Published in final edited form as:

Cognition. 2013 July ; 128(1): 82–102. doi:10.1016/j.cognition.2012.12.006.

Is Statistical Learning Constrained by Lower Level Perceptual Organization?

Lauren L. Emberson¹, Ran Liu², and Jason D. Zevin^{3,4}

¹Brain and Cognitive Sciences Department, University of Rochester

²Psychology Department, Carnegie Mellon University

³Sackler Institute for Developmental Psychobiology, Weill-Cornell Medical School

⁴Haskins Laboratories

Abstract

In order for statistical information to aid in complex developmental processes such as language acquisition, learning from higher-order statistics (e.g. across successive syllables in a speech stream to support segmentation) must be possible while perceptual abilities (e.g. speech categorization) are still developing. The current study examines how perceptual organization interacts with statistical learning. Adult participants were presented with multiple exemplars from novel, complex sound categories designed to reflect some of the spectral complexity and variability of speech. These categories were organized into sequential pairs and presented such that higher-order statistics, defined based on sound categories, could support stream segmentation. Perceptual similarity judgments and multi-dimensional scaling revealed that participants only perceived three perceptual clusters of sounds and thus did not distinguish the four experimenter-defined categories, creating a tension between lower level perceptual organization and higher-order statistical information. We examined whether the resulting pattern of learning is more consistent with statistical learning being “bottom-up,” constrained by the lower levels of organization, or “top-down,” such that higher-order statistical information of the stimulus stream takes priority over the perceptual organization, and perhaps influences perceptual organization. We consistently find evidence that learning is constrained by perceptual organization. Moreover, participants generalize their learning to novel sounds that occupy a similar perceptual space, suggesting that statistical learning occurs based on regions of or clusters in perceptual space. Overall, these results reveal a constraint on learning of sound sequences, such that statistical information is determined based on lower level organization. These findings have important implications for the role of statistical learning in language acquisition.

© 2013 Elsevier B.V. All rights reserved.

Correspondence to: Lauren L. Emberson, Brain and Cognitive Sciences Department, University of Rochester, Meliora Hall, Box 270268, Rochester, NY 14627-0268, USA, Phone: 607.342.1690, Fax: 585-442-9216, lemberson@bcs.rochester.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Language learning; cognitive development; statistical learning; Bayesian learning; speech perception; speech categorization; perceptual variability

1. Introduction

Starting in infancy, and continuing into adulthood, humans are highly sensitive to statistical regularities in their environment. From these regularities, it is possible to learn a large amount about the structure of the world without explicit feedback or innate knowledge. In a little over a decade, statistical learning has been implicated in the processing and acquisition of a variety of perceptual and cognitive skills, including knowledge of causal structure and human action (Baldwin, Anderer, Saffran, & Meyer, 2008; Oakes & Cohen, 1990; Oakes & Cohen, 1995; Sobel, Tenenbaum, & Gopnik, 2004), visual processing (Brady & Oliva, 2008; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Yuille & Kersten, 2006), and, the focus of this paper, language learning (Goldwater, Griffiths, & Johnson, 2009; Saffran, Aslin & Newport, 1996).

It is common practice for the learning tasks and models that populate this literature to investigate learning from statistical information at a single level of description. Studies either investigate statistical learning available at a relatively low level of the sensory signal (e.g. acquiring the internal structure of categories through distributional statistics) or characterize relationships at higher levels of description (e.g. segmenting streams of categorizable units, such as syllables, into higher-order structures, such as words). As a consequence, most of the existing statistical learning literature implicitly assumes that statistical information at lower levels is resolved *prior* to learning of statistics at higher levels of description. However, developmental trajectories suggest that this assumption is not always borne out and that there may be substantial interactions between statistical information at multiple levels of description. These interactions have important implications for our understanding of the specific mechanisms that support learning and development at each level.

In the current paper, we examine how statistical learning proceeds when information at lower and higher levels of description must be resolved *simultaneously*, a situation we argue is more akin to the natural task faced by language learners. The lower level learning challenge we investigate is the acquisition of auditory categories. Spoken language learners face the problem of having to categorize acoustically variable vocal utterances into functionally equivalent items (e.g. phonetic categories). The higher level learning challenge we chose to investigate is the acquisition of transitional probabilities defined *across* auditory categories that contain multiple acoustically-varying exemplars. To this end, we used a set of auditory categories (Wade & Holt, 2005), each containing acoustically-varying nonspeech exemplars that are novel to adult listeners. Participants heard familiarization streams characterized by transitional probabilities defined across these *experimenter-defined categories*. We then probed the nature of statistical learning resulting from this exposure. We contrast two broad theoretical possibilities: 1) statistical learning is primarily “top-down” – that is, statistics at higher levels can be optimally learned (see Section 3),

regardless of how lower level information is processed; 2) statistical learning is primarily “bottom-up” – that is, the processing of lower level information determines the statistics of higher levels regardless of whether those statistics are optimal at higher levels.

1.1 Statistical Learning across Levels of Description in Language

Statistical information has the potential to aid language learning at many different levels of description, including speech sound categorization, word segmentation and lexical development, and syntactic processing (for reviews, see Romberg & Saffran, 2010; Saffran & Thiessen, 2007). Most of the research identified by the term “statistical learning” has focused on the use of transitional probabilities to accomplish word segmentation and lexical development. To illustrate, take the phrase “pretty baby” (Saffran et al., 1996) which would typically be produced as a continuous utterance /prɪˈtɪbeɪˈbi/. In this phrase, as well as the ambient language, transitional probabilities—the probability that one will encounter Y given X—and other statistical regularities (e.g. co-occurrence frequency) are higher for syllables that cohere to form a word (e.g. “pre” and “ty”) than syllables that cross word boundaries (e.g. “ty” and “ba”). Not only can infants use transitional probabilities to segment words from a speech stream (Saffran et al., 1996; Aslin, Saffran & Newport, 1998), but they are also more likely to use syllables linked by high transitional probability as lexical labels (Graf-Estes, Evans, Alibali, & Saffran, 2007). Thus, learning from statistical regularities likely contributes to lexical development, characterized in part by the vocabulary explosion beginning around 14-months (Newman, Ratner, Jusczyk, Jusczyk, & Dow, 2006; Romberg & Saffran, 2010; Saffran & Thiessen, 2007).

However, the experiments that populate this literature and the corresponding models typically ignore a key challenge infants face in dealing with natural speech input: transitional probabilities are necessarily accumulated over many different utterances of the same phrase (e.g. “pretty baby” or /prɪˈtɪbeɪˈbi/), and it is well known that, even within the productions of a single speaker, there exists a large amount of acoustic variability across the multiple utterances of any given linguistic unit (Lieberman Cooper, Shankweiler, & Studdert-Kennedy, 1967; Peterson & Barney, 1952)¹. As such, the utterance “pretty baby” can be described as a sequence of unique exemplars drawn from the functional categories that specify the words at a more abstract, linguistic level of description.

Most statistical learning studies eliminate the perceptual variability across successive experiences by employing a corpus of acoustically identical repetitions of sounds (e.g. Saffran et al., 1996). Thus, this work includes the assumption that acoustic variation is already resolved, likely through the process of speech categorization—a phenomenon in which variable utterances within the same functional speech categories are treated equivalently—before transitional probabilities across the speech categories are learned.

Although literate adults can easily transcribe words and phrases as strings of symbols reflecting these functional categories, it is not at all clear that the perceptual abilities infants

¹While this paper focuses on acoustic variability, there are many other sources of information that vary across utterances including contextual and visual information and interaction with a caregiver. All of these factors have been shown to modify cognitive processing in infancy and thus variation of these factors will alter the informational content of each utterance.

bring to language learning (e.g. that might enable them to treat the /i/ at the end of “pretty” as an instance of the same category as the /i/ at the end of “baby,” or even as the end of “pretty” spoken by a different speaker). This issue is complicated by broader questions in the field of whether phonemes are the appropriate “unit” of speech processing. In the case of most speech sounds, there is a lack of sufficient acoustic cues available to reliably categorize and discriminate phonemes (Lotto, 2000, Peterson & Barney, 1952, Shankweiler, Strange, & Verbrugge, 1997). For example, at the acoustic level, there is no clear boundary between the entire array of possible utterances intended to be one speech sound (e.g. the syllable “ra”) and the array of possible utterances intended to be a different, but acoustically similar, speech sound (e.g. the syllable “la”). The distinction between all utterances of “ra” vs. “la” is defined only probabilistically, as is the case for many other speech sound contrasts. However, researchers from a broad spectrum of theoretical positions have questioned the primacy of overt, phoneme-level categorization in language use (Port, 2007; Lotto & Holt, 2000, Goldinger, 1998). The use of temporal windows of integration greater than the phoneme (Nusbaum & Henly, 1992; Poeppel, Idsardi, & Van Wassenhove, 2008) and dependence on communicative context (Flynn & Dowell, 1999; Obleser & Kotz, 2010) in successful speech comprehension are commonly observed. In extreme cases, people are even capable of understanding speech when the typical cues to phoneme identity are entirely unavailable (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Remez, Rubin, Pisoni, & Carrell, 1981). Thus, the issue of how functional equivalence in speech is used productively in higher-order cognitive processes — such as learning from transitional probabilities across syllables — remains an unsolved problem in language research.

There is evidence that reliable speech sound categorization is not fully developed until well beyond the age during which infants begin to segment and learn words. While it has been demonstrated that infants in the first year begin to preferentially discriminate the acoustic contrasts employed in their ambient language (e.g. Werker & Tees, 1984; Kuhl et al., 1992), it is also known that the development of speech sound categorization continues well beyond infancy (e.g. Hazan & Barrett, 2000; McGowan, Nittrouer, & Manning, 2004; Serniclaes, Van Heghe, Mousty, Carre, & Sprenger-Charolles, 2004). These results indicate that the emergence of speech sound categorization has a largely overlapping developmental time-course with infants’ abilities to use transitional probabilities between syllables to segment and learn words. Even in the case where infants show evidence of adult-like phoneme categorization and differentiation in isolation, they can fail to differentiate these sounds during language learning tasks such as word learning (e.g. Stager and Werker, 1997; Thiessen, 2007). Thus, even early demonstrations of infants’ speech sound categorization abilities may be highly task-dependent and might not be functionally applied when learning transitional probabilities between syllables or when acquiring other language skills (Zevin, in press).

The empirical evidence suggesting fragility of early infant speech categories, combined with the overlapping developmental trajectories of speech category acquisition and early lexical development, cast doubt on the assumption that speech sound categorization is resolved before infants can learn from higher order statistics across speech sounds. Instead, infants must be able to learn something about higher levels of description without the benefit of well-developed or supportive categorization abilities at lower levels of description. While

each aspect of this developmental challenge has been extensively examined separately (i.e. acquiring speech sound categorization and learning transitional probabilities between syllables), learning at a higher level of description without an assumed resolution of lower level regularities has largely gone unexamined.

1.2 The Current Paper

We investigate learning from regularities at a higher level of description, such as across sound categories, without assuming that perceptual variability within categories has already been resolved. To model the challenge faced by infant language learners who must learn transitional probabilities across speech categories while their speech sound categorization abilities are still developing, we gave adult learners exposure to four novel sound categories containing spectrotemporally complex nonspeech sounds. We examined a very simple model language where the four novel sound categories are organized into sequentially presented pairs, which we refer to as *sound-pairs*. These sound-pairs are presented in an auditory stream such that higher-level transitional probabilities, defined across these four experimenter-defined categories, are 100% for the sound categories within a sound-pair and 50% for the sound categories spanning two different sound-pairs.

