ambiguity effect diminishes from block 1 to block 2 for participants in the RC first condition, a linear mixed-effects model was fit using residual reading time as the dependent variable, item type and block as fixed effects, and participant and item as random effects. At the disambiguating region, a main effect of item type shows that ambiguous items are read slower than unambiguous items ($\beta = 102.39 \pm 24.76, p < 0.001$) and a main effect of block shows that items in block 2 are read faster than items in block 1 ($\beta = -155.28 \pm 27.86, p < 0.001$). An interaction between item type and block indicates that the ambiguity effect diminishes from block 1 to block 2 ($\beta = -106.75 \pm 39.82 p = 0.0075$), demonstrating an adaptation effect.
In block 3 we predict that the RC first group will be slower than the filler first group, as a result of having adapted to expect relative clauses rather than main verbs. A linear mixed-effects model was fit using residual reading time as the dependent variable, item type and condition as fixed effects, and participant and item as random effects. At the final word region, the interaction between item type and condition approached significance ($\beta = 41.05 \pm 22.79$, $p = 0.07$), indicating that participants in the RC first condition read ambiguous items slower than participants in the filler first condition.

Taken together, these results constitute a replication of the results reported in Fine et al., (2013). Figure 7 shows the results for block 2, in which both groups read relative clause continuations, and Figure 8 shows the results for block 3, in which both groups read main verb continuations.

Figure 7: Adaptation results, Block 2
In this chapter, we have demonstrated that AMT can be successfully used to conduct psycholinguistic research in which precise response time measurements are necessary. The effects we replicated here involve self-paced reading tasks designed to reveal differences in the processing of pronouns vs. DPs, filler-gap effects, agreement attraction, and rapid syntactic adaptation. While we were not able to convincingly find agreement attraction effects in grammatical $wh$-fronted questions, we were able to replicate agreement attraction in two more established contexts— with prepositional phrase and relative clause modifiers. We used similar numbers of trials and participants as traditional lab studies; the four experiments reported here used between 35–82
participants each, and between 96–120 trials each, well within the typical range for self-paced reading and other psycholinguistic tasks. Importantly, these experiments serve as proofs of concept for the types of experiments reported in Chapters 3 and 4.
3.1. Introduction

This dissertation explores the relationship between statistical learning ability and linguistic adaptation—both syntactic adaptation, as in syntactic priming, as well as phonetic adaptation, as in perceptual learning. If adaptation is the result of updating expectations based on probabilistic information in the linguistic input, then individual differences in adaptation should correlate with individual differences in statistical learning. The experiments reported in Chapter 2 demonstrate that participants recruited via AMT, who complete experiments presented dynamically over the web, perform comparably to laboratory participants in tasks that require precise response time data. Importantly, Experiment 4 replicated the syntactic adaptation effects reported in Fine, Jaeger, Farmer, & Qian, (2013) using a web-based methodology.

Previous research has shown on the one hand that language processing ability is related to statistical learning, and on the other hand that expectations about syntactic structure—which directly impact processing—can be rapidly updated. Experiment 5 investigates the hypothesis that adaptation in the syntactic domain is in fact the result of statistical learning. The latter is predicted to result in updated expectations or predictions about the current linguistic environment, which can lead to faster or slower processing depending on whether subsequent input is in line with these new expectations.
This dissertation tests this hypothesis by asking whether individuals’ ability to rapidly update expectations about syntactic structure is correlated with their performance on a statistical learning task. To rule out potentially confounding factors, the experiment also tests whether other measures of cognitive function, specifically verbal working memory, cognitive control, and print exposure, can predict performance on the syntactic processing task.

Results confirm the predicted correlation between statistical learning ability and syntactic adaptation to new distributional information involving reduced relative clauses. Importantly, none of the other measures of cognitive ability predict statistical learning performance.

3.2. Experiment 5.

As demonstrated in Chapter 2, Amazon Mechanical Turk is a viable resource for linguistic research, even when precise response time data is necessary, providing an efficient and cost-effective alternative to undergraduate laboratory participants. Thus for Experiment 5, 30 participants were recruited and compensated via AMT. An initial advertisement for the series of experiments briefly described each task, including information about time commitment and compensation. Participants completed each experiment in separate sessions, with the exception of the Author Recognition Task and the Stroop Task, which were administered in the same session. After completing one experiment, participants were emailed information about participating in the next experiment. In most cases, all five experiments were completed within a week. All tasks were displayed dynamically using JavaScript and posted to AMT as external HITs.
In this section, I will discuss each of the five sub-tasks, including procedure and results. In section 3.2 I will discuss correlations between tasks and overall results.

3.2.1. Statistical Learning Task.

The statistical learning measure used here is that of Misyak et al., (2010) who adapt a well-known non-adjacent dependency learning study (Gómez, 2002) using a serial reaction time (SRT) task. This task is intended to function as a measure of an individual’s online implicit statistical learning ability. The non-adjacent dependency to be learned is characterized by a grammar consisting of sequences of $a \ X \ b$, where $a$ and $b$ are one of three sets of paired monosyllabic CVC elements (conceptually similar to a prefix and a suffix) and $X$ is one of a set of 24 intervening elements. Learning non-adjacent dependencies in the absence of any phonological cues is relatively difficult—and thus, importantly, variation among individuals is expected. However, learning has been shown to be possible when the intervening material is highly variable, as in this design (Gómez, 2002; Newport & Aslin, 2004).

**Method**

*Participants*

30 participants were recruited via AMT for the statistical learning task. The task took about 35-40 minutes to complete and participants were compensated $3, in line with typical AMT rates. After accepting the HIT, workers were randomly assigned to one of the 30 test lists generated according to the grammar rules described below.

*Apparatus, Stimuli & Design*
Thirty different test lists were generated according to the following grammar rules. Three sets of dependencies (2-word pairs) were randomly assigned from a set of six phonetically monosyllabic words \((pel, dak, vot, rud, jic, tood)\). Each of the dependencies occurred with each of the 24 variable intervening two-syllable words \((wadim, kicey, puser, fengle, coomo, loga, gople, taspu, hiftam, deecha, vamey, skiger, benez, gensim, feenam, laeljeen, chila, roosa, plizet, balip, malsig, suleb, nilbo, wiffle)\), each with penultimate stress, three times per block, as described below. Each block was randomized, and each word was randomly assigned to either the top or bottom of the screen, with a foil word in the other position. Chi-square tests confirmed that within each test list, individual words were assigned to top and bottom positions an equal number of times (all \(ps\) ns).