One novel aspect of this experiment is that each sound category is instantiated in multiple, variable exemplars designed to model the natural acoustic variability that exists for any given linguistic utterance in human speech. Critically, the transitional probabilities across sounds in a sound-pair are substantially higher as defined at the category-level compared to the exemplar level. Within sound-pairs, no specific exemplar of the first category predicts any other specific exemplar from the second category. Thus, to reliably learn the intended higher-level transitional probabilities, participants would need to resolve the within-category acoustic variability through grouping or categorizing sounds. This task would be trivial if participants already had the ability to categorize or perceptually group the exemplars into the appropriate four experimenter-defined categories. However, the stimuli were designed to challenge participants' naïve perception of these four categories. Specifically, there are two "Easy" categories that have an invariant acoustic cue that can distinguish all exemplars from a given category, but also two "Hard" categories that do not have such an invariant cue and thus are less likely to be distinguished. We directly assess participants' naïve perceptual organization of the stimuli in Experiment 1.

This scenario creates a tension between participants' lower-level perceptual organization of sounds and the organization that they would need to have in order to learn from the intended higher-order statistical regularities. In a series of four experiments, we considered two broad theoretical possibilities for how learning might proceed. The first possibility is that statistical learning is "top-down." In this case, participants would be able to learn from higher-order statistical regularities even if their initial lower level perceptual organization does not support it. This would result in the most robust learning of the experimenter-intended transitional probabilities defined at the higher level. The second possibility is that statistical learning is "bottom-up." In this case, lower level organization is relied upon to define statistical information at higher levels. Since participants' naïve lower-level organization does not support learning the intended higher-order transitional probabilities in the current

experiment, participants will not have, or develop over the course of exposure, the best possible prediction of successive sounds or uncover the intended structure of the model language.

2. Experiment 1: Naïve perception of novel, spectrally complex non-speech sounds

Before examining how learning proceeds, we first examined in detail participants' naïve perception of the sound categories used in subsequent learning studies. Specifically, listeners heard six different exemplars in each sound category (stimuli are plotted in Figure 1). By *sound category*, we mean an experimenter-defined category of acoustically varying sounds. We will use the term *sound exemplar* to refer to a single sound heard by the participants during the experiment.

The novel sound categories were adapted from the training study of Wade and Holt (2005), which used a video game to implicitly train listeners on categories of novel, spectro-temporally complex non-speech stimuli. These sounds are carefully designed to capture key characteristics of natural speech categories without sounding speech-like (to listen: http://www.psy.cmu.edu/~lholt/php/gallery_irfbats.php). While previous research has shown that it is possible to develop some degree of specialized processing of these sounds with adequate exposure and training (Leech et al., 2009; Liu & Holt, 2011), participants in the present study have not experienced these sounds before and thus should not, *a priori*, categorize them into the four experimenter-defined categories. In fact, participants should have no specialized processing for these sounds. Thus, we are tapping into their naïve perception.

To characterize participants' perceptual organization of these sound categories, we asked naïve participants to perform a perceptual similarity judgment. Results from this task uncover how participants perceive the sounds in relation to each other. These judgments are defined both by the raw acoustic similarities in one or more cues characterizing the sounds, and by how much importance each particular cue carries for the listener. When analyzed with multidimensional scaling, similarity judgments permit us to characterize the *perceptual organization* or *space* that participants have for these stimuli. Stimuli that are perceived as similar tend to cluster tightly together in perceptual space, and are linearly separable from stimuli in other regions of perceptual space.

2.1 Methods

2.1.1 Participants—28 students participated in the current experiment (age: $M = 19.64$ years, $SD = 1.10$ years; 1 ambidextrous, 9 male). All participants reported in this paper were undergraduates at Cornell University who participated in exchange for course credit. Participants were asked to report any auditory, visual, or neurological deficits via post-experimental questionnaire; no participants reported any such deficits.

2.1.2 Sound Stimuli—We employed four categories adapted from Wade and Holt (2005), two of which were easily discriminable from each other (based on a single acoustic dimension, referred to as “Easy” categories) and two of which were not easily discriminable

based on a single acoustic dimension (referred to as “Hard” categories). All four sound categories were designed to have two spectral peaks, P1 and P2, both of which have a steady-state frequency portion and a transitioning frequency portion, similar to the acoustics of syllables containing a vowel and a semivowel or liquid (for a schematic diagram of the four categories of stimuli, see Figure 1). The two “Easy” categories (E1 and E2) each have a P1 that begins with a period of steady state frequency and then gradually decreases in frequency. Both Easy categories also have a P2 that begins with a steady state period and either consistently increases (E1) or decreases (E2) in frequency. In principle, the “Easy” categories are discriminable from each other based on the direction of the frequency transition of P2. The two “Hard” categories (H1 and H2) have a P1 that begin with a frequency transition followed by a steady state. They also both have a P2 that begins with a frequency transition followed by a steady state, but both H1 and H2 completely overlap in their steady-state frequencies *and both* contain rising and falling transitions within-category. Thus, only the interaction between two cues, the onset frequency and the steady state frequency, creates a perceptual space in which H1 and H2 are discriminable from each other (see Wade & Holt, 2005, for a comprehensive discussion). Note that it is reasonable for all stimuli from Hard categories to be easily discriminable from those of Easy categories because the P1 components have different carrier waves for Hard (sawtooth) vs. Easy (noise) sounds as well as different structures.

Indeed, the learning curves of participants trained to categorize these sounds (Liu & Holt, 2011; Wade & Holt, 2005) suggest that it is easiest to discriminate between the Easy stimuli and the Hard stimuli, and between the two Easy categories. Participants learn relatively quickly to categorize E1 and E2 sounds well above chance in the training paradigm, but substantially more training is necessary to produce above-chance categorization of sounds in H1 and H2. We employed six different exemplar sounds from each of the four categories, all of which share a common P1 but diverge in their P2 (Figure 1).

2.1.3 Stimulus presentation—All sounds were presented using over-the-ear headphones (Sony MDR-V150) at a comfortable, above-threshold volume. Instructions and stimuli were presented using PsyScope X B53 on MacMini computers. During sound presentation, participants observed blank, white screens on 17in CRT monitors. Sounds had a duration of 300 ms each. Each trial began with a 500 ms silent interval, followed by the first sound, and then, after a second 500 ms inter-stimulus interval, the second sound. After a 500 ms silent interval, participants were then prompted to respond.

2.1.4 Similarity Judgment—After hearing both sounds, participants were asked to report how similar the sounds were on a scale of 1 to 4 (1 = the same and 4 = completely different) on a keyboard. Participants were given an unlimited amount of time to make their responses.

For practical purposes, it was necessary to limit the number of trials by partitioning the full set of 24 exemplars (six from each of the four categories corresponding to exemplars 1, 3, 5, 7, 9, and 11 from the Wade and Holt stimuli, which we will refer to as exemplars 1 through 6 in the context of this study, e.g. E1₁, E1₂, ... E1₆) into two subsets; one subset contained exemplars 1, 3, 5 from each category and the other subset contained exemplars 2, 4, 6. Half

the participants performed similarity judgment on one subset, the other half on the other subset.

We performed Multi-Dimensional Scaling (MDS) of the aggregate data across participants using the IsoMDS routine, part of the MASS library for R (Venables & Ripley, 1997). We restricted the number of dimensions to two. Because the two participant halves experienced entirely disjoint sets of stimuli, in order to run the MDS analyses across all participants, perceptual similarity judgment values were interpolated between stimulus tokens for which they were collected (e.g. if judgments were collected for exemplars 1, 3, and 5, the value for exemplar 2 was interpolated between the judgments for exemplars 1 and 3, and the value for exemplar 4 interpolated between 3 and 5). This provides some overlap between the stimulus judgments across the two sets of participants that makes the group-level MDS analyses possible.

Additionally, ANOVAs were run on raw similarity scores, treating subjects and items (sound exemplars) as random factors to complement the MDS results. *Min F'* will be reported whenever both subject and item analyses were significant.

2.2 Results and Discussion

The MDS results of the aggregate data from all participants, presented in Figure 2, indicate that participants are able to perceive the Easy and Hard categories as distinct from each other. This is revealed through the linear separability of E1, E2 and H1, H2. The figure also reveals a distinction within the Easy categories: the exemplars from E1 are linearly separable from E2. This pattern of results indicates that participants also perceive a difference between the two Easy categories. E1 occupies a distinct region of perceptual space from E2, such that each E1 stimulus is on average more similar to all other E1 stimuli than to the E2 stimuli. This is not true for the H1 and H2 stimuli, which are more or less evenly distributed across the second perceptual dimension.

We then analyzed raw similarity scores in a series of ANOVAs. First, we tested whether participants separate E from H categories by examining the resulting judgments of three possible trial types: comparison within the Easy categories (e.g. E1₁ vs. E2₃, E1₁ vs. E1₅), within the Hard categories, and a comparison between Easy and Hard categories (e.g. E1₁ vs. H2₃). This analysis revealed a significant effect of trial type, *min F'* (2, 234) = 22.088, $p < 0.0001$. The means for these contrasts are higher for between-category types (E vs. H: 2.854) than within either E or H category trials (E vs. E: 2.243, H vs. H: 2.200) demonstrating that exemplars between Easy and Hard categories are rated as *less similar* than exemplars within either Easy or Hard categories.

Next, we investigated how participants perceive the four experimenter-defined categories (i.e. E1, E2, H1, H2), by crossing comparison type (Same vs. Different category) with category type (Easy vs. Hard). A significant interaction between these factors was observed, *min F'* (1, 119) = 27.955, $p < 0.0001$, such that there was a large difference in perceived similarity for Same vs. Different comparisons that was restricted to the Easy categories. That is, consistent with the MDS analysis, participants perceived the Easy items as forming two clusters roughly corresponding to the experimenter-defined categories, whereas the Hard

items were equally similar between and within these categories. There also was a main effect of Same-Different judgments, $\min F' (1, 119) = 25.509, p < 0.0001$ but no main effect of Easy-Hard trials. A planned comparison revealed that the effect of comparison type was significant only for stimuli from the Easy categories: $\min F' (1, 73) = 90.738, p < 0.0001$.

In sum, MDS (Figure 2), and ANOVAs on raw similarity scores reveal that participants perceive Easy and Hard categories as distinct and they further perceive the difference between the two Easy categories (E1 and E2) but do not perceive the difference between the two Hard categories (H1 and H2). These results confirm that participants do not perceive all sounds equivalently and that naïve perceptual organization does not conform to the experimenter-defined categories. Having characterized how participants naïvely cluster the exemplars from the current task, we next examine learning based on multiple exemplars of these novel sounds.

3. Experiment 2: Statistical learning across multiple exemplars of novel sound categories

Experiment 1 showed that participants' perceptual organization of the sound categories does not conform to the four experimenter-defined categories used to create the model language. The current learning paradigm creates an interrelationship between lower level perceptual organization and higher-order statistical information, and we examine whether statistical learning is more “bottom-up,” where perceptual organization determines what is learned, or “top-down,” where statistical learning proceeds based on the four experimenter-defined categories and participants achieve the best prediction of successive sounds.

Unlike previous statistical learning experiments, the current learning paradigm presents a novel method that creates an interrelationship between lower level perceptual organization and statistical information defined at a higher level of description. Specifically, participants are exposed to sound-pairs (e.g. E1-H2) where each sound category is instantiated in multiple, variable exemplars (e.g. E1₁, E1₂, ... E1₆). The inclusion of acoustic variability more closely simulates the ecological challenges faced by language learners, and highlights this interrelationship between learning at different levels of description that exists during language acquisition. The current paradigm makes learning based on the individual sound exemplars extremely unlikely. First, there are a large number of exemplars (24 exemplars across four categories of sounds). Second, each exemplar from one sound category is presented with all others in its paired category (e.g. if the sound-pair is E1-H2, then E1₁ is paired with all six H2 exemplars, E1₂ is paired with all six H2 exemplars, etc.). A single exemplar (e.g. E1₁) thus predicts six other exemplars with equal likelihood, and any given specific exemplar pair (e.g. E1₁-H2₂) is only presented together nine times during familiarization. In contrast, there are learnable transitional probabilities across sound categories (E1-H2) that occur 324 times for each sound-pair. Given the relative reliability of category-level statistics compared to exemplar-level statistics, if participants successfully learn the sound-pairs, they must learn based upon transitional probabilities calculated across groupings of exemplars and not across individual exemplars.

While the intended structure of the stream is based on the experimenter-defined categories (E1, E2, H1, H2), Experiment 1 established that participants' perceptual organization only supports three perceptual groups. E1 and E2 categories are separated in perceptual similarity space while the two remaining categories are not perceived as distinct, creating an omnibus H cluster. Given that participants do not readily distinguish the four experimenter-defined categories, this raises the following question: if participants are able to learn from higher-level statistical information defined across acoustically variable exemplars, does listeners' naïve perceptual organization determine what statistics are learned ("bottom-up"), or will participants be able to learn on-line according to the experimenter-defined categories because that yields greater statistical predictive power ("top-down")?

At this point, it is important to consider whether the goal of the learner is more in line with one of these theoretical outcomes than the other. We view the goal of the learner as two fold. The first goal is to uncover the intended structure. This goal is especially relevant in language learning where language structure is shared amongst members of a community and a learner must decipher the language such that they will be able to use these structures to communicate effectively to one another. However, it is unclear how a learner is able to determine what the intended structure is, beyond what they are able to learn (e.g. the presence of statistical information). It is possible that interaction between language learners and "teachers" (e.g. caregivers) is organized in such a way as to scaffold a direction of learning (e.g. sentence frames support word recognition, Fernald & Hertado, 2006; speaking in variation sets to support more effective word segmentation and verb learning; Onnis, Waterfall & Edelman, 2008; Waterfall, 2006). The current experiment doesn't provide this kind of interactive, guided experience for the learners. They are simply exposed to the raw statistics based on a randomized stream of sound category pairs.

The second, more cognitively reductionistic goal, is to have the most accurate prediction of the next sound. Learning can be broadly viewed as a progressive reduction of uncertainty between internal predictions and the outcomes of events (e.g. Schultz, Dayan & Montague, 1997; McClelland, 2002). Pursuing this goal would result in statistical learning and other forms of implicit learning, as uncovering structure through statistics allows for better prediction of successive experiences. This goal makes use of predictive processing to both reduce cognitive effort in processing correctly anticipated events and provide an error signal for incorrect predictions. This view has been widely applied in the domain of language processing (Altmann & Mirkovi, 2009; Kamide, Altmann, & Haywood, 2003) while such a link remains relatively unexplored in the domain of language learning.

Beyond the goal of learning the intended structure of the ambient language, if we characterize the goal of the learner as the best prediction, the learner should uncover the experimenter-intended structure of the model language (i.e. use the statistics based upon 4 sound categories). If a learner predicts the next sound based on any 6 exemplars from a given sound category, the learner has a 1/6 chance of correctly guessing the next sound and is predicting based on a smaller region of perceptual space. If the learner predicts based on the 3 categories in the naïve perceptual organization, the participant has a possible 12 exemplars to predict in the omnibus Hard category. Thus, there is less effective prediction when statistical information is determined is based on the naïve perceptual organization. An

alternative possibility is that participants could predict based on the category and not the successive sounds. If participants predict based on sound category, it would be advantageous to have larger categories thereby increasing the chance of a correct prediction. If there were no penalty for predicting based on larger categories, the best strategy for the learner would be to place all sounds into a single category, thus always correctly predicting the next category of sound. This strategy would result in no discovery of structure or statistical learning, and would not be beneficial in reducing processing costs as the predictions are too general. Also, there is no added benefit from predicting using groups of sounds that are smaller than the experimenter-defined categories. It is also not possible to gain better prediction by having groups smaller than the experimenter-defined categories. As summarized previously, there is no reliable information at the exemplar level, i.e. an individual exemplar will only predict the next experimenter-defined category of 6 possible exemplars. Thus, to gain the best possible prediction of the successive utterance in the model language, participants must use the experimenter-defined categories.

We examined the manner in which participants *actually* group sound exemplars in order to learn statistics across them. We will refer to these groups used to learn as a *functional grouping*. We differentiate functional grouping from naïve perceptual organization—even though participants group stimuli in a perceptual similarity task, this does not entail that these are the groups that participants will use to track statistical regularities at a higher level of description (e.g. infants difficulty using minimal pairs in a word learning task compared to a simple discrimination task; Stager & Werker, 1997). If statistical learning is more “bottom-up,” participants are expected to assume the three functional groups that comprise their naïve perceptual organization (E1, E2, and the combined ‘H’ category, as revealed in Experiment 1); however, if statistical learning is more “top-down,” participants are expected to use four functional groups that match the experimenter-defined categories (E1, E2, H1, and H2) to learn.

We are able to assess which functional grouping participants actually use to learn because using different functional groupings results in different higher-order statistical information in our learning paradigm. Specifically, there are two ways in which the learning paradigm can change based on differences in functional groupings. First, because the sound exemplars must be grouped to reveal higher-order statistical information, assuming different functional groupings will result in different higher-order transitional probabilities learned during exposure. As will be addressed in detail below, there are differences in both within and between-pair transitional probabilities depending on whether participants employ a “bottom-up” strategy using their naïve perceptual organization or a “top-down” strategy consistent with the four experimenter-defined categories. Second, as in previous statistical learning experiments, we examined whether participants can distinguish pairs heard during familiarization from novel pairs of sounds that violate the structure of the training exposure, known as *foils*. Different functional groupings modulate how readily a participant can discriminate a sound-pair from a foil during test. In short, we can use differences in patterns of learning across conditions/foils as a probe for participants’ functional groupings and, thus, assess whether statistical learning is more “bottom-up” or “top-down”.

3.1 Methods

3.1.1 Participants—Eighty-five students participated in the present experiment (age: $M = 19.42$ years, $SD = 1.11$ years; 9 left handed; 31 male). All participants received identical familiarization procedures, the only difference across participants being the specific sound-pairs they were familiarized on. In the post-familiarization behavioral test, there were two between-subject conditions that differed in the nature of the “foils” against which familiar sound-pairs were compared. Forty-five participants were assigned to the condition in which sound-pairs were pitted against *Type 1 foils*, and 40 participants were assigned to the condition in which sound-pairs were pitted against *Type 2 foils*. One participant in the Type 1 foil condition was excluded for failing to complete the entire experiment.

3.1.2 Sound-Pair Assignments—For each participant, the four experimenter-defined sound categories were grouped into two pairs (e.g. participant 1: E1-H2, H1-E2, participant 2: E1-E2, H2-H1, etc.), where across participants all possible pairings are represented. We will refer to any given set of pairings as a *Sound-Pair Assignment*. While each participant is only given one Sound-Pair Assignment, all possible Sound-Pair Assignments are used across participants in the experiment.

There are three basic structures that Sound-Pair Assignments could take: (1) E-E/H-H, (2) E-H/E-H, and (3) H-E/E-H. Within each of these structural classes, there are multiple *specific* Sound-Pair Assignments that are possible (for example, the E-E/H-H class can manifest in a sound-pairing of E1-E2/H1-H2, a sound-pairing of E2-E1/H2-H1, etc.). However, it is important to note that the transitional probabilities (of sound-pairs and foils) and learning predictions that are borne out of all the Sound-Pair Assignments in a particular structural class are equivalent. Thus, in our analyses, we collapse across the performance of all participants who received a particular class of Sound-Pair Assignments (e.g. participants who receive E1-E2/H1-H2 pairings and E2-E1/H2-H1 pairings are lumped together). Table 1 shows the three distinct Sound-Pair Assignment structural classes and one representative Sound-Pair Assignment example from that structural class.

3.1.3 Foil Types—Foils were constructed based on the four experimenter-defined sound categories to have zero transitional probability between sounds under the four-category assumption. Two different types of foils were tested against the sound-pairs in a between subjects design whereby each subject received either the comparison between sound-pairs and *Type 1 foils* or the comparison between sound-pairs and *Type 2 foils*.

Type 1 foils were constructed by creating two sequences as follows: (1) the initial item of the first sound-pair followed by the initial item of the second sound-pair, and (2) the latter item of the first sound-pair followed by the latter item of the second sound-pair. The “first” and “second” sound-pairs can be interpreted based on their arrangement in the left-most column of Table 1.

Type 2 foils were constructed by creating two sequences as follows: (1) the initial sound of the first sound-pair followed by the latter sound of the second sound-pair, and (2) the initial sound of the second sound-pair followed by the latter sound of the first sound-pair. See Table 2 for a concrete example of how Type 1 and Type 2 foils were generated for a

particular Sound-Pair Assignment class (E-H, E-H). The foils for the other Sound-Pair Assignment classes (E-E, H-H and H-E, E-H) were generated according to the algorithm exemplified in Table 2.

There were two distinct types of foils compared against sound-pairs because the different types of foils, interacting with the different Sound-Pair Assignments, generate distinct predictions depending on whether the participant assumes four functional groups (consistent with statistical learning being more “top-down”) vs. three functional groups (consistent with statistical learning being more “bottom-up”) and thus permits a direct test that distinguishes between these accounts (see Table 2).

3.1.4 Familiarization—Sounds were presented for 300 ms each with a 115 ms inter-stimulus interval (ISI). The familiarization stream comprised 648 pairs of stimuli generated by presenting each exemplar from each category with each exemplar from its paired category nine times. Sound-pairs (in terms of experimenter-defined categories, e.g., E1-H1) were presented in a pseudorandom order so that each pair type was equally likely to be followed by either itself or the other pair. The right-most column of Table 1 presents an example familiarization stream that might result from that specific Sound-Pair Assignment. This familiarization creates a distribution of transitional probabilities presented in Table 3 where within sound-pair transitional probabilities are very high (considering the case where four experimenter-defined categories are used, top row) and between sound-pair categories are much lower (50%). The duration of the familiarization period was approximately 9 minutes.

3.1.5 Cover Task—In order to encourage participants to pay attention to the familiarization stream without explicitly asking them to track the relationships between sounds heard, a cover task was employed in which participants had to detect the stimuli with attenuated volume. Six instances of each sound were presented at attenuated volume (without disturbing the sequence) for a total of 144 stimuli (of the 1296 sound presentations). Participants were instructed that they would hear a stream of sounds and to press the space bar when they heard the stream get quieter. Button presses within 1.3 secs of presentation of the attenuated sound were considered a correct response.

3.1.6 Test for Statistical Learning—After familiarization, participants were given a self-timed break. After the break, they were told that they would be presented with two pairs of sounds separated by a long pause (1000 ms) and, after hearing both, they would be asked to report which pair of sounds is more familiar based on their previous task. They used the ‘g’ and ‘h’ keys to indicate which pair was more familiar. They were also told that no new sounds would be introduced and encouraged to go with their intuition or “gut instinct”. The responses were self-timed.