Audio stimuli were recorded by a native English speaker in a sound-attenuated room. Files were then length-adjusted in Praat (Boersma & Weenink, 2012) such that each one-syllable word was 500 ms and each two-syllable word was 600 ms.

Procedure
Instantiated in a serial reaction time (SRT) task, the stimuli were presented aurally accompanied by a corresponding visual representation. The computer screen was divided into six boxes—three columns and two rows—with each column corresponding to a portion of the stimulus and each row presenting two alternative possible stimuli. After 250 ms of visual presentation, the audio files played with 250 ms between each word. Participants clicked on each box that corresponds with part of the stimulus, for a total of three clicks per trial, one for each element in \(a \times b\), as depicted in Figure 9 below.
Response times were collected for each click. After three clicks, the screen cleared and after 750 ms, the next trial began.

The task consisted of 8 blocks that were presented consecutively with no indication to the participant that one block was finished and another beginning. Blocks 1-6 and block 8 included each of the intervening syllables presented with each of the three dependencies, for a total of 72 trials per block. Following Misyak et al., (2010), Block 7 consisted of trials that do not obey the grammar of the language; specifically, the first and third elements do not follow the dependency. The purpose of this block is to distinguish between learning the grammar and task learning. If participants learn the dependency, click times for the third stimulus element should drastically slow down during the ungrammatical block, while if an increase in click times signifies task learning effects alone, Block 7 should not differ from surrounding blocks.

After 35 minutes of exposure to the artificial language in the SRT task, participants complete a forced-choice grammaticality judgment task. The task consists of
12 items, 6 of which are grammatical and 6 of which are ungrammatical. All items in the grammaticality judgment task include both visual and audio presentation.

**Results**

Following Misyak et al., (2010) individual scores for the SRT portion of the task were computed by first subtracting initial minus final click times for each stimulus, averaging this score by block, and then subtracting the time difference in Block 1 from Block 8. Scores for the grammaticality judgment portion were the correct number of responses, out of 12. The mean SRT score was 61.2, with a standard deviation of 207, meaning that some participants had positive scores, indicative of learning the dependency, while some had negative scores, indicative of failure to learn the dependency. The mean grammaticality judgment score was 6.4, approximately chance, with a standard deviation of 2.7. Importantly, these results are indicative of individual differences in ability to learn the dependency.

In line with scores reported in Misyak et al. (2010), an inspection of response times for “good” learners, i.e., those who scored above 8 on the grammaticality judgment task (7 of 29 participants, a bit lower than in Misyak et al. (2010)), indicates that response times for the third click (i.e., *rud* in the series *pel wadim rud*) increased steadily from Blocks 1-6, dropped in the ungrammaticality block, and rose again in the recovery block. This is illustrated in Figure 10. This suggests that participants who showed an increase in click times throughout the experiment did so as a result of learning the dependency, rather than simply becoming faster at the task.
3.2.2. Syntactic Adaptation Task

The syntactic adaptation task was based on Fine et al., (2013), which manipulates the probability of encountering a verb functioning as a main verb or introducing a reduced relative clause, as described in Chapter 2.

**Method**

*Participants*

After completing the first task, participants received an email inviting them to participate in the syntactic adaptation task. Only participants who completed the first task were eligible to participate. The task took approximately 20 minutes, and participants were paid $2.
Apparatus, Stimuli & Design

Stimuli consist of 40 critical items of which 20 are ambiguous (10 main verbs (MV), 10 reduced relative clauses (RC)) and 20 are unambiguous (10 main verbs, 10 relative clauses) along with 50 filler items, for a total of 90 sentences. Twenty verbs generated 40 ambiguous items—20 MVs and 20 RCs—which were distributed between two counterbalanced test lists to which participants were randomly assigned. Eight of the 20 verbs come from Fine et al., (2013), with the other 12 created following the same template. We added new stimuli to ensure we were measuring reliable individual differences rather than statistically expected noise. Each critical item consists of 10 words, including 3 in the initial area, 4 in the ambiguous region, 3 in the disambiguating region, and 1 in the final word region. A sample stimulus set is shown in Table 7.

Table 7: Sample stimulus set for the syntactic adaptation task.

<table>
<thead>
<tr>
<th>Stimulus Sentence</th>
<th>Sentence Type</th>
<th>Item Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>The experienced soldiers warned about the dangers</td>
<td>Ambiguous</td>
<td>Main Verb</td>
</tr>
<tr>
<td>before the midnight raid.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The experienced soldiers warned about the dangers</td>
<td>Ambiguous</td>
<td>Relative</td>
</tr>
<tr>
<td>conducted the midnight raid.</td>
<td></td>
<td>Clause</td>
</tr>
<tr>
<td>The experienced soldiers who were told about the</td>
<td>Unambiguous</td>
<td>Relative</td>
</tr>
<tr>
<td>dangers conducted the midnight raid.</td>
<td></td>
<td>Clause</td>
</tr>
<tr>
<td>The experienced soldiers spoke about the dangers before</td>
<td>Unambiguous</td>
<td>Main Verb</td>
</tr>
<tr>
<td>the midnight raid.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Procedure

Participants read each sentence one word at a time, pressing the spacebar to hide the previous word and reveal the next word. Other words in the sentence do not remain
masked on the screen. A yes/no comprehension question follows each sentence, with feedback given for incorrect responses.

**Results**

Response times for one participant exceeded 2.5 standard deviations from the mean, therefore data from 29 participants were analyzed. Syntactic adaptation is measured by the difference in mean reading time between ambiguous and unambiguous RCs in the disambiguating region in the first third of the stimulus list minus the difference in mean reading time between ambiguous and unambiguous RCs in the disambiguating region in the final third of the stimulus list. The mean syntactic adaptation score was 79.2 ms, with a standard deviation of 313.4 ms. Just as the statistical learning task showed evidence of individual differences, these results indicate significant individual differences in adaptation to a novel distribution of reduced relative clauses and main verb continuations.

Figure 11 shows reading times for all ambiguous and unambiguous sentences. At the disambiguating region, reading times are slower for ambiguous items than for unambiguous items, as predicted based on participants’ previous linguistic experience. Although this difference is not significant (p = 0.1) because the number of participants in the current study (n = 30) is smaller than in the original study (n = 80), the results are in the right direction, suggesting that the task is in fact capturing the desired effects.
Figure 11: Residual reading times for ambiguous and unambiguous sentences for the entire stimulus set.

3.2.3. Print Exposure Task

Print exposure refers to an individual’s cumulative reading experience. It is related to vocabulary knowledge and has been shown to be a significant contributor to reading comprehension (Acheson, Wells, & MacDonald, 2008), which may explain individual differences that arise in the self-paced reading task. To assess this potential confound, we administered an Author Recognition Task (Stanovich & West, 1989).

Method

Participants

After completing the first two tasks, 30 participants received an email inviting them to participate in the third task, which was composed of both the author recognition task and the Stroop task. Only participants who completed the first two tasks were eligible to
participate. The two experiments combined took about 15 minutes, and participants were paid $1.

**Apparatus, Stimuli & Design**

The print exposure task used here is an updated version of the Author Recognition Task (Stanovich & West, 1989) from Acheson, et al. (2008). The update replaces some of the original authors with more modern authors that participants are more likely to be familiar with, and it includes more items than the original. The task consists of 130 names, half of which are authors and half of which are foils. The task is presented as a static web page with all 130 items visible.

**Procedure**

Participants click on the names that they believe are real authors, with instructions that selecting incorrect items will count against their score.

**Results**

Print exposure is measured as the number of authors that participants correctly identify plus the number of foil names that participants do not select. The mean author recognition score was 90.1, with a standard deviation of 13.9.

**3.2.4. Cognitive Control Task**

Recent research suggests that cognitive control mechanisms are recruited during syntactic ambiguity resolution (Farmer, Misyak, & Christiansen, 2012; January, Trueswell, & Thompson-Schill, 2009; Novick, Trueswell, & Thompson-Schill, 2005). Because the stimuli in the syntactic adaptation task involve ambiguity resolution, a Stroop task was administered to all participants to determine whether individual differences in the
syntactic processing task can be attributed to differences in cognitive control rather than
differences in syntactic adaptation.

Method

Participants

30 participants completed the Stroop task at the same time as the author recognition task.
Both tasks combined took about 15 minutes, and participants were paid $1.

Apparatus, Stimuli & Design

The Stroop Task used here included five color words and five fonts (red, blue, green, yellow, and purple). On each trial, a word was presented on the screen in bold in either
the matching font color (e.g., the word RED appears in red font) or mismatching font
color (e.g., the word RED appears in blue font).

Procedure

Participants were instructed to identify the font color, as opposed to the word, with an
example give. In other words, if the word RED appeared in blue font, the participant’s
task was to name the color “blue”, not the color “red”. Participants pressed the key
corresponding with the first letter of the font color, so given the previous example,
participants pressed the “b” key to indicate blue font. Response times were collected for
the interval between stimulus presentation and key press.

Results

Cognitive control was measured here as the mean response time for incongruent trials
minus the mean response time for congruent trials. The mean cognitive control score wss
316 ms, with a standard deviation of 338.7 ms, again indicating significant individual differences in performance.

3.2.5. Verbal Working Memory Task

Verbal working memory actively holds temporary information in the mind, and is often measured in individual differences studies (e.g., King & Just, 1991). It has been shown to correlate with native language comprehension, particularly reading. To assess this as a potential confound for performance on the self-paced reading task, we included a version of the reading span task (Daneman & Carpenter, 1980) designed by van den Noort, Bosch, Haverkort, & Hugdahl (2008).

Method

Participants

After completing the first three tasks, participants received an email inviting them to participate in the final experiment, which was the reading span task. Only participants who completed all previous experiments were eligible to participate. The task took approximately 20 minutes, and participants were paid $5. The higher rate of compensation was intended to incentivize participants to complete the final HIT in the set.

Apparatus, Stimulus & Design

The stimuli were those used in van den Noort et al., (2008). These stimuli were matched in word length, character length, and sentence-final word frequency.

Procedure
The reading span task required participants to read a set of sentences while simultaneously holding in memory the last word of each sentence. Participants were instructed to read each sentence that appeared on the screen out loud, while trying to remember the final word of each sentence within a block. Each sentence appeared on the screen until either the participant presses the spacebar key or 6.5 seconds elapses. Participants were explicitly told that the task would be difficult and that they must not write anything down or read silently instead of aloud. At the end of each block, ranging from 2-6 sentences, the word “Recall” appeared on the screen and participants were instructed to type as many sentence-final words as they could remember into a textbox.

**Results**

The reading span score was measured as the total number of sentence-final words correctly recalled across all blocks. The mean reading span score was 78, with a standard deviation of 14.8. This is much higher than the mean score of 65.6 (SD = 5.8) reported in van den Noort et al. (2008), suggesting the possibility that the online participants used illicit methods to keep track of sentence-final words. A post-test questionnaire asked participants whether they used any “tricks” to help them remember the words, and no participants reported using illicit methods such as writing words down or reading the sentences silently. Nevertheless, since the scores are much higher than others reported for the same task, this remains a possibility. As noted in Chapter 2, one of the potential drawbacks of using web-based participants rather than laboratory participants is that the researcher gives up some control of the experimental conditions.
3.3. Overall Results

If a general mechanism like statistical learning underlies both language acquisition and later processing and use, a clear prediction is made: performance on an independent measure of statistical learning should correlate with ability to adapt native language expectations based on novel information. Thirty participants completed 5 distinct tasks to assess this prediction. As mentioned above, on the self-paced reading task response times for one participant exceeded 2.5 standard deviations from the mean. Therefore, data in the overall analysis come from 29 participants.

A simple correlation test indicates that performance on the statistical learning task as measured by the offline grammaticality judgment score correlates with performance on the syntactic adaptation task ($r = 0.41$, $p = 0.025$). This relationship is illustrated in Figure 12.
A stepwise multiple regression was conducted to evaluate whether scores on all of the tasks were necessary to predict syntactic adaptation scores. Results indicate that the model which best fits the data is the model including only grammaticality judgment as a predictor, and that it is a significant predictor ($t = 2.36, p = 0.026$). Model comparison also indicates that there is no significant difference in predictive power between the simplest model including only grammaticality judgment and the full model including all factors ($F = 0.54, p = 0.71$). Taken together, these results show that performance on an offline statistical learning task predicts performance on a syntactic adaptation task, while cognitive control, print exposure, verbal working memory, and an online measure of statistical learning do not.

3.4. Discussion

If syntactic adaptation results from rapidly updating expectations about the distribution of linguistic structures based on novel information, then performance on a statistical learning task should predict performance on a syntactic adaptation task. Results from the set of tasks reported in Experiment 5 indicate that this is in fact the case, suggesting that individual differences in syntactic adaptation are driven by individual differences in statistical learning ability. The experiments presented in this chapter indicate that statistical learning ability predicts syntactic adaptation ability, while no
significant correlations were found between performance on these tasks and various other
cognitive measures, i.e., verbal working memory, cognitive control, and print exposure.