All participants were given 48 test trials. In each trial, participants heard one sound-pair (consistently experienced during familiarization) and another pair that was either a Type 1 or Type 2 foil. Half of the participants heard sound-pairs vs. Type 1 foils only, and the other half heard sound-pairs vs. Type 2 foils only. All exemplars within the sound categories that

were experienced during familiarization were heard in the test phase, and each pair was compared to each foil an equal number of times and with the order counterbalanced.

3.1.7 Perceptual Similarity Judgment—After completion of the statistical learning test, participants were asked to perform a post-experiment perceptual similarity judgment, as described in Experiment 1.

3.2 Results

3.2.1 Cover task results—Participants across both foil-type conditions responded correctly to the ‘soft’ sound with an average of 75% accuracy. No participants were excluded based on Cover Task performance.

3.2.2 Perceptual Similarity Judgments—The MDS analyses of the aggregate data across participants in both foil-type conditions of Experiment 2 reveal that participants make a distinction between E1 and E2 categories but not between H1 and H2, similar to results seen in Experiment 1 (Figure 2). The same inferential tests were employed as those reported for Experiment 1 and together with MDS analyses, a striking similarity is found. See Appendix 2 for specific analyses and MDS plots.

3.2.3 Test for Statistical Learning: Overall—We begin by examining the post-familiarization test results for evidence of overall statistical learning. Performance in the test was evaluated against chance (24 out of 48, or 50%). Results indicate that participants are able to reliably distinguish the sound-pairs heard during familiarization from foils: Type 1 foil condition, mean performance = 27.93 (58.2%), $SD = 6.65$, $t(43) = 3.93$, $p < 0.001$; Type 2 foil condition, mean performance = 26.4 (55%), $SD = 4.63$, $t(39) = 3.31$, $p < 0.01$. Thus, participants in both foil-type conditions demonstrated sensitivity to the statistical information presented during familiarization.

3.2.5 Test for Statistical Learning: Pattern of Learning by Sound-Pair

Assignments—In subsequent analyses, participants were organized according to three types of Sound-Pair Assignment structures, summarized above, in order to examine whether learning is more “bottom-up” where participants assume 3 functional groups corresponding to their naïve perceptual organization or a “top-down” where participants assume the 4 experimenter-defined categories as their functional groups (see Table 1). To do this, we derived different learning predictions for each Sound-Pair Assignment structure using statistical information determined based on the 4 experimenter-defined categories (“top-down”) and the 3 perceptual groups seen in the naïve organization (“bottom-up”). These predictions are presented in the bottom panel of Figure 3. This section presents a brief overview of the predictions. For an in-depth treatment of how these predictions were derived, see Appendix 1.

Predictions of learning outcomes were made based upon the transitional probabilities during familiarization as well as those of the foils during test. For familiarization, the robustness or reliability of statistical information indicating the sound-pair structure was considered as the *relative* reliability of the within- and between-pair transitional probabilities. The statistical information is considered more robust if there is a large difference of within and between-

pair transitional probabilities. However, as the values of within- and between-pairs become quantitatively closer, the statistical information is considered to be less reliable, and we predict that participants will not show evidence of learning during test. For the present experiment we assumed that, for learning to take place, within-pair transitional probabilities must be *much higher* than the between-structure transitional probabilities (see Appendix 1 for more detailed reasons behind this assumption). There is one Sound-Pair Assignment structure (H-E, E-H) that we considered to have unreliable statistical information at familiarization, under the bottom-up assumption, because one of the within-pair transitional probabilities was at 50% (third column of Table 3). Thus, we did not expect reliable learning to emerge for this Sound-Pair Assignment. If, however, statistical information *is* reliable during familiarization (as is true for the other Sound-Pair Assignment structures), we then considered the relationship between the within-pair transitional probabilities and the transitional probabilities of the foils employed at test.

The transitional probabilities of the foils were calculated relative to the familiarization stream (i.e. how likely is that combination of sounds given familiarization experience). If the transitional probabilities of the foils are lower than the within-pair transitional probabilities, then we predict participants will be able to discriminate pairs from foils at test and show evidence of learning. However, if the foil transitional probabilities are not lower than the within-pair transitional probabilities, we predict that it will not be possible for participants to distinguish pairs from foils at test and thus, they will not show evidence of discrimination at test.

We employed two different types of foils to produce two distinct patterns of predictions (see Table 3 for a summary of the transitional probabilities for within-pairs and the foils employed during test). It is important to note that the between-pair transitional probabilities are higher than typical statistical learning experiments (around 50%). This is because the structure of the stream is simple (two pairs of sounds; see Kirkham et al., 2002 for a study with similar statistical design).

After deriving these predictions for learning based on sound-pair assignment, we examined how these predictions relate to the pattern of learning performance. First, if participants assumed a structure based on the four experimenter-defined categories (“top-down” statistical learning), learning would be uniform across sound-pair assignment structures. Thus, there should be no effect of sound-pair assignment (see first row of the table in Table 3 and Figure 3). However, for both Type 1 and Type 2 foil conditions, a one-way ANOVA revealed a significant effect of Sound-Pair Assignment ($F(2, 41) = 7.71, p = 0.001$ for Type 1 foils; $F(2, 39) = 5.46, p < 0.01$ for Type 2 foils).

We next examined learning in each of the three Sound-Pair Assignment general structure types (E-E/H-H, E-H/E-H, H-E/E-H). Focusing first on the results for Type 1 foils. This analysis revealed that, for sound-pair assignments E-E/H-H and E-H/E-H, participants reliably discriminated sound-pairs from Type 1 foils (sound-pair assignment E-E/H-H: mean = 28.5 (59.4%), $SD = 7.06, t(13) = 2.37, p < 0.05$; sound-pair assignment E-H/E-H: mean = 31.8 (66.3%), $SD = 5.83, t(14) = 5.18, p < 0.001$) whereas participants in sound-pair assignment H-E/E-H failed to discriminate sound-pairs from Type 1 foils: mean = 23.53

(49%), $SD = 4.27$, $t(14) < 1$, *n.s.* Comparing this pattern of results to our predictions (bottom of Figure 3), this provides evidence that participants' learning was constrained by their naïve lower level organization of the sound categories.

For Type 2 foils, again “top-down” statistical learning would result in consistent learning across sound-pair assignments, however, if statistical learning were “bottom-up,” learning would only be predicted in the E-E/H-H sound-pair assignment structure. Thus, while the previous analysis with Type 1 foils predicted learning for the E-H/E-H sound-pair assignments, the current analysis with Type 2 foils does not predict learning for this group (see Table 2 for a summary of the differences in foils for this sound-pair assignment structure). Examination of learning revealed that participants selected sound-pairs over Type 2 foils above chance for only the sound-pair assignment E-E/H-H (mean = 29.54 (62%), $SD = 4.63$, $t(12) = 431$, $p = 0.01$), but not for the other conditions, both $t_s < 1$ (Figure 3). This pattern of results is consistent only with the behavioral predictions for the 3-functional-group assumption (second row of the bottom, Figure 3).

Thus, the findings from both foil conditions of the current experiment support the hypothesis that participants assume a functional grouping consistent with their naïve perceptual similarity judgments when they engage in higher-order statistical learning.

3.3 Discussion

Participants are able to learn from the temporal structure of a sound stream comprising novel, spectro-temporally complex sounds with which they had no prior experience. This result confirms that participants are able to learn from higher level statistical information despite perceptual variability within higher level units. The model language was based on four experimenter-defined categories (E1, E2, H1, and H2). Learning based on this organization of sound exemplars would produce the best learning outcome (see discussion of the goals of the learner in Section 3). However, the pattern of learning observed suggests that participants' learning was constrained by their naïve perceptual organization of the stimuli.

4. Experiment 3: Generalizing Learning to Novel Exemplars

Experiment 2 demonstrated that participants' learning of higher-order statistical information is constrained by their lower level perceptual organization of these sounds. However, it is unclear exactly *how* participants' are using their naïve perceptual organization to learn. These adult participants have no prior experience with, and thus lack any specialized or explicit categorization abilities for, these sounds. However, we do find that participants have a non-uniform organization of perceptual space that supports the grouping of some exemplars and a failure to differentiate others. We found that this organization constrains statistical learning but does this mean that participants learn from higher-order statistical information defined based upon these *regions* of perceptual space? In the current experiment, we examine a crucial prediction of this view: if learning is indeed based on functional groups that match participants' naïve perceptual clusters, participants should be able to generalize the statistics learned across these functional groups to acoustically novel exemplars drawn from the within participants' naïve regions of perceptual space. To test this

hypothesis, we familiarized a new group of participants using identical methods to the Type 1 foil condition from Experiment 2. At test, however, both the target and foil stimuli were generated using unfamiliar exemplars from the four experimenter-defined categories. If participants are able to generalize their higher-order statistical knowledge to novel exemplars, we predict that they will exhibit the same pattern learning across Sound-Pair Assignment structures as seen for the Type 1 foil condition in Experiment 2.

4.1 Methods

Thirty-seven additional participants were recruited for this experiment (age: $M = 19.68$ years, $SD = 1.25$ years; 5 left handed; 9 male). We used the same experimental procedures as Experiment 2 for Type 1 foils, except for one crucial difference: different exemplars from the categories were introduced at test than were used during familiarization. In Figure 1, there are six exemplars shown for each of the four experimenter-defined sound categories. We used these six exemplars during both familiarization and test in Experiment 2. These were the odd numbered exemplars from the Wade and Holt study (1, 3, 5, 7, 9, and 11). In the current experiment, we employed the same six exemplars shown in Figure 1 during familiarization but used five *novel* exemplars, which were stimuli interpolated between the P2 components of the original six exemplars, in the *test* portion (c.f Wade & Holt, 2005 for more details on the acoustics of the novel exemplars). These five novel exemplars were the even numbered exemplars (2, 4, 6, 8, 10). Thus, we used these acoustically novel sounds that fit within the general structures of the experimenter-defined categories to test for generalization of higher-order statistical knowledge to novel sounds and to demonstrate that the statistics were learned at a higher across-grouping level as opposed to across acoustically specific exemplars only.

To verify that these novel exemplars are in fact grouped in the same manner as the familiarization exemplars, participants performed the same perceptual similarity test as previous experiments with the novel exemplars. In order to reduce the length of perceptual similarity task, each participant received a subset of exemplars for each category to compare with all other selected exemplars. Because there was one fewer novel exemplar, the middle exemplar (6) was used in both subsets (either 2, 6, 8 or 4, 6, 10).

The statistical learning test was identical to the Type 1 foil condition of Experiment 2, with the exception of using the novel exemplars and a reduction of total test trials due to the smaller number of novel exemplars (40 test trials). Thus, if participants are able to generalize their knowledge to novel exemplars which occupy the same perceptual similarity space, we should observe the same pattern of results as in the Type 1 foil condition of Experiment 2: we predict above-chance learning in Sound-Pair Assignments E-E/H-H and E-H/E-H but not in the assignment H-E/E-H.

4.2 Results

4.2.1 Cover Task—Participants performed the cover task with an average of 78% accuracy. No participants were excluded based on cover task performance.

4.2.2 Perceptual Similarity of Novel Exemplars—Participants provided perceptual similarity judgments for the stimuli used in the test phase of the statistical learning experiment. Results from the perceptual similarity judgments are consistent with our assumption that the novel exemplars are grouped in the same way as the exemplars tested in Experiments 1 and 2. See Appendix 2 for specific analyses and MDS plot.

4.2.3 Test for Statistical Learning—Test performance is consistent with the hypothesis that participants are able to generalize knowledge to novel sound exemplars. Overall, participants demonstrate significant learning, mean = 22.4 out of 40 (56%), $t(36) = 2.044$, $p = 0.048$. The predictions for learning according to Sound-Pair Assignment and level of perceptual grouping are identical to the Type 1 foil condition in Experiment 2 (see the left panel of Figure 3). Broken down by Sound-Pair assignment, significant learning was observed in Sound-Pair Assignment E-E/H-H (mean = 24.08 (60.2%), $t(11) = 2.512$, $p = 0.029$) and in Sound-Pair Assignment E-H/E-H (mean = 25.31 (63.3%), $t(12) = 2.284$, $p = 0.041$). For Sound-Pair Assignment H-E/E-H, performance was marginally below chance (mean = 17.5 (43.8%), $t(11) = -2.147$, $p = 0.055$).

4.3 Discussion

The current experiment essentially replicates Experiment 2 (using one type of foil), demonstrating that statistical learning during familiarization generalizes to stimuli drawn from the same regions of perceptual space. Generalization to new sounds is a crucial prediction of the hypothesis that participants use regions in their naïve perceptual organization to learn statistical regularities defined at a higher level of description in a stream of novel sounds.

There is one notable difference in the current set of results: for one sound-pair assignment, we find a marginally lower than chance performance rather than finding performance at chance. The learning predictions for generalization were to have no learning for this sound-pair assignment. While any deviation from chance (either greater than or less than chance) could be considered evidence of knowledge of the structure of the sound-pairs, it is difficult to explain how knowledge of the sound-pairs would result in a systematic rating of non-familiarity at test. Thus, we interpret this result as evidence that learning did not take place for this sound-pair assignment, similar to the finding of chance-level performance.

5. General Discussion

The current study examined a scenario where participants' lower level perceptual organization does not readily support the learning of experimenter-defined higher order statistics. This novel statistical learning paradigm captures an important feature of the task faced by the infant language learner: infants learn the words of their native language based in part on statistical information defined over sequences of speech categories, before their ability to categorize speech sounds is fully developed. Our goal was to examine whether statistical learning is constrained by lower level organization ("bottom-up") or if learning will proceed despite the current lower level organization, uncover the intended structure, and produce the best prediction of successive sounds ("top-down").

To our knowledge, this current study is the first investigation of statistical learning where participants must resolve information across multiple levels of description. Specifically, the use of variable exemplars of novel sound categories created a tension between the participants' naïve lower level perceptual organization of the sounds and the experimenter-defined higher-order statistical information. While two recent studies employed variable exemplars of visual stimuli, these stimuli were drawn from categories that participants have had considerable experience with: Brady and Oliva (2008) used visual scenes organized by well-learned semantic categories (e.g. kitchens and beaches), while Baldwin et al., (2008) used pictures of common actions such as pointing and grasping. In each of these cases, it is very likely that participants' naïve organization of these exemplars would conform to the experimenter-defined categories providing lower level, perceptual support for the experimenter-defined, higher-order statistics. In the auditory domain, a small number of statistical learning studies have used variable natural productions of speech sounds (e.g. Thiessen, Hill, & Saffran, 2005), but, to our knowledge, neither the acoustics nor the perceptual similarity of the stimuli have been systematically controlled, and participants had already had substantial experience categorizing the sounds used (speech syllables). Thus, it is unclear how these variable productions were perceptually organized by the learner at the time of statistical exposure and how that perceptual organization interacted with learning higher level statistics. In contrast, the current study explicitly probed participants' perceptual (or lower level) organization of the multiple exemplars and found that the organization did not conform to the experimenter-defined categories used to produce higher-order statistical information. Using this novel paradigm, we were able to examine how information across levels of description affects statistical learning.

If participants were able to learn based on the higher-order statistics determined by the experimenter-defined categories ("top down"), they would have been able to 1) demonstrate learning uniformly across all sound-pair assignments, 2) uncover the intended structure of the acoustic stream, and 3) achieve the best possible prediction for successive sounds. Instead, we find that statistical learning of higher-order regularities is non-uniform across sound-pair assignments. In fact, the pattern of learning across assignments is consistent with the predictions borne from assuming a functional grouping that matches the three groups of exemplars in participants' naïve lower level perceptual organization. This finding suggests that statistical learning, at least in the current task, is constrained by learners' perceptual organization and thus is more "bottom-up" than "top-down". In other words, the intended and optimal learning of higher-order statistics is restricted by lower level perceptual organization.

6.1 Implications for the Development of Language Learning

The current learning paradigm is motivated by the following developmental problem: how are infants able to segment and learn their first words, using higher-order statistical information, without the use of established, adult-like speech sound categorization abilities? The developmental trajectories of word learning and speech sound categorization substantially overlap. While the possibility of simultaneous development of speech sound categorization and word learning has been widely acknowledged, the canonical account of how statistical learning contributes to language acquisition often assumes a sequential

development from lower to higher levels of description in the speech signal (e.g. Romberg & Saffran, 2010; Swingley, 2008; Saffran & Thiessen, 2007). Thus, an understanding of how statistical learning contributes to language acquisition depends on understanding how learning proceeds when statistically-defined structure must be resolved across multiple levels of description and, specifically, when there is tension between lower and higher levels of description is essential.

The current study finds that statistical learning can still occur despite conflict between lower level perceptual organization and higher level statistical information. Specifically, our findings suggest that participants gain some knowledge of the structure of the familiarization stream by relying on their naïve lower level perceptual organization to guide higher-order statistical learning. First, it is important to note that participants do not have specialized or speech-like processing abilities for the current sounds.² Thus, we find that learning can proceed over a variable acoustic stream without specialized categorization abilities. In lieu of these categorization abilities, we find evidence that higher-order statistical knowledge is constrained by the organization of listeners' naïve perceptual similarity space. Specifically, Experiment 3 presents evidence that participants derive their functional groupings (across which higher-order statistics are calculated) from their perceptual similarity organization and can generalize knowledge gained from the learned statistics to novel sounds that occupy the same regions of perceptual similarity space.

Considering this result in relation to language acquisition, one of the immediate implications is that the statistical information present in the speech stream must be considered in relation to the infants' current sound categorization abilities. Previous modeling work has attempted to objectively determine whether the statistical information available to the infant (e.g. the CHILDES corpus) is sufficient for certain aspects of language learning (e.g. Christiansen, Onnis & Hockema, 2009). However, these models have processed the speech signal according to categorization abilities that the infant does not possess. Despite positive findings in these studies, it is likely that this statistical information is not immediately available to the language learner because of their current perceptual organization. Thus, an infant's perceptual organization and categorization abilities must be considered when considering what abilities statistical information might support during language acquisition.

If the uncovering of language structure is indeed constrained by the organization of speech sounds, the early stages of language development might be more dynamic than has been previously characterized. We find evidence that statistical information does affect learning tasks across levels of description with lower levels constraining higher levels. Thus, at any given moment, an infant's knowledge of the structure of language is filtered based on their organization of speech sounds. However, we know that lower level organization has a protracted development with changes occurring well beyond infancy (e.g. Hazan & Barrett,

²One key assumption of this work is that adult learning from novel, complex sound categories is at some level analogous to learning from speech in the first year of life. Broadly, this assumption is based on the view that speech sound categorization is a form of perceptual expertise. It follows that the speech categorization abilities possessed by adults do not effectively transfer to novel, non-speech sound categories. Thus, the cognitive processes of an adult listening to novel sound categories are matched to an infant listening to speech sounds. While this assumption is shared amongst many studies in the statistical and language learning literature, the authors acknowledge that it is important for future research to establish these same effects using speech sounds in infant language learners.

2000; McGowan, Nittouer, & Manning, 2004; Serniclaes Van Heghe, Mousty, Carre, & Sprenger-Charolles, 2004). Current findings suggest that each change in speech sound categorization abilities could potentially have a cascading impact on the statistical information available to the learner. This view presents the statistical language learner as having a much more iterative and dynamic development than a sequential view of language learning from lower to higher level statistical information (see Werker & Curtin, 2005 for a similar view).

In this dynamic view of language learning, perceptual organization, rather than explicit or specialized categorization, may have a larger role in language learning than has previously been acknowledged. Moreover, perceptual organization is likely based, at least in part, on participants' orientation towards acoustically salient features. Recent work has found acoustic contrasts with large, perceptually separable differences are easily distinguished by infants across the first year of life, whether or not these acoustic differences are part of their native environment (Narayan, Werker, & Beddor, 2010) whereas less salient contrasts are slower to emerge (Polka & Werker, 1994). While these results suggest that infants might be able to use salient acoustic differences to distinguish functional categories of sounds, it has yet to be determined whether infants can use relative perceptual similarity to group variable exemplars of sound categories and to subsequently apply this in an on-line task of tracking statistical regularity. However, current results suggest that it is possible for perceptual salience, in supporting lower level organization, to provide an initial entry into learning the structure of the speech stream.

Learning based upon perceptually salient differences could drive changes of functionally-equivalent contrasts that are not supported by initial perceptual organization. A recent study with infants suggests such a possibility: Thiessen (2007) demonstrated that phonemic contrasts that are indistinguishable on their own in a word learning task (daw/taw) are learnable when they have been previously associated with more salient phonemic distinctions (dawbow/tawgoo). According to Thiessen, the acoustically salient difference between 'bow' and 'goo' provides a differential context which can in turn support cognitive separation of the minimal pair of 'taw/daw' necessary for subsequent word learning.³ These results provide some initial evidence that learning based on perceptually distinct sounds can support the acquisition of more difficult phonetic contrasts.

According to Thiessen (2007), this finding is consistent with a distributional account of the development of speech perception. On this view, distributional statistics, based on lower level occurrence of the sound tokens, can shape the perceptual representation of sound categories (see Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Maye, Werker, & Gerken, 2002; Yoshida et al., 2010). It could also be that functional differentiation of sounds is supported by higher-order statistical information in the language, such as the kind of co-occurrence statistics manipulated here (e.g. Beckman & Edwards, 2000). Feldman, Griffiths & Morgan (2009) present evidence from a Bayesian model that lexical acquisition can aid in

³The addition of the perceptually salient endings ('goo' and 'bow') cannot readily be considered as adding a new cue during word learning as these endings were not present during the word learning portion of the task or during an initial familiarization to the minimal pair ('taw/daw').

distinguishing the highly overlapping speech categories. In other words, the higher-order statistical information (i.e. reflective of lexical structure) could be used to shape lower-order speech categories. This model presents one exciting alternative to sequential learning across levels of description in a perceptually and informationally complex signal like language and is potentially complimentary to a distributional account of the development of speech sound categorization.

Like the model presented by Feldman et al., (2009), the higher-order statistical information in the current learning paradigm functionally distinguishes experimenter-defined sound categories. However, we did not find evidence that exposure facilitated perceptual distinction between H1 and H2; this is not surprising given the difficulty of learning this contrast and the length of training required to shift behavior and perceptual treatment of these sounds in previous work (e.g. Wade & Holt, 2005; Leech et al., 2009; Liu & Holt, 2011). Thus, this lack of perceptual shift could be due to the relatively short exposure period used in the experiments or to the fact that the presentation of within-category sounds had a flat distribution. Natural speech categories exhibit a more distinctive within-category distributional structure (e.g. centroid exemplars of a category tend to be more frequently uttered than category outliers), which may make it easier to acquire categories in general. A third possibility is that passive exposure to higher-order transitional probabilities simply do not facilitate sound category learning or need to be supplemented by a more active category learning mechanism. Previous research (Wade & Holt, 2005; Leech et al., 2009; Liu & Holt, 2011) that trained listeners to successfully differentiate H1 from H2 engaged a multi-modal, active, and implicit-reward-based videogame training paradigm. In general, our results suggest that learning functional equivalence classes based on higher-order transitional probabilities is at least harder than learning transitional probabilities derived from perceptual similarity spaces.