Following (Misyak et al., 2010), we collected both an online measure (SRT score) and an offline measure (grammaticality judgment) of statistical learning. Although (Misyak et al., 2010) found a correlation between SRT score and sentence processing ability, we found a correlation between grammaticality judgment score and syntactic adaptation, while SRT score did not correlate. The magnitude of the effect that the SRT task measures is quite small, typically around 10 ms. Our task included two additional sources of variation that may have uniquely contributed to performance on the online portion of this task. First, participants used multiple different input methods (e.g., mouse, trackpad, touchscreen, etc.), which likely contributed to the individual differences in scores. Second, while Misyak and colleagues use undergraduate students as participants (age: $M = 19.8$, $SD = 1.5$), our participants ranged in age from 23-60 years ($M = 35.2$, $SD = 9.4$). A correlation test showed that age was negatively correlated with SRT scores among our participants ($r = -0.41$, $p = 0.025$), meaning that younger participants performed better at the mouse click portion of the task than older participants. Importantly, age does not correlate with grammaticality judgment ($r = 0.08$, $p = 0.66$) or syntactic adaptation ($r = 0.03$, $p = 0.88$). The SRT task is affected by these other sources of variation while other tasks (e.g., spacebar pressing, author recognition, grammaticality judgment) would not be affected by such factors.

These results suggest that the same mechanism that underlies learning from
distributional cues during acquisition can be used to dynamically impact the linguistic
system at any age. Comparable to implicit learning accounts of syntactic adaptation, phonetic adaptation (perceptual learning) is often discussed in terms of updating expectations and error-driven learning. For example, research has demonstrated that listeners adjust perceptual categories given novel input (Maye, Aslin, & Tanenhaus, 2008) or unreliable cues (Clayards, Tanenhaus, Aslin, & Jacobs, 2008).

Given that both syntactic adaptation (priming) as well as phonetic adaptation (perceptual learning) have been demonstrated to result from incremental expectation updating, it seems plausible that the same cognitive mechanism, namely statistical learning, underlies both. Since Experiment 5 has demonstrated that statistical learning ability correlates with syntactic adaptation, the final experiment will test whether it also correlates with phonetic adaptation. If this prediction is borne out, a further novel prediction is made: despite surface differences, syntactic adaptation and phonological adaptation should be correlated. If statistical learning is a general cognitive mechanism used across different domains, then it could in principle predict both syntactic and phonetic adaptation ability in a single individual.
4.1. Introduction

Statistical learning is typically claimed to be a cognition-general mechanism—that is, not specific to a particular linguistic domain, and indeed not specific to the language faculty (Conway & Christiansen, 2005; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Saffran, Johnson, Aslin, & Newport, 1999). If this is so, then the general hypothesis pursued in this dissertation makes the following novel prediction: superior statistical learning ability should generalize from the syntactic to the phonological domain. In other words, both syntactic adaptation behavior and phonetic adaptation or perceptual learning should be predicted by participants’ statistical learning ability.

Chapter 4 explores this prediction using the same statistical learning measure as in Chapter 3 and a task involving phonological adaptation to foreign-accented speech. Results confirm the predicted correlation between statistical learning ability and phonological adaptation. As in the previous experiment, other potentially confounding individual differences, in this case including auditory working memory, were not significant predictors of phonological adaptation.

4.2. Experiment 6.

For Experiment 6, 30 participants were recruited and compensated via AMT. As with Experiment 5, an initial advertisement for the series of HITs briefly described each
task, including information about the time commitment and compensation. Four tasks were included in this series: a statistical learning task, a syntactic adaptation task, a phonetic adaptation task, an auditory working memory (digit span) task. Participants completed each task in separate sessions, with the exception of the phonetic adaptation task and the digit span task, which were administered in the same session. After completing one experiment, participants received an email with information about participating in the next experiment. Participants were required to complete each previous task before completing subsequent tasks, and participants were not allowed to skip tasks or complete them out of order. Workers who participated in Experiment 5 were prohibited from participating in Experiment 6. In most cases, all three HITs were completed within one week. All tasks were dynamically displayed using JavaScript and posted to AMT as external HITs.

In this section, I will discuss the design and procedure each of the four tasks as well as the results for each task. In section 4.2 I will report tests of correlations between tasks and overall results.

4.2.1. Statistical learning task.

The statistical learning task is a non-adjacent dependency learning task from Gómez (2002), instantiated in a serial reaction time task as in Misyak, Christiansen, & Tomblin (2010), as in Chapter 3.

Method

Participants
Thirty participants recruited via AMT completed the statistical learning task described in Chapter 3. The task took about 35-40 minutes to complete, and participants were paid $6.

Apparatus, Stimuli & Design

The test apparatus and stimulus lists are the same as reported in Chapter 3.

Procedure

The procedure for the serial reaction time task and grammaticality judgment task are the same as described in Chapter 3.

Results

One participant self-reported as a non-native English speaker after completing all tasks, therefore, data come from 29 participants. Individual scores on the SRT task and the grammaticality judgment task were computed in the same manner as in Chapter 3. The mean SRT score was 30.7 ms, with a standard deviation of 271 ms. As in Chapter 3, these scores are indicative of individual variation in task performance. The mean grammaticality judgment score was 7.5, with a standard deviation of 2.7. This is slightly higher than the scores from Experiment 5 (M = 6.4), suggesting that more participants successfully learned the non-adjacent dependency in Experiment 6. The standard deviation was the same in both experiments, indicating individual variation in performance on the grammaticality judgment task as well.

4.2.2. Syntactic adaptation task.

The syntactic adaptation task used temporarily ambiguous stimuli manipulating the distribution of reduced relative clauses versus main verb continuations, presented as a self-paced reading task, as in Experiment 5.
Method

Participants

Thirty participants completed the syntactic adaptation task, which took about 15 minutes to complete, and were compensated $3.

Apparatus, Stimuli & Design

The stimulus lists are the same as those used in Chapter 3.

Procedure

The self-paced reading procedure is the same as described in Chapter 3.