Our results do suggest that if a participant's perceptual organization is different, that different higher-order statistical information might be available. Due to the limited perceptual similarity judgment collected for each participant, we were unable to examine whether individual differences in perceptual organization translated into differences in statistical learning. However, in future research, we plan to investigate whether prior training to differentiate H1 from H2 based on this more active videogame paradigm will allow listeners to subsequently treat H1 and H2 as separate functional groups in the statistical learning paradigm used in the present study.

Discussions of how statistical language learning might contribute to language learning have implicitly assumed a sequential order of learning from lower to higher levels of description. If this assumption is not true, as suggested by the overlapping developmental trajectories of word learning and speech sound categorization, it opens up new questions about the potential interaction of learning at multiple levels of description in the speech signal, and the role of perceptual organization and perceptual salience of contrastive features in word learning. The current results demonstrate that the issue of perceptual salience and the interaction of statistical regularities defined at different levels of description are in fact important to consider in statistical learning experiments, and, potentially in language learning.

6.2 Learning from Statistical Regularities in a Perceptually Variable World

Understanding how a mechanism like statistical learning might function in a highly variable and complex natural environment is a central problem in establishing the role of experience in the development of complex perceptual-cognitive abilities. It is equally important to understand how knowledge gained from statistical learning can be generalized. In the current experiments, we demonstrate that adults learn from higher-order transitional probabilities based on perceptually variable exemplars from previously unheard sound categories, and are able to generalize this knowledge to other stimuli that occupy the same region of perceptual space. Further, this learning does not seem to depend on employing symbols or specialized categorical knowledge. Thus, these results are relevant to two issues – the role of variability in learning and the role of symbolic processing in generalization – that are of increasing relevance to statistical learning.

6.2.1 Perceptual Variability and Statistical Learning

In the current experiment, learners received exposure to variable information throughout learning, and possibly because of the variability of experience, they were able to generalize based on their perceptual organization. This is counterintuitive if we conceive of statistical learning as a process by which transitional probabilities are tracked between relatively abstract, symbolic units. We might imagine that perceptual variability would impede such learning by interfering with the learner's ability to identify subsequent instances of the same "unit." A number of recent studies suggest, however, that learners may be better able to extract generalizable or higher-order relationships between stimuli with increased perceptual variability during exposure (Lany & Saffran, 2010; Rost & McMurray, 2009). Specifically, variability in experience might allow infants to distinguish what aspects of the sensory signal are an integral part of a higher-order pattern or invariant (Gibson, 1969) without having to determine *a priori* what part of the stimulus array is relevant as it changes over time.

An alternative possibility is that, regardless of the degree of perceptual variability during experience, learners will be able to generalize knowledge gained through experience with statistical regularities to stimuli of the same perceptual space. This would result in a "generalization gradient" for this statistically-based knowledge. In learning theory, as a stimulus becomes more perceptually distinct from the conditioned stimulus, conditioned responses fall off according to a generalization gradient (e.g. Guttman & Kalish, 1956). It is possible that an analogous effect exists for knowledge gained through exposure to higher-order statistical information: as novel sounds decrease in perceptual similarity, generalization of learning to those sounds will decrease in a related fashion and, once a certain boundary of perceptual space is exceeded, participants will fail to generalize altogether.

Across the lifespan, the learner exists in an environment rich in sensory information, arriving simultaneously from multiple sensory modalities with information across multiple levels of description and dimensions of relevance. It has been unclear how statistical learning mechanisms can operate given the richness of this sensory input and the seemingly unconstrained nature of statistical learning (Romberg & Saffran, 2010). One possibility

suggested here is that the variability of experience that learners experience in their everyday environment may support generalizable learning by drawing the learner's attention to the relevant – or predictive – higher-order information in the sensory input. The relationship between perceptual similarity, variability of experience, and generalization of knowledge in learning from statistical regularities is an important topic for future research.

6.2.2 Perceptual Variability and Symbolic vs. Non-Symbolic Processing

Rule-learning studies tend to engage extreme levels of perceptual variability during exposure and ability to generalize knowledge. These studies typically present infants with a higher-order sequence where each new instance is perceptually novel, and find that infants generalize this sequence to untrained sounds or pictures (e.g. Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Pollack, Siebel, and Shkolnik, 2007). Specifically because of the degree of generalization required by rule-learning studies, it has been suggested that statistical learning and rule-learning represent two different learning mechanisms with the latter requiring abstract, symbolic representations (e.g. Marcus 2003, but see Seidenberg, MacDonald, & Saffran, 2002).

Even in studies of statistical learning, there is an underlying assumption that the statistics are tracked at a relatively abstract level of description. For example, two statistical learning studies have employed variable exemplars of stimuli and saw evidence of generalization (Baldwin et al., 2008; Brady & Oliva, 2008). Both of these studies employed stimuli drawn from categories that participants had considerable experience with: Brady and Oliva (2008) used visual scenes organized by well-learned semantic categories (e.g. kitchens and beaches), while Baldwin et al., (2008) used pictures of common actions such as pointing and grasping. In these studies, participants learned to segment continuous streams of stimuli based on transitional probabilities defined at the category level, with variable exemplars. Importantly, the categories were transparent to the participants, supporting generalization to novel stimuli despite gross differences in the surface forms of the training set. Thus, these studies are most readily interpreted in terms of symbolic learning, where the physical stimulus is first abstracted to a categorical symbol, and transitional probabilities are learned at this level of representation. On this view, the wide variability in the surface forms of categories of symbols is important because it signals that there is a general statistical pattern at work, and not simply a literally repeated series of stimuli to be memorized.

In contrast, the stimuli employed in the current study are not supported by specialized or abstracted categories that already exist for learners. This demonstrates that generalization across variable exemplars is possible without requiring that the exemplars fall into an existing or known symbolic category. Instead, we find some initial evidence that generalization, in this context, can proceed based upon the perceptual organization of novel, unfamiliar stimuli. While it is not clear how perceptual organization alone could support the extreme degree of generalization observed in rule-learning studies, this finding does suggest that generalization without the aid of existing symbolic categories, and influenced by perceptual similarity, may play a role in some phenomena that are considered examples of rule-learning.

6.3 Conclusions

In order for statistical information to aid in complex developmental tasks such as language acquisition, learning from higher-order statistics must be possible while organization at lower levels of description is still developing. This is the first study to examine how such learning would proceed. Broadly, our results suggest some constraints on the statistical learner in this scenario: if higher-order structure is not supported by lower level organization, then the structure is not readily uncovered. In this circumstance, statistical learning appears to be more “bottom-up” than “top-down” (i.e., what is learned at higher levels of description is dependent upon the perceptual organization at lower levels). This finding has important implications for the conceptualization of how statistical learning contributes to language acquisition and more broadly, for how statistical learning mechanisms support the translation of experience into perceptual-cognitive knowledge and abilities.

In relation to language learning: first, statistical information must be considered based on the current categorization abilities of the learner not according to adult categorization abilities because the differences in categorization abilities might obfuscate higher-order statistical information. This entails that the statistical information about the ambient language may be radically different for an infant and an adult. Second, the interrelation of lower level perceptual organization and statistical information present at higher levels suggest a more dynamic and re-iterative process of language learning than previously conceived. Specifically, each change in lower level organization can potentially reveal new statistical information at higher levels of description. Finally, the current results suggest an important role for perception and specifically, perceptual salience as an early and lasting influence on language learning.

These results also have implications for the understanding of how statistical learning mechanisms operate in the unconstrained, variable environment outside the laboratory and what learning outcomes might be expected in the ‘real-world.’ Specifically, this study suggests that statistical learning is robust to perceptual variability even in the absence of symbolic, abstract representations. Moreover, evidence for generalization based on perceptual organization bares the relationship between rule-learning and statistical learning and specifically indicates that generalization is possible without symbolic processing. Finally, the current results also point to a relationship between perceptual variability during experience and generalization of knowledge as a result of that experience and join a handful of other studies suggesting that increased variability might support more robust learning rather than obfuscate statistical information.

Acknowledgments

Thanks to Dani Rubinstein, David Kalkstein, Noa Hertzfeld, Andrew Webb for data collection. Thanks to Drs. Lori Holt and Travis Wade for use of their stimuli and special thanks to Dr. Michael Spivey. This work was funded by in part by R01 DC007694 and R21 DC008969. Thank you to two anonymous reviewers for their very helpful comments guiding a revision of this work.

7.0 Appendix 1: Deriving Learning Predictions based on Familiarization and Foil Transitional-Probabilities

In Experiments 2 and 3, we compare predictions of learning performance and patterns of actual performance to determine what functional groupings participants are using to uncover statistical information. Crucial to this procedure is that different statistical information is uncovered in the stream if one employs 3-functional-groupings vs. 4-functional-groupings. The bottom panel of Figure 3 presents the predictions for two different foil types and across sound-pair assignments. Predictions for learning are based upon both the reliability of transitional probabilities during familiarization and the transitional probabilities of the foils used at test. This appendix presents in detail how the values of transitional probabilities were used to derive predictions.

Differences in the functional groups used by participants during familiarization to uncover higher-order statistical information affect the resulting higher-order statistical information and correspondingly statistical learning outcomes. In brief, we first evaluate the transitional probabilities during familiarization and determine the robustness or reliability of statistical information. Transitional probabilities are considered more robust if there is a large difference between the within- and between-pair probabilities. If statistical information (transitional probabilities) is not reliable during familiarization, we predict that there will be no evidence of learning during test. If statistical information *is* reliable during familiarization, we then considered the relationship between the within-pair transitional probabilities and the transitional probabilities of the foils employed at test. The transitional probabilities of the foils were calculated relative to the familiarization stream (i.e. how likely is that combination of sounds given familiarization experience). If the transitional probabilities of the foils are lower than the within-pair transitional probabilities, then we predict participants will be able to discriminate pairs from foils at test and show evidence of learning. However, if the foil transitional probabilities are not lower than the within-pair transitional probabilities, we predict that it will not be possible for participants to distinguish pairs from foils at test and thus, they will not show evidence of discrimination at test. We employed two different types of foils to produce two distinct patterns of predictions. See Table 3 for a summary of the transitional probabilities for within-pairs and the foils employed during test. It is important to note that the between-pair transitional probabilities are higher than typical statistical learning experiments (around 50%). This is the case for both 3- and 4-functional-grouping cases and is due to the simplicity of the stream's structure (two pairs of sounds; see Kirkham et al., 2002 for a study with similar transitional probabilities).

We now move onto the application of this method to specific conditions and generating the pattern of predictions presents in the bottom of Figure 3. Starting with the “top-down” account, if participants use four functional categories, that conform to the experimenter-defined categories, transitional probabilities highlight the sound-pair structure with a within-pair transitional probability of 100%. This familiarization also produces foils with 0% transitional probabilities at test for all sound-pair assignment structures. This is consistent with the experimenter-intended organization of the stream; a more detailed discussion of

these methods is presented in 3.1.3 and 3.1.5. These two factors—the transitional probabilities during familiarization and the transitional probabilities of test foils, determined relative to familiarization—determine the prediction for learning and are summarized in Table 3. The first row of Table 3 presents this information for four functional groups consistent with a “top-down” view of statistical learning. This information is presented for all three sound-pair assignments. We predict successful statistical learning when participants use four functional groups uniformly in all sound-pair assignments. In other words, if statistical learning is more “top-down,” participants will show evidence of learning for all sound-pair assignments structures (see the top-row of the table in Figure 3).