Results

Response times for one participant exceeded 2.5 standard deviations from the group mean, and one participant was not a native speaker of English, therefore data for the analysis come from 28 participants. Individual measures of syntactic adaptation were computed in the same manner as in Experiment 5. The mean syntactic adaptation score was 82 ms, with a standard deviation of 271 ms, again indicating significant individual differences in adapting to a novel distribution of reduced relative clauses and main verb continuations.

At the disambiguating region, reading times were slower for ambiguous items than for unambiguous items, as predicted based on participants’ presumed previous linguistic experience, and this difference was significant (t = −8.1, p <001). Reading times for ambiguous and unambiguous sentences is illustrated in Figure 13.
Figure 13: Residual reading times for ambiguous and unambiguous items across the entire stimulus set.

**4.2.3. Phonetic adaptation task.**

The phonetic adaptation task was based on Clarke & Garrett (2004), which demonstrated that participants can adapt to foreign-accented speech with only 16 sentences of exposure. Based on the assumption that foreign-accented speech contains regularities that a listener can exploit, those listeners who are better able to track these regularities should be more effective and efficient at processing the accented speech after a very short period of exposure. In this task, participants heard a sentence spoken with an accent, and were then asked to decide whether a visually-presented word was the same or different from the last word in the spoken sentence.

**Method**

**Participants**
Thirty participants completed the phonetic adaptation task. This task and the phonological working memory task were completed in the same HIT, which took about 10 minutes to complete, and for which participants were paid $3.

*Apparatus, Stimuli & Design*

Stimuli came from the scripted sentence reading portion of the Wildcat Corpus (Bradlow & Alexander, 2007; Van Engen et al., 2010) spoken by a native Russian speaker. Sixteen sentences from the low-predictability final word condition were used. Example sentences include *Mom talked about the pie* and *We pointed at the bird*. In both cases, the final word, probed in the task, is not predictable from the context. The predictability of the sentence-final words in the low-predictability condition was normed, as described in Bradlow & Alexander (2007). Each sentence from the corpus was copied into its own file with silences at the beginning and end trimmed. Following Clarke & Garrett (2004), foils for the sentence-final words were generated by changing one sound either in the onset, vowel, or coda of the word. For example, the foil word for the stimulus ending in *pie* was *pay*, and the foil for the stimulus ending in *bird* was *board*. Four blocks of four sentences each were presented to each participant with no breaks between blocks. Two test lists were generated, such that final words and foils were counterbalanced across lists. Participants were randomly assigned to one of the two lists upon accepting the HIT.

*Procedure*

On each trial, participants heard a sentence uttered by a speaker with an accent. Immediately after the audio ended, a word appeared on the screen. Participants indicated
with a mouse click whether the word on the screen was the same as the last word they heard. For half of the trials, the correct answer was “yes”, and for half it was “no”.

Results

As noted above, one participant was not a native English speaker, therefore results are computed for 29 participants. No participants reported any level of proficiency in Russian, the native language of the speaker. Individual scores were computed in two ways. First, a response time score was computed by subtracting the mean response time in Block 4 from the mean response time in Block 1. Second, an accuracy score was computed by counting the number of trials for which the participant correctly answered “yes” or “no”. The mean response time score was 156 ms, with a standard deviation of 327 ms. Like the other tasks reported here, this task showed significant individual differences in performance. The mean accuracy score was 23.25, with a standard deviation of 1.2. This indicates somewhat limited variation, but, as discussed below, these differences prove to be important.

Mean response times by block are plotted in Figure 14 below. Results indicate a drop in response time from block 1 to block 2 of 320 ms, and then fairly consistent response times for subsequent blocks, indicative of very rapid adaptation to foreign-accented speech. Clarke & Garrett (2004) report a somewhat smaller difference between blocks 1 and 2 (98 ms), although the difference between block 1 and block 4 reported here (139 ms) is comparable to the difference between block 1 and block 4 reported in Clarke & Garrett (2004) (168 ms). An ANOVA was fit using response time as the dependent variable and click time as the predictor variable to test whether the differences
between blocks is significant. Tukey-adjusted pairwise comparisons indicate that no between-block differences are significant (all $p$s ns). Although this difference is not significant ($p = 0.14$), likely because the number of participants in the current study ($n = 30$) is smaller than in the original study ($n = 48$), the results are in the right direction, with the difference between block 1 and block 2 being the largest, suggesting that the task is in fact capturing the desired adaptation effects.

![Mean response times by block for the phonetic adaptation task. Error bars represent standard error of the mean.](image)

Figure 14: Mean response times by block for the phonetic adaptation task. Error bars represent standard error of the mean.

Accuracy by block is plotted in Figure 15. Results show that Block 1 accuracy is the lowest (93%), with accuracy increasing and remaining about at ceiling through subsequent blocks (98%, 98%, and 96%, respectively). Clarke & Garrett (2004) include
an accuracy measure in their task, with accuracy rates somewhat lower than reported here (block 1 = 86%, block 2 = 95%, block 3 = 90%, block 4 = 95%). An ANOVA was fit using response time as the dependent variable and block number as the predictor variable to test whether the differences between blocks is significant. Tukey-adjusted pairwise comparisons indicate that no between-block differences are significant (all ps ns).

Figure 15: Phonetic adaptation accuracy scores by block.

4.2.4. Phonological working memory task.

As discussed in Chapter 3, working memory actively holds information in the mind and is often reported in individual differences studies (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992). Phonological working memory tasks typically require participants to temporarily store auditory information which they later repeat back, either verbally or in
writing. The working memory task used in this set of experiments is the digit span task, which is a widely used example of this type of test (Baddeley, Gathercole, & Papagno, 1998), and is comparable in procedure to the reading span task used in Chapter 3.

Method

Participants

Thirty participants completed the digit span task in the same session as the phonetic adaptation task. The two tasks combined took about 10 minutes to complete, and participants were paid $3.

Apparatus, Stimuli & Design

Stimuli were created by recording the numbers 1-9 spoken in a monotone voice. Each spoken number was then copied into its own audio file with silences at the beginning and end trimmed.

Procedure

Participants completed the digit span task in the same HIT as the phonological adaptation task. Participants heard a set of numbers spoken in a monotone voice with two seconds of silence between each number. After each set of numbers, participants typed the numbers they remembered into a textbox on screen. The task began with a set of two numbers, and incrementally increased in length every two trials. The task continued until the participant was unable to successfully repeat back either trial of a given length. Number sets were randomly generated using JavaScript.

Results
The number of digits in the largest set that the participant could successfully repeat back was used as the participant’s digit span score. The mean digit span score was 8.8, with a standard deviation of 3.8, again demonstrating individual differences in task performance.

4.3. Overall Results

The results from Experiment 5 demonstrated that performance on a statistical learning task predicts performance on a syntactic adaptation task. Research on perceptual learning is often couched in the same terms; tracking and encoding regularities in the speech signal leads to more effective and efficient adaptation to new information (Clarke & Garrett, 2004; Kleinschmidt & Jaeger, 2012; Maye, Aslin, & Tanenhaus, 2008). Experiment 6 tested the prediction that statistical learning performance also predicts phonetic adaptation, i.e., perceptual learning. Thirty participants completed 4 tasks to assess this prediction. As mentioned above, one participant had response time scores on the syntactic adaptation task that exceeded 2.5 standard deviations from the mean, and one participant was not a native English speaker. Therefore, data from 28 participants are included in the overall analysis.

Based on the findings from Experiment 5—that grammaticality judgment scores in an artificial language learning task correlate with response time changes in a syntactic adaptation task—it is predicted that in this experiment, performance on the grammaticality judgment task will correlate with response time changes in a phonetic adaptation task. Since the experiment tests whether the same underlying mechanism, namely statistical learning, underlies both adaptation processes, a further prediction is
that performance on the two adaptation tasks will co-vary by individual. However, because mouse click response times did not correlate with offline scores in the statistical learning task in Experiment 5, both response time and accuracy measures will be used to investigate correlations between tasks.

A simple correlation test indicates that performance on the statistical learning task, as measured by grammaticality judgment score, correlates with performance on the phonetic adaptation task, as measured by number of correct responses ($r = 0.45$, $p = 0.02$). This relationship is illustrated in Figure 16. The response time measure of phonetic adaptation (initial block RT minus final block RT) does not correlate with performance on the grammaticality judgment task ($r = -0.11$, $p = 0.57$). Importantly, it also does not correlate with accuracy ($r = -0.11$, $p = 0.59$), indicating that participants with lower accuracy scores are not necessarily slower, reducing the likelihood that lower performance is due to lower effort.
A stepwise multiple regression was conducted to evaluate whether scores on all of the tasks (grammaticality judgment score, SRT score, syntactic adaptation score, phonetic adaptation RT score, digit span score) were necessary to predict phonetic adaptation scores. Results indicate that the model which best fits the data is the model including grammaticality judgment and SRT scores, with grammaticality judgment being the only significant predictor. Results of the simpler model are reported in Table 8. Model comparison also indicates that there is no significant difference in predictive power between the model including only grammaticality judgment and SRT scores and the full model (F = 0.35, p = 0.79).

Table 8: T statistics and p-values for each factor in the simplified regression model.

<table>
<thead>
<tr>
<th>Factor</th>
<th>T statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammaticality Judgment</td>
<td>2.35</td>
<td>0.03</td>
</tr>
<tr>
<td>SRT</td>
<td>-1.6</td>
<td>0.12</td>
</tr>
</tbody>
</table>

These results indicate that statistical learning ability predicts phonetic adaptation, suggesting that the same underlying cognitive process drives both syntactic and phonetic adaptation.
Experiment 5 showed that statistical learning correlates with syntactic adaptation. Among the participants in Experiment 6, the correlation between reading time difference at the disambiguating region did not correlate with grammaticality judgment score ($p = 0.8$). However, the full replication of Fine, Jaeger, Farmer, & Qian (2013) reported in Experiment 4 showed that while some effects emerged in the disambiguating region, some did not emerge until the final word region. Among the participants in Experiment 6, grammaticality judgment score correlates with reading time difference in the final word region ($r = 0.38$, $p = 0.047$), rather than in the disambiguating region. Figure 17 shows this relationship.

![Figure 17: Correlation between performance on the syntactic adaptation task (at the final region) and the grammaticality judgment task.](image-url)
There are two factors that may lead to the effects emerging later than the disambiguating region. First, some of the stimuli may not be fully disambiguated by the end of the disambiguating region (Jaeger, p.c.). For example, the stimulus *Several angry workers warned about low wages decided to file complaints* could be followed by a third verb phrase, resulting in a conjunction reading rather than a reduced relative clause reading, such as in *Several angry workers warned about low wages, decided to file complaints, and succeeded in their goal*. Second, the self-paced reading tasks reported here are presented as a single word on the screen at a time, rather than in a moving window paradigm, meaning that participants are unable to see how many words are remaining in each sentence. This may increase the potential that participants do not revise their predictions until later in the sentence.

Overall, the results in Experiment 6 indicate that statistical learning ability correlates with phonetic adaptation, comparable to the way in which statistical learning ability correlates with rapid syntactic adaptation.

4.4. Discussion

Chapter 3 demonstrated that syntactic adaptation resulting from rapid updating of expectations can be predicted from performance on a statistical learning task. Chapter 4 explored whether a functionally similar process occurring in a different linguistic domain, phonetic adaptation, is driven by the same underlying mechanism. The results found in Experiment 5 led to two predictions: performance on a statistical learning task will correlate with performance on a phonetic adaptation task, and syntactic adaptation and phonetic adaptation will co-vary within an individual. The first prediction was borne out;
performance on a phonetic adaptation task—as measured by accuracy—was predicted from performance on a statistical learning task, suggesting that syntactic adaptation and phonetic adaptation are indeed fundamentally related processes.

Interestingly, the second prediction was not borne out here. Although performance on a syntactic adaptation task and a phonetic adaptation task both correlate with statistical learning ability, syntactic adaptation scores and phonetic adaptation scores do not correlate among these participants. Neither the disambiguating region score nor the final word score correlates with the phonetic adaptation response time score or the accuracy score (all ps ns). This may be because there are other sources of individual variation involved in both syntactic and phonetic processing. Many cognitive factors impact language processing, and there are certainly differences between the types of noise that occur in syntactic tasks and phonetic. For example differences in auditory acuity may add a source of noise to the phonetic adaptation task alone, while print exposure is clearly only relevant to syntactic adaptation. These additional sources of variation may obscure any correlation between the two tasks in this experiment despite the same cognitive mechanism underlying both processes. Another possibility is that phonetic adaptation occurred so rapidly in the particular task used here that the measure is not fine-grained enough to capture differences in the time course of phonetic adaptation effects. I discuss possible alternative tasks which could achieve this in Chapter 5.