Turning to the learning predictions for “bottom-up” statistical learning where participants employ only three functional groups corresponding to E1 and E2 experimenter-defined categories and an amalgamated H category (including both H1 and H2 exemplars), the non-uniform grouping of exemplars by a “bottom-up” statistical learner results in different learning predictions across sound-pair assignments. The transitional probabilities during familiarization and for the foils at test are presented in the bottom row of Table 3.

First, we determined whether participants would receive reliable or robust statistical information during familiarization for each sound-pair assignment (note: across foil-types, the same familiarization was employed and thus, there are not differences in predictions made based on familiarization across these conditions). For the predictions consistent with the “top-down” condition, transitional probabilities during familiarization would be 100% within-pairs and 50% between pairs. We assume that as the values with within- and between-pair become quantitatively closer, the statistical information is considered to be less reliable or robust and participants will not be able to learn about the structure of the stream from transitional probabilities. To our knowledge, it has not been systematically studied how different within- and between-structure transitional probabilities must be for statistical learning to take place. The majority of statistical learning experiments employ a very high within-pair or within-structure (e.g. triplets) transitional probabilities, typically 100%. Between-pair or between-structure transitional probabilities are typically much lower. In the canonical paradigm of the literature, first employed by Saffran et al., 1996, within-structure transitional probabilities are 100% and between-structure transitional probabilities are 33%. Rarely have paradigms employed transitional probabilities that are closer in value than this experiment. Thus, we assume that for learning to take place, transitional probabilities for within-structure (in this case, within-pair) must be *much higher* than the between-structure transitional probabilities. Consistent with these requirements, the sound-pair assignments E-E/H-H and E-H/E-H present higher-order transitional probabilities that highlight the sound-pair structure by having very high transitional probabilities within sound-pairs and much lower transitional probabilities between-pairs (left two columns in Table 3). Note that there are slightly weaker transitional probabilities for the latter sound-pair assignment structure. However, the H-E/E-H sound-pair assignment structure presents transitional probabilities that do not readily and reliably highlight the sound-pairs. Specifically, one sound-pair only has a transitional probability of 50% and is comparable to the between-pair transitional probabilities. These transitional probabilities are much less reliable than previous statistical learning studies and correspondingly, we do not predict any learning for this sound-pair

assignment structure for the H-E/E-H sound-pair assignment structure (presented in the last column of Table 3). To summarize, using 1) the principal that in order for learning to take place within-pair transitional probabilities must be much higher than between-pair transitional probabilities and 2) 100% within-structure and below 50% between-structure transitional probabilities, as per previous statistical learning experiments, we predict that there will not be learning in the H-E/E-H sound-pair assignment and that there is a possibility of learning in the first two sound-pair assignments, E-E/H-H and E-H/E-H.

For the two sound-pair assignments that have reliable statistical information during familiarization, we then evaluated the foils employed at test. In order to differentiate foils from sound-pairs, the foils must have much lower transitional probabilities, similar to between-pair words. Again, transitional probabilities for the foils are determined relative to the familiarization stream. The foil transitional probabilities are only relevant for the learning prediction for the EH/EH or EE/HH sound-pair assignment structures as both of these groups received learnable transitional probabilities during familiarization. Type 1 foils have a very low transitional probability for both EH/EH and EE/HH sound-pair assignment structures and thus we predict learning for both of these structures. Among Type 2 foils, only the foils for the EE/HH sound-pair assignment structure are lower than the sound-pairs. For the EH/EH group, the transitional probabilities of the foils are indistinguishable from the sound-pairs and thus we predict no learning. See the bottom-row of Table 3 for presentations of the foil transitional probabilities but see Table 2 for a clearer illustration of the differences in predictions for Type 1 and Type 2 foils for the EH/EH sound-pair assignment structure. Thus, there are different predictions for learning across foil types if statistical learning is “bottom-up” or employs three functional groups.

8.0 Appendix 2: Perceptual organization from Experiments 2–3

The current study examines how “bottom-up” influences of perceptual organization affect learning from higher-order statistical regularities. To this end, all experiments include a perceptual similarity task designed to directly examine participants’ perceptual organization of the stimuli. Experiment 1 assessed perceptual organization for participants without prior exposure to the stimuli (i.e., naïve participants) while Experiments 2 assessed perceptual organization after exposure during a statistical learning task. Experiment 3 employed novel exemplars at test and determined whether participants had the same perceptual organization, assessed after statistical learning test, as the stimuli employed during familiarization and the only stimuli employed in Experiments 1 and 2. Identical inferential tests, accompanied with multidimensional scaling (MDS) analyses, were employed for each experiment separately. Overall, we find very consistent results across groups and before and after exposure. This appendix contains the inferential tests and graphical representations of the MDS analyses for Experiment 2 (separately for each Foil Type) and Experiment 3.

8.1 Perceptual similarity judgments for Experiment 2

The same analyses as those used in Experiment 1 were performed separately for the participants in each foil-type condition. For participants in the Type 1 foil condition, a oneway ANOVA examining the effects of three types of perceptual similarity trials (within

Easy or Hard categories: E vs. E, H vs. H, and between Easy and Hard categories: E vs. H) revealed a significant effect of contrast type, $\min F' (2, 176) = 20.858, p < 0.0001$. The means for these contrasts are higher for between-category types (E vs. H: 2.878) than within-category trials (E vs. E: 2.398, H vs. H: 2.167) establishing that participants distinguish between Easy and Hard exemplars. Next, we examined how participants perceive the four experimenter-defined categories (i.e. E1, E2, H1, H2), by crossing comparison type (Same vs. Different category) with category type (Easy vs. Hard). As with Experiment 1, a significant interaction between these factors was observed, $\min F' (1, 125) = 26.322, p < 0.0001$, such that there was a large difference in perceived similarity for Same vs. Different comparisons that was restricted to the Easy categories. That is, consistent with the MDS analysis, participants perceived the Easy items as forming two clusters roughly corresponding to the experimenter-defined categories, whereas the Hard items were equally similar between and within these categories. There also was a main effect of Same-Different judgments, $\min F' (1, 67) = 52.640, p < 0.0001$ but no main effect of Easy-Hard trials. A planned comparison revealed that the effect of comparison type was significant only for stimuli from the Easy categories: $\min F' (1, 73) = 90.738, p <$

For participants in the Type 2 foil condition, a significant effect on perceptual similarity judgments was also found in analyses considering the Easy and Hard stimuli, $\min F' (2, 189) = 41.421, p < 0.0001$. The means for these contrasts are higher for between-category types (E vs. H: 3.000) than within-category trials (E vs. E: 2.288, H vs. H: 2.040). These results indicate that participants make a distinction between Easy and Hard exemplars. Investigating whether participants make further distinctions by examining the effects of same-different judgments *within* Easy and Hard categories revealed a main effect of Same-Different judgments, $\min F' (1, 135) = 18.223, p < 0.0001$ but no main effect of Easy-Hard trials. In addition, there was an interaction of Easy and Hard with Same-Different judgments: $\min F' (1, 133) = 21.910, p < 0.0001$. Examining Same vs. Different judgments separately for Easy and Hard categories, there was no effect of Same vs. Different judgments for the Hard category trials, but a significant effect of Same vs. Different trials in the Easy categories: $\min F' (1, 82) = 42.169, p < 0.0001$. These results indicate that participants perceive the difference between the two Easy categories but not the difference between the two Hard categories, consistent with both Experiment 1 results and the results from the Type 1 foil condition.

8.2 Perceptual similarity judgments for Experiment 3

While using the same stimuli as familiarization in Experiments 1 and 2, Experiment 3 introduced novel stimuli during the test of statistical learning. These stimuli were derived from the same novel categories but had not been heard before by participants in the experiment. After completing the statistical learning task, participants completed a perceptual similarity task for these novel stimuli (5 exemplars per category). We conducted the same inferential tests on these novel stimuli as on all previous perceptual similarity data sets. Again, we find a striking similarity between perceptual organization of these novel stimuli and perceptual organization of familiarization stimuli after exposure (Experiment 2) and prior to exposure (Experiment 1).

A oneway ANOVA examining the effects of three types of perceptual similarity trials (within Easy or Hard categories: E vs. E, H vs. H, and between Easy and Hard categories: E vs. H) revealed a significant effect of contrast type, $\min F'(2, 181) = 35.314, p < 0.0001$. The means for perceptual similarity judgments are higher for between-category types (E vs. H: 2.901) than within either E or H category trials (E vs. E: 2.051, H vs. H: 2.092). We then examined how participants perceive the four experimenter-defined categories (i.e. E1, E2, H1, H2), by crossing comparison type (Same vs. Different category) with category type (Easy vs. Hard). As with previous experiments, a significant interaction between these factors was observed, $\min F'(1, 101) = 19.486, p < 0.0001$, such that there was a large difference in perceived similarity for Same vs. Different comparisons that was restricted to the Easy categories. There also was a main effect of Same-Different judgments, $\min F'(1, 71) = 41.327, p < 0.0001$ but no main effect of Easy-Hard trials.

References

- Altmann G, Mirković J. Incrementality and prediction in human sentence processing. *Cognitive Science*. 2009; 33:583–609. [PubMed: 20396405]
- Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics of 8-month-old infants. *Psychological Science*. 1998; 9:321–324.
- Baldwin D, Andersson A, Saffran J, Meyer M. Segmenting dynamic human actions via statistical structure. *Cognition*. 2008; 106:1382–1407. [PubMed: 18035346]
- Beckman ME, Edwards J. The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development*. 2000; 71:240–249. [PubMed: 10836579]
- Brady TF, Oliva A. Statistical learning using real-world scenes: extracting categorical regularities without conscious intent. *Psychological Science*. 2008; 19:678–685. [PubMed: 18727783]
- Christiansen MH, Onnis L, Hockema SA. The secret is in the sound: From unsegmented speech to lexical categories. *Child Development*. 2009; 12:388–395.
- Clayards M, Tanenhaus MK, Aslin RN, Jacobs RA. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*. 2008; 108:804–809. [PubMed: 18582855]
- Feldman, NH.; Griffiths, TL.; Morgan, JL. Learning phonetic categories by learning a lexicon. In: Taatgen, N.; van Rijn, H., editors. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society; 2009. p. 2208-2213.
- Fernald A, Hurtado N. Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*. 2006; 9:F33–F40. [PubMed: 16669790]
- Fiser J, Aslin RN. Statistical learning of new visual feature combination by infants. *Proceedings of the National Academy of Sciences*. 2002; 99:15822–15826.
- Flynn MC, Dowell RC. Speech perception in a communicative context: An investigation using question/answer pairs. *Journal of Speech, Language, and Hearing Research*. 1999; 42:540–552.
- Gibson, EJ. *Principles of Perceptual Learning and Development*. East Norwalk, CT, US: Appleton-Century-Crofts; 1969.
- Goldinger SD. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*. 1998; 105(2):251–279. [PubMed: 9577239]
- Goldwater S, Griffiths TL, Johnson M. A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*. 2009; 112:21–54. [PubMed: 19409539]
- Graf-Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? *Psychological Science*. 2007; 18:254–260. [PubMed: 17444923]
- Guttman N, Kalish HI. Discriminability and stimulus generalization. *Journal of Experimental Psychology*. 1956; 51:79–88. [PubMed: 13286444]
- Hazan V, Barrett S. The development of phonemic categorization in children aged 6 to 12. *Journal of Phonetics*. 2000; 28:377–396.