One additional aspect of the results deserves mention here. As was the case with the statistical learning task, the response time measure in the phonetic adaptation task was
not a reliable measure of individual differences in adaptation. There are (at least) two potential explanations for this. First, recall that results of the SRT task in Chapter 5 demonstrated effects of age on mouse movement combined with mouse click. It is possible that the same source of noise arises to some extent in this case, although here age did not correlate significantly with RT. The second potential explanation is that task learning was confounded with adaptation effects in this task. The syntactic adaptation task measured the difference in click time between ambiguous and unambiguous items in both the first and last block. This difference score controlled for task learning effects. Similarly, the serial reaction time portion of the statistical learning task included an ungrammatical block and a recovery block to tease apart task learning effects from artificial language learning effects. In an attempt to capture what is known to be a very rapid effect (Clarke & Garrett, 2004), the phonetic adaptation task did not include a separate measure of task learning. For example, the task could have included both low-predictability and high-predictability sentences, which would have allowed for a difference measure comparable to the syntactic adaptation task rather than a simple response time measure. However, because this would have provided participants with twice as much exposure, and in order to directly replicate the method used by Clarke & Garrett (2004), the high-predictability sentences were not included in this task. If task learning is particularly important for changes in RT—as opposed to accuracy—then this would explain why RT scores are not clearly reflecting individual differences in adaptation.
CHAPTER 5: DISCUSSION AND CONCLUSION

5.1. Introduction

Interpreting variable input across a noisy channel to comprehend a speaker’s intended message is one of the primary challenges of comprehension. To accomplish this, listeners derive a set of expectations about upcoming linguistic structures from information sources such as grammatical knowledge and prior linguistic experience, and these expectations influence how linguistic input is processed and understood. Various cognitive resources are recruited to accomplish the task of language comprehension, and while all normal, healthy adults possess native speaker linguistic competence, many of these cognitive resources exhibit significant individual differences. One such cognitive mechanism is statistical learning—the ability to track distributional information available in the input. Listeners’ expectations can be dynamically updated (Fine, Jaeger, Farmer, & Qian, 2013), suggesting that statistical learning is not a process recruited solely during language acquisition, as it is typically understood, but is actively used during language comprehension as well.

This dissertation proposes that statistical learning drives language adaptation, and as a consequence, individual differences in statistical learning predict differences in adaptation. Three sets of experiments were run to address this hypothesis. The first set
of experiments serves as proof of concept, demonstrating that response time data can be collected accurately and efficiently over the web using crowd-sourcing services like Amazon Mechanical Turk. The second set of experiments explores the relationship between statistical learning ability and syntactic adaptation. Performance on a language processing task involving temporarily ambiguous main verb / relative clause stimuli correlated with performance on a long-distance dependency artificial language learning task. These results suggest that the two processes—statistical learning and syntactic adaptation—co-vary within an individual. The third set of experiments tests whether statistical learning also underlies phonetic adaptation, addressing the novel prediction that syntactic adaptation and phonetic adaptation are driven by the same underlying cognitive mechanism. Again, performance on an adaptation task—in this case, adaptation to foreign-accented speech—correlates with performance on a statistical learning task of non-adjacent dependencies. This suggests that both types of adaptation, despite occurring in different linguistic domains, are driven by the same cognitive mechanism, namely statistical learning.

5.2. Implications

These findings have implications for current theories of adaptation and related effects, like syntactic priming. As discussed in Chapter 1, syntactic priming is typically discussed under one of two accounts. One account argues that syntactic priming is a form of error-driven implicit learning, suggesting that priming occurs as the result of updating expectations about the distribution of syntactic structures to better match the distribution of the input (Jaeger & Snider, 2013; Thothathiri & Snedeker, 2008). By contrast, a short-
term activation boost account of priming attributes syntactic priming solely to a short-term boost in activation of recently processed material (Pickering & Branigan, 1998). The observed correlation reported here between statistical learning and syntactic adaptation is consistent with error-driven implicit learning accounts of syntactic priming, but not with short-term activation accounts. Unexpected distributions result in rapid adaptation, even when different stimuli types are mixed together and would both be activated. Short-term activation boost accounts predict similar reduction in reading time for both ambiguous and unambiguous stimuli as the task progresses, which was not the case.

This dissertation demonstrates that statistical learning underlies perceptual learning as well, indicating that it is active over multiple linguistic domains. While perceptual learning is a comprehension-based phenomenon, phonetic accommodation represents a similar process in production. Discussions of phonetic accommodation address whether it is an automatic process, or one that can be consciously mediated by social factors (Babel, 2009). Social factors may influence the degree of attention people pay to their interlocutors, modulating the extent to which they learn about the characteristics of their interlocutor’s speech. Nevertheless, I would argue that the effects of such factors on phonetic accommodation do not constitute evidence that statistical learning is not at play. Indeed, the results reported here suggest that phonetic accommodation is, at least to some extent, automatic.

The correlation between statistical learning and both syntactic and phonetic adaptation suggests a reformulation of the traditional generative notion of a grammar. In most generative linguistic theories, a grammar is understood to be a finite set of rules
acquired during initial language acquisition, which is then static throughout the lifespan (e.g., Chomsky, 1995). Taken together with the growing body of evidence for linguistic adaptation, the findings reported here suggest that this is overly simplistic; instead, an individual’s grammar is, at least to some degree, malleable and will be affected by further language experience—particularly when it is unexpected. This is in line with recent work in sociolinguistics (e.g., Sankoff, 2005; Sankoff & Blondeau, 2007), which has argued that language change does not just occur between generations of speakers, rather individuals exhibit language change within their lifetime. For instance, Sankoff & Blondeau (2007) investigated production of ‘r’ in Montreal French. They found that while many speakers are stable in their production as either /r/ or /R/ after a critical period, a sizeable minority of speakers demonstrated substantial changes from one to the other realization across the lifespan. Of course, there are limits to the malleability of grammar, particularly in the case of second language acquisition (Birdsong, 1999; Lenneberg, Chomsky, & Marx, 1967). For example, despite the fact that people show rapid adaptation to a novel accent (Clarke & Garrett, 2004; Maye, Aslin, & Tanenhaus, 2008), and adapt production to match their interlocutor (Babel, 2009, 2012; Babel & Bulatov, 2012), adult language learners are rarely able to produce fully native-like phonology. This suggest that there are limits to how radically a grammar can be updated, although the extent to which is this due to changes in the statistical learning mechanisms is not clear.