- Kamide Y, Altmann G, Haywood SL. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*. 2003; 49:133–156.
- Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*. 2002; 83:B35–B42. [PubMed: 11869728]
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 1992; 255:606–608. [PubMed: 1736364]
- Lany J, Saffran JR. From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science*. 2010; 21:284–291. [PubMed: 20424058]
- Leech R, Holt LL, Devlin JT, Dick F. Expertise with artificial non-speech sounds recruits speech-sensitive cortical regions. *Journal of Neuroscience*. 2009; 29:5234–5239. [PubMed: 19386919]
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:431–461. [PubMed: 4170865]
- Liu R, Holt LL. Changes in mismatch negativity reflecting the acquisition of complex, non-speech auditory categories. *Journal of Cognitive Neuroscience*. 2011; 23:683–698. [PubMed: 19929331]
- Lotto AJ. Language Acquisition as Complex Category Formation. *Perceptual Processing*. 2000; 57:189–196.
- Lotto, AJ.; Holt, LL. The illusion of the phoneme. In: Billings, SJ.; Boyle, JP.; Griffith, AM., editors. *Chicago Linguistic Society*. Vol. 35. Chicago: Chicago Linguistic Society; 2000. p. 191-204. The Panels
- Marcus GF, Vijayan S, Bandi Rao S, Vishton PM. Rule learning by seven-month-old infants. *Science*. 1999; 283:77–80. [PubMed: 9872745]
- Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*. 2002; 82:B101–B111. [PubMed: 11747867]
- McClelland JL. Prediction-error driven learning: The engine of change in cognitive development. *Proceedings of the 2nd Annual IEEE Conference on Development and Learning*. 2002:43.
- McGowan R, Nitttrouer S, Manning C. Development of [r] in young, Midwestern, American children. *Journal Of The Acoustical Society Of America*. 2004; 115:871–884. [PubMed: 15000198]
- Narayan CR, Werker JF, Beddor PS. The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*. 2010; 13:407–420. [PubMed: 20443962]
- Newman R, Ratner NB, Jusczyk AM, Jusczyk PW, Dow KA. Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*. 2006; 42:643–655. [PubMed: 16802897]
- Nusbaum, HC.; Henly, AS. Listening to speech through an adaptive window of analysis. In: Schouten, MEH., editor. *The Auditory Processing of Speech: From Sounds to Words*. Berlin: Walter de Gruyter & Co; 1992. p. 339-348.
- Oakes LM, Cohen LB. Infant perception of a causal event. *Cognitive Development*. 1990; 5:193–207.
- Oakes, LM.; Cohen, LB. Infant causal perception. In: Rovee-Collier, C.; Lipsitt, LP., editors. *Advances in Infancy Research*. Vol. 9. Norwood, NJ: Ablex Publishing Corporation; 1995. p. 1-54.
- Obleser J, Kotz SA. Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*. 2010; 20(3):633–540. [PubMed: 19561061]
- Onnis L, Waterfall HR, Edelman S. Learn locally, act globally: Learning language from variation set cues. *Cognition*. 2008; 109:423–430. [PubMed: 19019350]
- Peterson GE, Barney HL. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Polka L, Werker JF. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*. 1994; 20:421–435. [PubMed: 8189202]
- Port R. How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*. 2007; 25:143–170.

- Poeppell D, Idsardi WJ, Van Wassenhove V. Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2008; 363:1071–1086.
- Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science*. 1981; 212:947–949. [PubMed: 7233191]
- Rost GE, McMurray B. Speaker variability augments phonological processing in early word learning. *Developmental Science*. 2009; 12:339–349. [PubMed: 19143806]
- Romberg AR, Saffran JR. Statistical learning and language acquisition. *Wiley Interdisciplinary Review of Cognitive Science*. 2010; 1:906–914.
- Saffran JR, Aslin R, Newport EL. Statistical learning by 8-month-old infants. *Science*. 1996; 274:1926–1929. [PubMed: 8943209]
- Saffran JR, Pollack SD, Seibel RL, Shkolnik A. Dog is a dog is a dog: Infant result learning is not specific to language. *Cognition*. 2007; 105:669–680. [PubMed: 17188676]
- Saffran, JR.; Thiessen, ED. Domain-general learning capacities. In: Hoff, E.; Shatz, M., editors. *Handbook of Language Development*. Cambridge: Blackwell; 2007. p. 68-86.
- Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science*. 1997; 275:1593–1599. [PubMed: 9054347]
- Serniclaes W, Van Heghe S, Mousty P, Carre R, Sprenger-Charolles L. Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*. 2004; 87:336–361. [PubMed: 15050458]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995; 270:303–304. [PubMed: 7569981]
- Sobel DM, Tenenbaum JB, Gopnik A. Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. *Cognition*. 2004; 28:303–333.
- Shankweiler, D.; Strange, W.; Verbrugge, R. Speech and the problem of perceptual constancy. In: Shaw, R.; Bransford, J., editors. *Perceiving, acting, knowing: Toward an ecological psychology*. NJ: Erlbaum; 1977. p. 315-345.
- Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*. 1997; 388:381–382. [PubMed: 9237755]
- Swingley D. The roots of early vocabulary in infants' learning from speech. *Current Directions in Psychological Science*. 2008; 17:308–312. [PubMed: 20523916]
- Thiessen ED. The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*. 2007; 56:16–34.
- Thiessen ED, Hill EA, Saffran JR. Infant-directed speech facilitates word segmentation. *Infancy*. 2005; 7(1):53–71.
- Thiessen, ED.; Saffran, JR. Domain-general learning capacities. In: Hoffman, E.; Shatz, M., editors. *Blackwell Handbook of Language Development*. Oxford: Blackwell Publishing; 2007. p. 68-86.
- Venables, WN.; Ripley, BD. *Modern applied statistics with S-Plus*. 2. Springer; New York: 1997.
- Wade T, Holt LL. Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of Acoustical Society of America*. 2005; 118:2618–2633.
- Waterfall, HR. A little change is a good thing: Feature Theory, Language Acquisition and Variation Sets. Chicago, IL: University of Chicago, Department of Linguistics; 2006. Ph.D. Dissertation
- Werker JF, Curtin S. PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*. 2005; 1(2):197–234.
- Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*. 1984; 7:49–63.
- Yoshida KA, Pons F, Maye J, Werker JF. Distributional phonetic learning at 10 months of age. *Infancy*. 2010; 15:420–433.
- Yuille A, Kersten D. Vision as Bayesian inference: analysis by synthesis. *Trends in Cognitive Science*. 2006; 10:301–308.
- Zevin JD. A Sensitive Period for Shibboleths: The Long Tail and Changing Goals of Speech Perception Over the Course of Development. *Developmental Psychobiology*. (in press).

- We examine two competing predictions (bottom-up vs. top-down) about how listeners' will learn statistical relationships across experimenter-defined categories composed of acoustically-varying exemplars
- Listeners' naïve perceptual organization of sound exemplars are measured
- Results support the predictions of a bottom-up model in which naïve perceptual organization of exemplars constrains the nature of the higher-level statistics that are learned

Figure 1. Schematic diagram of the spectrotemporal properties of the stimuli employed in all experiments. Each sound has two components: P1 (constant over all stimuli in a category) and P2 (varies for each stimulus).

Figure 2. Perceptual Distance between stimuli, averaged across all subjects. Similarity judgments were entered into an MDS analysis with two dimensions.

Figure 3.

Bottom: Predictions of post-test performance (checkmarks indicate expected above-chance performance, Xs indicate expected performance no better than chance) that result from either employing a top-down (4-functional-group) or bottom-up (3-functional-group) approach. Top: Actual behavioral data, grouped by Sound-Pair Assignment type. The data are consistent with the predictions derived from the bottom-up account (e.g. participants are assuming 3 functional groups). Each bar represents results from 13–15 participants and a total of 85 participants.

Figure 4. Perceptual Distance between stimuli, averaged across all subjects for Foil Type 1 in Experiment 2. Similarity judgments were entered into an MDS analysis with two dimensions.

Figure 5. Perceptual Distance between stimuli, averaged across all subjects for Foil Type 2 in Experiment 2. Similarity judgments were entered into an MDS analysis with two dimensions.

Figure 6. Perceptual Distance between novel stimuli employed at test, averaged across all subjects in Experiment 3. Similarity judgments were entered into an MDS analysis with two dimensions.

Table 1

Example sound-pair assignments from each of the three basic classes of sound-pair ‘structures’, along with corresponding examples of a randomly generated familiarization stream (subscript numbers indicate which of the six exemplars per category was randomly selected for that position in the stream). Although only one example per sound-pair structure is shown here, we included conditions for both possible sound-pairs that conformed to each structure type (e.g. for the EE, HH structure type, we had one condition where the sound-pairs were E1-E2 and H1-H2 as well as a condition where the sound-pairs were E2-E1 and H2-H1).

General Sound-Pair Structure	Example Sound-Pair Assignments	Example Familiarization Stream
E-E, H-H	E1-E2 H1-H2	E1 ₂ - E2 ₄ - H1 ₁ - H2 ₂ - H1 ₆ - H2 ₅ - E1 ₁ - E2 ₃ - H1 ₃ - H2 ₆ ...
E-H, E-H	E1-H1 E2-H2	E2 ₄ - H2 ₁ - E1 ₁ - H1 ₅ - E2 ₆ - H2 ₂ - E2 ₅ - H2 ₃ - E1 ₂ - H1 ₁ ...
H-E, E-H	H1-E2 E1-H2	E1 ₆ - H2 ₃ - H1 ₂ - E2 ₁ - E1 ₃ - H2 ₅ - H1 ₅ - E2 ₅ - H1 ₄ - E2 ₂ ...

Table 2

Transitional probabilities for one example sound-pair assignment (E1-H1, E2-H2). Of critical importance is that post-test performance behavioral predictions differ, when comparing Sound-pairs vs. Type 2 Foils, depending on what functional grouping (that is, grouping into three vs. four categories) participants are assuming in learning transitional probabilities during familiarization. This provides a clear way to use the post-test performance patterns to assess whether participants are employing a top-down or bottom-up approach to learning the statistics.

Functional grouping	Sound-pairs	Type 1 Foils		Type 2 Foils		Predicted post-test performance
Top-Down (4-group)	E1 → H1	100%	E1 → E2	0%	E1 → H2	0%
	E2 → H2	100%	H1 → H2	0%	E2 → H1	0%
E1/E2/H1/H2						SP vs. Type 2 Foils: above chance
Bottom-Up (3-group)	E1 → H	100%	E1 → E1	0%	E1 → H	100%
	E2 → H	100%	H → H	0%	E2 → H	100%
E1/E2/H						SP vs. Type 2 Foils: below chance

Table 3

Transitional probabilities of Sound-pairs vs. Type 1 and Type 2 Foils resulting from each combination of (1) word-pair assignment structure and (2) perceptual grouping assumption. Each cell contains the transitional probabilities of both sound-pairs and both foils used in each condition.

	E-E, H-H			E-H, E-H			H-E, E-H		
	Sound Pairs	Type 1 Foils	Type 2 Foils	Sound Pairs	Type 1 Foils	Type 2 Foils	Sound Pairs	Type 1 Foils	Type 2 Foils
Functional grouping									
Top-Down (4-group) E1/E2/H1/H2	100% 100%	0% 0%	0% 0%	100% 100%	0% 0%	0% 0%	100% 100%	0% 0%	0% 0%
Bottom-Up (3-group) E1/E2/H	100% 75%	0% 25%	0% 0%	100% 100%	0% 0%	100% 100%	100% 50%	25% 50%	0% 25%