Given the finding that adaptation is a result of statistical learning, there are broad implications for those fields this research brings together, namely language acquisition,
cognitive psychology and psycholinguistics. First, while it is widely believed that statistical learning plays a role in first language acquisition, these results bear on the important idea that learning from distributional cues can be used to dynamically impact language at any age. This includes acquiring a second language and adapting to new information about an already acquired one. These findings serve as additional evidence that statistical learning is indeed a general cognitive mechanism, tying together two types of behavior—syntactic priming or adaptation and perceptual learning—which on the surface appear quite distinct. As this dissertation has bearing on our conceptualization of grammar and how it is malleable after initial acquisition, it also furthers our understanding of the source of individual differences in linguistic performance, which has implications for second language teaching and applied linguistics. These results suggest that the mechanism behind these differences is active across the lifespan, therefore the kinds of tasks used here could conceivably be developed as diagnostic tools for assessment of language ability, and how it changes with age. Additionally, they could also be used to test whether statistical learning is being engaged during second language acquisition. If adult language learners are trying to do very explicit learning, it may not, potentially addressing some of the limits on how malleable a grammar is. Lastly, this project involved developing new web-based tools which make conducting cognitive psychology and psycholinguistics research faster and less resource intensive. Some of these tools are reported in Enochson & Culbertson (2015) and are publicly available at https://code.google.com/p/enochson-amt/.

5.3. Future research
The reaction time results on the statistical learning task suggest that participants may not understand the instructions. Response times remain consistent across all blocks, perhaps initially suggesting that participants are unilaterally unable to learn the difficult non-adjacent dependency. However, grammaticality judgment scores indicate otherwise, with about a third of participants performing better than chance. Alternatively, participants may be misinterpreting the instructions in the absence of a demonstration or example and clicking all three boxes after listening to the stimulus. This is supported by performance on the grammaticality judgment task as well as reports from participants regarding how long the task takes. Many participants report the task taking appreciably longer than expected, indicating misinterpretation of the task requirements. One possibility for future work could be to alter the directions or include a sample set of words to ensure participant click times represent processing times.

The phonetic adaptation task results in adaptation that occurs very rapidly (Clarke & Garrett, 2004), and a task measuring individual differences in this kind of adaptation would need to capture very fine-grained response time measures. It may be that collecting response times after every sentence is not often enough, and a significant amount of adaptation occurs within one sentence of exposure. One possible alternative to this task is to instead use stimuli with just one distribution altered, for example shifted VOTs, in order to control how much or little exposure participants get. A task such as in Kraljic & Samuel (2006), where each stimulus is one word, with the t/d distribution manipulated, may provide more fine-grained response time data regarding individual differences in phonetic adaptation.
Section 4.2 discusses the possibility that not all the stimuli used in the syntactic adaptation task are fully disambiguated by the end of the disambiguating region. In order to address this possibility, future work should consider using a moving-window paradigm in the self-paced reading task. This would potentially allow participants to realize the end of the sentence is upcoming and revise predictions sooner. Finding effects in two different regions obscures reliable effects and results in less precise individual differences measures.

Misyak & Christiansen (2012) found that long-distance dependency learning and adjacent dependency learning did not correlate with each other, and that each type of artificial language learning correlated with different language processing and cognition factors. Rather than being evidence of a different underlying mechanism, this may instead be indicative of other sources of variation and noise, such as those suggested to explain why syntactic adaptation and phonetic adaptation did not correlate in the studies reported here. While this dissertation argues that general statistical learning ability—i.e., not specifically non-adjacent learning—underlies adaptation, it is possible that the artificial language learning task used here is too difficult, and a somewhat easier statistical learning task, e.g., an adjacent dependency learning task, may result in more individual variation in statistical learning performance. Artificial language learning tasks involving non-adjacent dependency learning are known to be difficult (Gómez, 2002), perhaps resulting in fewer individual differences than easier statistical learning tasks because of a floor effect.
The syntactic adaptation task used here manipulates the distribution of main verbs and reduced relative clauses, which differ in overall frequency but also in terms of structural (and processing) complexity (Hawkins, 1999; Just & Carpenter, 1992). Future steps in this line of research could tease apart these two factors by using a distribution that is based purely on frequency differences, and is conditioned on particular verbs, such as the dative alternation. The dative alternation—which refers to the fact that some verbs in English can form the dative using both prepositional object (...gave X to Y) and double object (...gave Y X) constructions. Jaeger & Snider (2013) demonstrated that the strength of the priming effect for verbs that participate in the dative alternation is related to the prediction error associated with that verb. For example if a verb like owe, which is heavily biased toward the double object dative, is presented to participants in the prepositional dative, a stronger priming effect is found relative to an unbiased verb. This suggests that production priming, like the adaptation found in Fine et al. (2013), results from expectations formed based on prior input. Thus this research would extend the correlation between statistical learning and adaptation to a new syntactic distribution.

If statistical learning is indeed a domain-general mechanism, and in fact not limited to the language faculty (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Saffran, Johnson, Aslin, & Newport, 1999; although see Conway & Christiansen, 2005 for a discussion of modality-specific differences), then performance on a non-linguistic statistical learning task should also predict performance on adaptation tasks. Tasks involving ordered sequences of shapes, colors, or animals can be used in place of language-based statistical learning tasks as a measure of language-independent statistical
learning ability. If the correlation between statistical learning and linguistic adaptation tasks persists, this would support the argument that statistical learning is a domain-general cognitive mechanism.

5.4. Conclusion

This dissertation investigated the hypothesized relationship between statistical learning ability and rapid adaptation, both in the syntactic domain and the phonetic domain. Chapter 2 reported the results of a series of replication studies serving to illustrate the validity web-based research methods in gathering response time data. These experiments successfully replicated robust psycholinguistic effects in Amazon Mechanical Turk, using comparable numbers of stimuli and participants as laboratory studies. Chapter 3 then made use of these general tools to test whether individual differences in statistical learning would predict performance on a syntactic adaptation task. Results revealed that statistical learning ability did indeed predict syntactic adaptation—operationalized as a change in read-time speed of previous unexpected constructions—while verbal working memory, cognitive control, and print exposure did not. Chapter 4 extended this finding to the domain of phonetic adaptation, with results again showing that statistical learning ability predicted performance on a phonetic adaptation task—operationalized as accuracy in identification of foreign-accented words. These results inform our understanding of the function and domain-generality of statistical learning beyond language acquisition, how grammars are (at least somewhat) malleable beyond initial acquisition, and form the basis for exploring these relationships further.
REFERENCES


BIOGRAPHY

Kelly Enochson holds a Bachelor of Arts degree in English from the University of Virginia, and a Master of Arts degree in English with a concentration in Linguistics from George Mason University.